

西南交通大学

硕士学位论文

大庆市PM10污染预测的研究

姓名：王顶

申请学位级别：硕士

专业：环境工程

指导教师：杨巧艳

201105

摘 要

可吸入颗粒物 PM_{10} 是大庆市大气中首要污染物,为了更好的反映出 PM_{10} 的污染变化规律、加强污染防治工作和预防严重污染事件的发生,及时开展 PM_{10} 污染预报工作意义重大而深远。

本文以 2008 年 1 月 1 日至 2010 年 6 月 30 日的全市 PM_{10} 逐日 API 值和对应的逐日气象数据资料为基础,对大庆市 PM_{10} 污染预报进行了研究。

首先采用概率统计方法分析了大庆市 PM_{10} 污染的年度、季节、月的污染变化特征及趋势,重点对采暖期与非采暖期进行比对分析,对沙尘暴的影响、过年期间烟花爆竹的影响进行了单独分析,并且探讨了污染变化的成因。

其次通过对大庆市近年来的气象要素数据进行统计分析,总结出各季节气象因子的变化规律和 PM_{10} 污染高值日气象因子变化规律,以便从中找出 PM_{10} 污染与气象因子之间的关系。同时采用 SPSS 软件对 PM_{10} 污染指数与气象因子之间的相关性进行了分析,初选出对 PM_{10} 污染影响显著的气象要素,作为后续进行 PM_{10} 污染预报的依据。

最后在初选因子的基础上利用逐步回归法和主成分分析法进行了相关因子的精选,并分别建立了大庆市四季 PM_{10} 污染的多元回归预报模型和 B-P 神经网络预报模型,同时对这两种模型的预报准确率进行了检验和比对研究。得出多元回归模型对四季 PM_{10} 污染的 API 值的预报准确率分别为 73.9%、77.6%、77%、79.7%,预报的精确度评分分别为 76.8、82.3、80.7、83.3。B-P 神经网络模型对四季 PM_{10} 污染的预报准确率分别为 77.9%、83.2%、75.9%、81.6%,预报的精确度评分分别为 83.6、86.6、81.5、84.6。由此可见,本文所建立的两两种大庆市 PM_{10} 污染预报模型都可以满足实际预报的需要,但 B-P 神经网络所建立的污染预报模型要比传统的统计方法拥有更好的预测能力,这也反映出 B-P 人工神经网络在处理大气污染预报这类有着较强非线性变化特点的问题时具有明显优势。

关键词: PM_{10} ; 大气污染预报; 逐步回归; 主成分分析; B-P 神经网络

Abstract

Inhalable particulates(PM_{10}) has been the primary pollutant in the air of Daqing city for a long time. For better reflecting the pollution change tendency, enhancing pollution control and preventing serious pollution accidents, develop pollution forecasting work in time are of great significance. In this paper, the forecasting of PM_{10} pollution in Daqing city had been studied according to the city PM_{10} daily API values that monitored from 1 January 2008 to June 30, 2010 and the corresponding daily meteorological data for the data base.

Analyzed the Daqing PM_{10} pollutant's variations and trends of pollution by annually, seasonal, monthly with probabilistic statistical method, with emphasis on heating period and non-heating period for comparison and analysis, at the same time to do a separate analysis on the impact of dust storms and the impact of fireworks during the Chinese New Year, and to explore the causes of pollution changes.

PM_{10} pollution is closely-related with meteorological factors, through the statistical analysis of recent years meteorological elements data in Daqing city, summed up the seasonal changes of meteorological factors. And analyzed the high API values day's variation of meteorological factors in order to better to find out the relationship between PM_{10} pollution and meteorological factors. Using SPSS software to make a correlation analysis between meteorological factors and PM_{10} pollution, then choose the significant impact meteorological factors as a basis for prediction.

Based on the process of primary factor, the stepwise regression method and principal component analysis method are used on related factors selection. And then we tried to establish two PM_{10} pollution's prediction model of four seasons in Daqing city, one model is multiple regression prediction model, another prediction model is based on B-P artificial neural network. The models' prediction accuracy rates were examined in the paper and the differences between the two prediction models are also compared in the paper. By means of the simulation and practice, the multiple regression prediction model's precision of spring is 73.9%, summer's is 77.6%, autumn's is 77% and winter's is 79.7%. Four seasons forecast accuracy scores were 76.8、82.3、80.7 and 83.3. The B-P artificial neural network prediction

model's precision of spring is 77.9%,summer's is 83.2%,antumn's is 75.9% and winter's is 81.6%. Four seasons forecast accuracy scores were 83.6、86.6、81.5 and 84.6.The results shows that two established PM_{10} pollution prediction model can meet the needs of the actual forecast.But the B-P neural network prediction model have better predictive ability than traditional statistical method,it also reflects the B-P artifical neural network have obvious advantages in dealing with non-linear problems such as pollution forecast problems.

Key words: PM_{10} pollution; atmospheric pollution forecasting; stepwise regression;
B-P neural networks; principal component analysis

第一章 绪 论

1.1 论文选题背景

1.1.1 大气污染及其危害

现如今由于全球经济社会的飞速发展，工业化、城市化的不断推进，人口数量的持续增加，人类的可持续发展意识淡漠等原因导致资源能源过度消耗，生态环境污染加剧。人类已经走到了该仔细审视资源环境保护与自身发展之间关系的重要关口，怎样使人类的发展与环境的承载力相适应，更好的生存繁衍，已经成为全世界面临的主要问题。

大气环境保护作为生态环境保护的一个重要切入点与结合点，正面临着前所未有的巨大挑战，以二氧化硫、氮氧化物、颗粒物为主要污染物的大气污染问题日趋严重，威胁着可持续发展的基础。根据国际标准化组织（ISO）的概念，大气污染是指由于人类活动和自然过程引起某些污染物或其二次污染物进入大气，累积到了足够的浓度、持续了足够长的时间，并因此而危害了人类的健康舒适和危害了自然环境的现象。它分为自然过程（如火山喷发、森林大火等）和人为因素（如汽车尾气、生活燃煤、工业废气、核爆炸等）两种，并且主要以后者为主，尤其是工业污染和交通运输污染所造成的。其过程主要由污染物的排放，大气介质传播，人、物、环境的受害这三个环节构成。

20 世纪以来，世界上相继发生了多起严重的大气污染突发事件，比如 1930 年的比利时马斯河谷工业大气污染事件，1952 年的伦敦烟雾惨案，1953 年的洛杉矶光化学烟雾污染等，早期的这些大气污染，很短时间内就会在局部地区形成高浓度的污染物，给人和动物造成毁灭性伤害。这些灾难为人们敲响了警钟，得到了各国政府的高度重视。20 世纪 60 年代以来，一些国家采取了污染物控制措施，采用控制污染物排放和建造高烟囱使污染物扩散等方式，使污染物的近地面浓度有所降低，但是却造成了远离污染源的大范围大气污染。同时由于内燃机的广泛使用，石油燃烧造成的污染成为了新的污染源，大气环境保护趋势越来越不容乐观。近几十年来，随着汽车工业的快速发展，大量汽车尾气的排放及其在阳光下形成的光化学烟雾又进一步加重了大气污染程度，大气污染的防治已经到了刻不容缓的地步。进入 21 世纪，形势更加严峻，一次能源加剧消耗造成的传统污染，植被遭破坏造成的沙尘污染，长期累

积的大气环境问题, 以及许多潜在的新型污染使我们赖以生存的大气环境风险剧增。同时各国进行的大气核试验产生的放射性物质, 火山喷发产生的火山灰等也大大增加了全球性大气污染的隐患。因此, 大气环境污染是目前全球最为重视的环境问题之一。

目前对环境和人类产生危害的污染物约有 100 种, 其中影响范围较广的污染物主要分为以下两类, 其一是有害气体, 包括 SO_2 、 NO_x 、碳氢化合物、光化学烟雾和卤族元素等, 其二是颗粒物, 包括粉尘、酸雾和气溶胶等。这些大气污染物对人体的主要危害表现为较易引起急慢性中毒、咽喉炎、支气管炎、肺气肿、哮喘、鼻炎、皮肤方面疾病以及神经系统疾病、增加患癌症几率等^[1 2]; 对植物可使其生理机制受压抑, 成长不良, 抗病虫能力减弱, 甚至死亡; 大气污染还能对气候产生不良影响, 如导致能见度降低, 太阳辐射减少, 温室效应产生等。

1.1.2 PM_{10} 概念及其危害

大气环境中的悬浮颗粒物因其对人体健康、生物生长发育、气候等方面的负面影响巨大而倍受人们关注。在我国的环境空气质量标准中, 根据粉尘颗粒的大小分为总悬浮颗粒物(Total suspended particles)和可吸入颗粒物(Inhalable particles)。总悬浮颗粒物(TSP)指能悬浮在空气中, 空气动力学当量直径 $\leq 100 \mu\text{m}$ 的颗粒物。而所谓可吸入颗粒物, 即 PM_{10} , 是指悬浮在空气中, 空气动力学当量直径 $\leq 10 \mu\text{m}$ 的颗粒物^[3]。由于其极易随人们的呼吸而进入呼吸系统, 故也称其为可吸入颗粒物。

近年来, 随着各国对 PM_{10} 的化学组成、存在状态、物理化学特性以及 PM_{10} 在大气环境中的时空分布、迁移特性等方面研究的深入, 人们越来越意识到 PM_{10} 的危害性。本文在参阅大量文献资料的基础上, 对其危害性归纳如下:

(1) 严重危害人体健康。颗粒物根据其粒径大小的不同, 被人吸入沉积于呼吸系统的部位也不同, 所造成的危害也不同。 $5 \sim 10 \mu\text{m}$ 的颗粒物, 大部分沉积于呼吸道, 被分泌的粘液吸附并排出。粒径小于 $5 \mu\text{m}$ 的颗粒物, 大部分进入肺部, 当颗粒粒径小于 $2.5 \mu\text{m}$ 时, 则会深入沉积于肺腔内甚至进入血液循环, 对人体危害极大。同时, 由于较细小的颗粒物组成的复杂结构集合体比大颗粒组成的简单结构集合体比表面积大, 因而更加容易吸附一些对人体健康有害的重金属及有机物, 因而毒性更大^[4]。有关研究

表明, 滞留于上呼吸道的颗粒物, 会与进入人体的 SO_2 等有害气体联合作用刺激和腐蚀呼吸道粘膜、纤毛, 引起发炎症状和增加气道阻力。滞留在细支气管与肺泡的颗粒物也会与 NO_2 等有害气体联合作用而损伤肺泡和粘膜, 引发支气管与肺部炎症。长此以往, 还会诱发肺部疾病并出现继发性感染, 进而导致肺病与心脏病死亡率上升。

Zanobetti 等人^[5]通过研究分析了 1985~1994 十年间心脏病、慢性肺病、肺炎等疾病住院患者的发病情况, 发现患有呼吸系统疾病并受 PM_{10} 影响的心血管病患者, 住院率显著增高, 受 PM_{10} 危害的哮喘患者、传导和节律障碍患者、心率衰竭患者的发病住院率都明显增高。同时, PM_{10} 的危害与性别、年龄、人种等都无关。Lippmann 等人^[6]通过分析加拿大温莎市多年统计数据, 对颗粒物浓度与发病率、住院与死亡率之间的关系进行了研究, 采用泊松回归方法分析后得出, PM_{10} 与总悬浮颗粒物(TSP)浓度与呼吸系统、心血管系统疾病的发病率联系密切。

(2)对气候影响巨大。首先, 颗粒物浓度直接影响空气能见度, 由于颗粒物对太阳光的散射和吸收作用, 导致能见度的降低。在干净的大气环境中, 能见度最高可达 30 公里, 而在大气污染严重的区域, 能见度仅能达到 5 公里左右。其次, 大气颗粒物浓度影响太阳辐射, 它将太阳辐射散射到外层空间, 减少到达地面的紫外线强度, 进而影响地表的热收支平衡, 直接参与影响全球气候。这样会影响人类骨骼的生长发育, 导致佝偻病的患者的增加, 同时还可能引起农作物的减产, 对国家社会的长远发展产生影响。再次, PM_{10} 还能形成雨滴的凝结核, 颗粒物吸附的酸性物质能溶解于水, 形成酸性降雨而污染环境。

(3)不确定性。大气颗粒物在形成、迁移过程中会吸附许多化学物质, 这些化学成分在特定条件下可能发生化学变化。比如不同颗粒物会吸附不同波长的阳光, 使以颗粒物为载体的化学成分发生光化学变化, 进而形成毒性更强的衍生物, 会对人类健康产生不确定的潜在危害。而且某些气溶胶中包含的元素碳的表面活性可以影响 SO_2 、 NO_x 、 O_3 和其他气态化合物的均相与非均相化学反应, 对对流层和低层平流层臭氧的非均相转化消耗有着重要贡献, 对全球气候变化起到重要作用, 成为气候预测中最不确定的因子之一^[7]。

基于 PM_{10} 的严重危害性, 我们应高度重视对 PM_{10} 的来源、分布污染趋势、影响因素及 PM_{10} 污染预报等方面的研究, 及时认识到大气环境中 PM_{10} 的防治是一项具有全

局意义，关乎国计民生的系统工程，应该置于战略高度。

1.1.3 我国城市环境空气质量及其主要污染物

随着我国国民经济社会的跨越式发展，工业化、城市化进程的不断推进，许多大中城市的空气质量恶化情况也日趋突出。随着居民生活水平的不断提高，人们对所居住城市的空气污染情况也是越来越关注，城市空气质量已经成为衡量居民幸福感的一个重要指标。

根据对环保部公布的中国环境状况公报^[8]与中国环境状况概况^[9]中给出的大气环境状况的分析，近年来我国开展环境空气质量监测城市的城市空气质量总体逐年好转。如图 1-1 所示，从 2003 年到 2009 年的七年间，全国各大中城市城市空气质量达到或优于二级的城市比例趋于逐年上升。至 2009 年，全国空气质量达到或优于二级标准的城市比例已经达到 82.5%，空气质量劣于三级标准的城市也已经下降至 1.3%。PM₁₀ 浓度达到或优于二级标准的城市比例已达到 86.3% (见图 1-2)，远远超过 2003 年的 45.6%，这表明近年来我国加大力度整治大气污染的行动卓有成效。

基于对环保部重点监测城市近年的空气质量数据分析可知，我国各大中城市的首要污染物仍为 PM₁₀。本论文对 80 个重点监测城市 PM₁₀、SO₂、NO₂ 作为首要污染物出现的频率做了对比，各城市污染物中 PM₁₀ 作为首要污染物出现的频率达 75% 以上，一些受 PM₁₀ 污染较严重的城市，如宝鸡、石家庄、西安、兰州、长治、长春、赤峰等地，PM₁₀ 作为首要污染物出现的频率则高达 90% 以上。据统计，近年来我国 PM₁₀ 污染较为严重的城市主要分布于山东、陕西、新疆、山西、内蒙古、湖北、江苏、甘肃、湖南等省市。

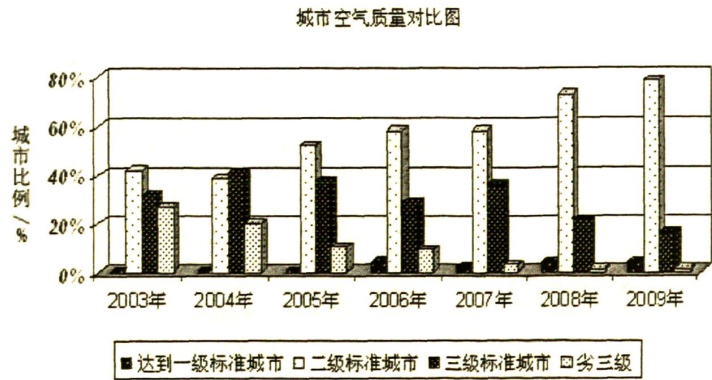


图 1-1 各年城市空气质量对比图

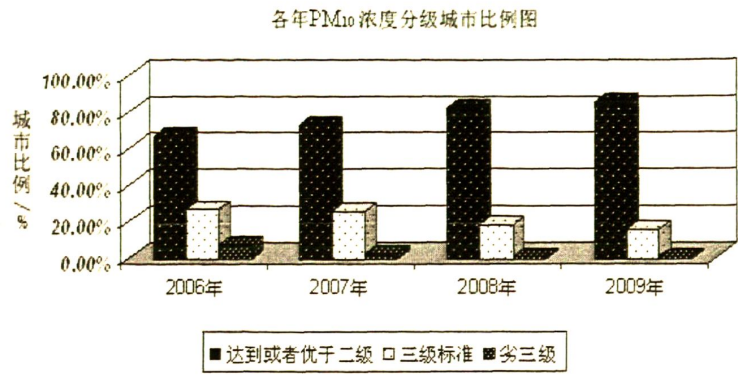


图 1-2 各年PM₁₀浓度分级城市比例图

总之，PM₁₀是造成城市大气环境污染，尤其是重度污染最主要的污染物。虽然通过努力我国大中城市空气质量趋于好转，但是部分地区、部分城市仍是PM₁₀污染的重灾区，我们对PM₁₀的防治依然任重道远。

1.2 国内外空气污染治理研究现状

1.2.1 空气污染治理及其主要方法

近三四十年来，随着公众环保意识的提高，大气环境的保护不断受到重视。为了满足民众的环保需求，同时也为了给大气污染物的综合治理和环境管理与决策提供科学依据，空气污染治理技术应运而生并迅速发展。空气污染治理是指根据某些特定方法对某个区域未来的污染浓度及空间分布作出估计，若有严重污染出现或污染浓度超过某一限值时则发出警报，供有关部门提前采取有效措施防止污染危害事件的发生的一种实用技术。按预测的模式与性能不同，可以分为以下几类^[10 11 12]：(1)大气污染潜势预测。大气污染潜势预测是以气象上的天气形势和气象因子为依据，对未来的大气环境质量状况进行定性或者半定量分析的“二次预测”。基本原理是归纳出以往发生污染事件时的天气形势与气温、气压、能见度、总云量、辐射量、风速、风向等气象因子指标，对这些数据进行总结分析以预测未来大气污染的气象条件。它是以气象因子的某一临界值作为依据的，并不考虑污染源的位置及污染物排放量，故在空气质量很好的地方可能也会产生高污染潜势，误差率较高^[13]。

(2)大气污染统计预测。大气污染经典统计预测是依据事物的规律与联系进行预测的方法，它需要一个地区长期的气象因素逐日实测资料与这个地区的污染物逐日实测浓度资料，分析各种气象因子，将气象因子的参数与相应的污染物浓度参数建立起各种定量或者半定量的关系。如采取多元回归、逐步回归等方法，最后建立出此地区污

染物预报的数学模型。这种方法的优点是简单、易普及。缺点是监测数据需长期、准确、大量，较难获取。

(3) 大气污染相似预报。现实操作中，由于有时难以获取长期、大量的空气污染浓度资料和气象监测资料，或是由于污染源的变化较大，以往的历史资料参考价值不大，只能选用近期的监测数据。这时，大气污染相似预报方法就较统计预报方法表现出了更好的实用性。这种预报方法是将预报日前一日的气象数据资料汇总，在历史数据库中寻找相似数据，找出相似日后，以相似日后一日的大气污染浓度值作为预报值。如将预报日前日的多个气象因子汇总，在历史数据库中逐日计算这些气象因子数据与历史数据的相关系数，相关系数最大即为相似日，然后以相似日次日浓度数据为预报值。或如香港环境保护署的做法，在做相似性分析的过程中，把先前的污染物浓度与相关的气象因子作为一个矢量单元，如下所示：矢量 $X=[\text{污染物浓度}, (\text{各相关气象因子 } x_1, x_2, x_3, \dots, x_i)]$ ，当日矢量与数据库中数据的逐日矢量差用欧几里得距离计算有，距离

$(x, y) = \sqrt{\sum (x_i - y_i)^2}$ ，计算出的与当天矢量差最小日即为相似日，然后以相似日次日浓度为预报值。这种方法的优点是方便快捷，缺点是防突变能力差，难以反映突变情况。

(4) 大气污染数值预测。大气污染数值预测方法是在大气化学、大气物理学、大气动力学等学科基础上使用数值计算方法直接在环境介质中求解污染物浓度的方法。它可以直接定量的描述预测污染物的浓度，但是需要了解污染物在大气中所经历的物理、化学、生物变化即污染物在迁移过程中的演变规律。常用的模模型有拉式(Lagrange)模型、欧式(Euler)模式或者混合模型。能定量预报大气污染物浓度的模型有烟羽(Plume)模型、烟团(Puf)模型、箱式(Box)模式及求解物质守恒方程的数值模型等^[14]。这种方法的优点是科学性极强，缺点是难度大，费时。

(5) 扩散模式预报。这种方法是根据污染物的排放量及气象参数，依物理原理和实验得出的经验参数，设计出数学模型，进行空气污染分布的预报。这种方法适用单个比较大的污染源，或者高度低、面积较小，不适合统计方法处理的面源。我国在进行原子弹爆炸实验时，进行放射性污染预报时，就是使用扩散模式。

(6) 人工神经网络预报。由于大气污染预报具有突发性、非线性等特点，一般的预测方法，如统计预报、潜势预报等都是基于气象数据资料并且一般情况下对这些数据都是做线性处理，故对特殊气象条件下的大气污染物浓度变化并不能很好的反映。而人工神经网络有着很好的处理非线性问题的能力和学习记忆能力，因此非常适合用于

空气质量预报。目前, 人工神经网络在空气污染预报领域主要有以下几个研究方向: 其一是对污染物浓度的预报, 其二是对空气污染指数的预报, 其三是对人工神经网络预报建模的优化改进, 其四是关于人工神经网络与其他预报方法的预报效果对比分析, 其五是人工神经网络与其他预报方法的集成预报的研究。

1.2.2 国外空气污染预报的发展与现状

国外对空气污染预报的研究开展较早, 美国早在 70 年代初就开始了大气污染潜势预报^[15], 现在其大气污染预报模式已从第一代的拉格朗日模式进化为第二代的欧拉模式, 现在又开发出第三代模式系统, 命名为 Model-3 CMAQ 模式。美国国家气象局 (NWS)^[16]根据天气预报的风、天气情况、大气稳定度等气象要素, 用箱式模型和污染源源强来对未来 24 小时进行扩散计算, 以对空气污染情况进行预报。Yilamaz Yildirim 等^[17]对土耳其 Erzurum 市市中心的上空空气建立了一个基于连续性方程的简单非线性模型, 对该市污染物浓度进行预报, 预测结果与实测值吻合度较高。日本的 Ota M^[18 19]利用广岛 1974 年的污染物浓度资料与同期气象资料, 建立了氧化物浓度与气象因子之间的统计学关系, 并结合每日的天气预报, 对广岛的氧化物浓度进行预报。Jorquera 等^[20]将线性的时间序列、模糊理论和人工神经网络建模结合起来预测圣地亚哥每日臭氧水平的最大限度, 取得了很好的效果。Petros Koutrakis 等^[21]利用智利圣地亚哥市 1989 至 2001 年监测的 PM_{2.5}、PM₁₀、PM_{2.5-10} 浓度数据与同期的气象数据资料, 建立了颗粒物浓度的回归模型, 以此对圣地亚哥市的可吸入颗粒物浓度进行了预测。荷兰的 Var Der Wal^[22]利用时间序列分析法和主成分分析法研究了 PM₁₀ 的日变化趋势并对 PM₁₀ 的浓度做了预测。Gong Yuh Lin^[23]对污染物浓度与气象因子之间进行了相关分析并研究了污染发生过程的统计学关系, 由此来预报加利福尼亚南海岸地区的大气污染过程, 精确度令人满意。Martin Branis^[24]等人对捷克中部一个小村镇大气环境中的 PM₁₀ 浓度进行了为期一年的监测, 得到 PM₁₀ 的浓度和黑碳 (BS) 的浓度在采暖期有着强相关性, 而在非采暖期两者的关系则不明显, 且 PM₁₀ 的浓度无论在采暖期还是非采暖期都与最低气温有着强相关性。

自上世纪 90 年代起, 人工神经网络应用于大气污染预测的研究开始逐步深入, 目前已有较大的发展。Boznar M^[25]等人最早将人工神经网络应用于大气污染预报领域, 他们采用了一个 MLP (Multi-Layer Perception, 即多层感知器) 模型, 对位于斯洛文尼亚绍什塔尼的全国最大的热电厂附近大气环境中的 SO₂ 浓度进行了短期预测。Yi.J^[26]等采

用 MLP 模型预测北美工业区的 O_3 浓度, 模型的输入层包括风速、温度、压强等 9 个因子, 得到的预测结果优于同输入因子下的回归模型和 ARIMA 时间序列模型。Perez^[27]等人将 MLP 模型、回归模型、持续性模型应用于智利圣地亚哥固定地点 $PM_{2.5}$ 浓度预测, 比对后得出应用神经网络的预测效果最佳。P.viotti^[28]等人用 BP 模型对 NO_x 、 O_3 、CO 的浓度进行了预测, 预测结果与实测值拟合度很高。Sang^[29]等应用 BP 神经网络预报汉城大气环境中的 O_3 浓度, 以 8 个输入因子的时间序列总计 30 个变量为输入层, 以为来时间序列的 O_3 浓度为输出层, 设一个隐含层, 对该地区 O_3 浓度的时间及空间上的变化趋势做出了准确的预报。

1.2.3 国内空气污染预报的发展与现状

我国开展对大气污染预测的研究较晚, 但发展迅速。1997 年中国气象科学院徐大海^[30]等利用箱式预报模型并引入空气污染潜势指数, 率先开展了城市大气污染的潜势预报。王迎春^[31]、王川等^[32]、杨民等^[33]、刘实等^[34]分别建立了北京、西安、兰州、长春等地的大气污染潜势预报的方法。在大气污染数值预报方面, 雷孝恩^[35]于 1998 年研究推广了大气污染数值预报模式系统。中国气象科学研究院朱蓉, 徐大海^[36 37 38]等人研究开发了 CAPPS 模式系统, 该系统对北京和上海的大气污染进行预测, 预测结果准确率平均可达 60%。韩志伟等^[39]、董亚龙等^[40]分别应用此系统对济南市、铜川市区的大气污染物浓度进行了预报, 预报结果令人满意。在大气污染统计预报方面, 王淑英^[41]对北京市的 PM_{10} 污染物与同期气象条件进行了相关性分析, 并且建立了多元回归分析模型对 PM_{10} 污染物浓度进行了预报。黄磊, 王赐震等^[42]对青岛市污染监测资料进行分析, 采用了时间序列分析方法, 建立了多种预报模型, 最后提出了一种综合模型。柴微涛, 宋述君等^[43]对成都市城区的空气污染指数进行了时间序列分析, 并建立了相应的预报模型, 得出的模拟值与实测值之间拟合效果良好。郭尚媛^[44]对西安采暖季的 PM_{10} 浓度与大量气象因子进行了相关性分析, 通过主成分分析法选出数个预报因子, 建立了多元线性回归模型进行污染物浓度预测, 得出的平均准确率可达 79.6%。周变红^[45]建立了基于因子分析法的宝鸡市 PM_{10} 浓度的四季预报模型, 四季平均准确率达到 78%。

我国关于人工神经网络应用于大气污染预测方面的研究较晚。1997 年, 李祚泳^[46]应用基于 BP 算法的人工神经网络, 提出了建立大气污染浓度预测的方法, 并且通过对某市 SO_2 浓度的神经网络预测实例, 证实了人工神经网络应用于空气污染预测的可行性。王俭等^[47]利用 1999 年沈阳市秋季 SO_2 的数据, 根据大气污染物浓度与气象因子之

间的关系,建立了 BP 神经网络模型进行污染物的预测,得到的平均相对误差仅为 10.95%。后来又应用遗传算法对 BP 模型进行了改进,使 BP 神经网络的预测精度得到了提高^[48]。王灿星等^[49]对 1999 年至 2001 年杭州市区大气中的颗粒物实测数值进行分析,构建并选用了合适的 BP 神经网络进行训练,结果表明 BP 神经网络应用于大气污染的预测是可行的。邓伟妮^[50]建立起了西安市 PM_{10} 污染物浓度的 BP 神经网络预测模型,并对动量 BP 算法、BFGS 拟牛顿法、SCG 算法等六种 BP 算法的训练效果进行了对比分析,确定了最佳算法并得出预报的吻合率为 80%。王兰霞^[51]对西安市采暖季和非采暖季分别建立了 RBF 神经网络模型,对污染物浓度进行预测,得到采暖季的平均准确率为 74.47%,非采暖季的平均准确率为 79.48%。王芳等^[52]为了对北京市 PM_{10} 浓度进行预测,建立了遗传神经网络模型并将该模型与美国第三代空气质量模型 CMAQ 的数值模拟结果进行了对比,表明两者预测结果精度相当。雷蕾^[53]采用伦敦市 $PM_{2.5}$ 的小时平均浓度数据,利用 BP 神经网络技术建立了预报模型,定量的预测了伦敦市 $PM_{2.5}$ 的小时平均浓度。

1.3 论文的研究目的与意义

我国现阶段处于经济社会发展的高速时期,发展与环境保护之间的博弈日趋激烈。目前我国还存在着能源结构不合理,环保设备技术落后,民众环保意识淡薄等严重问题,部分地方政府还只重视经济增长这个“硬抓手”,而忽略环保这个“软指标”。这些因素综合起来导致了我国环保事业的发展缓慢,部分城市大气污染严重。在这个关键时期,如何保护好我们赖以生存的大气环境,使人民能在享受改革开放成果的同时也享受到良好的空气质量,检验着我们党和政府的执政能力和行政伦理。在这种形势下,开展对大气污染预测的研究尤其是开展对多数城市首要污染物 PM_{10} 的预测研究意义深远而重大。主要表现在以下几个方面:(1)关系到民众的切身利益,会对民众的生活、出行产生直接影响。(2)为环境管理部门的决策提供有效的支持,能使环境管理部门对污染迅速做出反应,并采取防止和消减措施。同时能使环境管理部门对城市环境规划要素(功能分区、污染源分布、城市规模、交通网络)提出更科学、更明确的控制和管理方案。(3)关系到“十二五”期间科学发展、可持续发展理念的践行。“十二五”期间,政府要继续加大对环境保护的投入力度,加大对生态文明建设的投入力度。同时,大气污染预测的开展也表明了我们政府敢于面对环境现实的信心,表明了我们政府治理大气污染的决心,更表明了我们政府为人民群众撑起一片蓝天的诚心。

本论文研究的城市大庆市位于黑龙江省西部，是举世闻名的石油城市。大庆市始终将环境保护工作纳入城市经济社会发展的总体规划，不断加大对环境保护的投资力度，不断加强城市环境的综合整治，环境质量趋于稳定。先后被命名为“国家卫生城市”，内陆第一家“国家环保模范城市”，这对于一座典型的资源能源型城市来说实属不易。大庆市区的总体大气环境质量多年来始终保持在国家环境空气质量二级水平，基本上每日的首要污染物都是 PM_{10} 。基于此，大庆市政府近年来实施了以低空大气污染专项治理为重点的“蓝天工程”，大庆市的大气质量又上了一个新的台阶，2008 年，全市全年 API 均值为 51，城市环境空气优良天数为 356 天，占全年天数的 97.3%。2009 年，全市全年 API 值为 49，城市环境空气优良天数为 361 天，占全年天数的 98.9%，在全国大气质量重点监测城市中名列前茅。但是，随着经济社会的不断发展，大庆市的环保形势依然严峻，存在以下几方面的问题^[54]：(1) 油田的特征污染物总烃排放量大，对大气环境质量影响较大并形成一定程度的热岛效应。(2) 机动车数量增长迅速，使汽车尾气成为新的大气污染源。(3) 城市化及人口增加导致的城区供热、供暖小煤炉、油炉的增加，油电改煤工程的开展，使城市 PM_{10} 的污染问题日趋突出(尤其在采暖期)。(4) 大庆市周围分布大量的风沙土和盐碱地，植被覆盖少，生态环境脆弱，导致风沙和二次扬尘形成的降尘和粉尘成为影响大庆市空气环境质量的不确定因素。以上问题的都会导致大庆市 PM_{10} 的污染加剧，由此可见，大庆市 PM_{10} 污染的防治任重而道远，故及时尽早的开展大庆市大气中 PM_{10} 的预测研究意义重大。

本论文旨在对大庆市 PM_{10} 的污染趋势及气象因子进行探讨，并对近年的 PM_{10} 浓度监测数据与同期的气象因子数据进行比对分析，探究两者之间的相关性，据此构建出大庆市 PM_{10} 污染预测的多种模型并对模型进行比对以找出最佳模型。以期为大庆市制定合理的 PM_{10} 防治规划提供科学依据，并为产业转型期大庆的大气环境、生态环境保护提供参考。

1.4 论文主要内容及研究方法

本论文综合运用了环境学、气象污染学、数学、统计学、计算机科学等学科原理，重点分析了大庆市环境监测中心站提供的 5 个监测点近年的 PM_{10} 空气污染指数日报以及同期的气象因子监测资料。以 2008 年 1 月至 2010 年 6 月的 PM_{10} 空气污染指数为数据基础分析了大庆市 PM_{10} 的污染特征及污染趋势，以 2006 年 1 月 1 日至 2010 年 6 月 30 日的气象因子数据为基础分析了大庆市历年气象要素的基本情况及变化趋势，对

2008 年 1 月 1 日至 2010 年 6 月 30 日的 PM_{10} 污染指数数据资料及同期气象因子数据进行相关分析, 建立起大庆市的 PM_{10} 污染预报模型体系。本论文主要包括以下几个部分:

(1) 大庆市 PM_{10} 污染特征与污染趋势的分析。

通过对历年 PM_{10} 的污染资料进行汇总分析, 采用概率统计方法研究大庆市 PM_{10} 污染的年、季节、月的污染变化特征及趋势, 并重点对采暖期与非采暖期进行比对分析, 对沙尘暴的影响、过年期间烟花爆竹的影响进行单独分析。

(2) 大庆市气象要素特征分析及气象要素与 PM_{10} 污染的相关性分析。

对大庆市近年的气象要素数据进行分析, 探讨气象因子与 PM_{10} 污染之间的关系, 并使用 SPSS 软件对 PM_{10} 污染指数与气象因子之间做相关性分析, 找出对大庆市 PM_{10} 污染影响较大的气象因子并汇出相关趋势图。

(3) 大庆市 PM_{10} 污染预测模型的建立。

分别采用主成分分析法、逐步回归法、B-P 人工神经网络等方法建立起大庆市 PM_{10} 污染预报的数学模型和神经网络模型, 对两模型进行检验分析并通过模型预测值与 API 实测值的对比确定其准确性, 以期找出最适合大庆市的 PM_{10} 污染预报模型。

第二章 基本理论

2.1 人工神经网络基本原理

2.1.1 人工神经网络概述

人工神经网络(Artificial Neural Networks,ANNs),也称神经网络。是一个由人工构造的模拟生物神经网络的抽象数学模型。它由大量的神经元节点广泛连接组成,用于模拟大脑的某些机理与机制。它能通过对大量样本的训练学习,获得这些数据内在的规律,从而实现复杂逻辑操作和非线性信息处理的目的^[55]。

人工神经网络的特点有:可以处理并行分布问题和非线性问题;具有自学习的能力;拥有联系想象的功能。因为其特殊的特点,它已经成为突破科学技术瓶颈,深入探索非线性复杂现象的有力工具,广泛应用于各个科学领域。人工神经网络在环境领域应用极为广泛,近年来国内外学者在这方面也有较多研究。如对城市综合环境质量评价的研究^[56 57],对地表水水质、地下水水质、大气质量等方面评价的研究^[58 59 60 61],对生态环境质量的评价的研究^[62 63]及对环境质量预测^[64]方面的研究等,也取得了大量的研究成果。

对神经网络的研究起源于20世纪40年代,至现在已经历以下三个阶段:

(1)20世纪50~60年代。这段时期为神经网络研究的第一次高潮。学者 McCulloch 和 Pitts 于1943年共同提出了神经元的数学模型,即著名的MP模型,这一研究成果打开了神经网络时代的大门。之后的接近20年里,Hebb、Rosenblatt、Widrow 等人先后提出了Hebb规则,感知器模型,自适应线性元件等模型,因而形成了神经网络、脑模型研究的第一次高潮。

(2)20世纪60~70年代。这段时期为神经网络研究的过渡时期。这期间内,人们研究发现了感知器模型存在的一些不足,如不能解决异或问题,这导致研究工作趋向低潮,但科学家们仍然进行了大量研究。如Kohonen提出的自组织映射,Fukushima提出的神经认知机网络理论,Webos提出的BP理论,这些理论为神经网络的第二次研究热潮做了铺垫。

(3)20世纪80年代至今。这段时期为神经网络的第二次研究高潮。美国的学者Hopfield在80年代初期,提出了以自己名字命名的Hopfield神经网络模型,他引入了能量函数的概念,给出了网络的稳定性判断,有力的推进了神经网络的研究。1984年,

Hinton 等人又提出了 Boltzman 模型，该模型借用了统计物理学的概念和方法，提出了多层网络的学习算法。两年之后，McCulland 和 Rumelhart 等人提出了 BP 算法，这是一种由多层网络构成的误差反向传播算法，这种算法能够不断地学习样本，误差，把得到的结果再反馈到隐含层的隐单元，通过不断改变权系数矩阵来达到期望学习的目的。这也是迄今为止最常用、最普通的网络。Powell M.J.D 于 1985 年提出了径向基函数(RBF)，1988 年 Broomhaced 将 RBF 网络和 BP 网络进行了比对，揭示了二者关系。迄今为止，研究者们已经提出了多种 RBF 网络的学习算法。RBF 神经网络的优良特性，使其成为 BP 神经网络之后又一种应用广泛的神经网络。RBF 网络是一种局部逼近网络，即对输入空间的某一个局部区域只存在少数的神经元用于决定网络的输出，BP 网络则是典型的全局逼近网络，即对每一个输入输出数据对，网络的所有参数都要调整。由于二者的构造本质不同，径向基函数网络相比在函数逼近能力、分类能力和学习速度等方面均由于 BP 神经网络，因而将它应用于复杂的储层参数会取得很好的效果^[65]。

2. 1. 2 人工神经元模型

人工神经元是人工神经网络的基本信息处理单元，其模型如下图 2-1 所示。

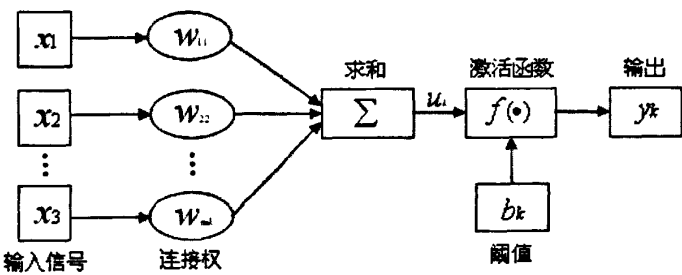


图 2-1 人工神经元结构图

从图中可以看出，一个人工神经元模型应由三部分组成：

- (1) 一组突触或连接。用权值表示各人工神经元之间的连接强度，其值可取正值可取负值，正值代表激活神经元，负值代表抑制神经元。
- (2) 一个信号累加器。作用是求输入信号对各神经元的相应突触加权求和。
- (3) 激活函数。用于将输出信号限制到允许范围内的一定值，从而限制神经元的输出。一般情况下，一个神经元的输出扩充范围可写成[0,1]或[-1,+1]闭区间。

除此之外，还可以给上述神经元模型加一个外部偏置 b_k ，其作用是可根据神经元的正负来增加或者降低激活函数的网络输入。故一个人工神经元模型可抽象为以下数学

模型:

$$U_k = \sum_{i=1}^m w_{ik} x_i \quad (2-1)$$

$$y_k = f(u_k + b_k) \quad (2-2)$$

式中 x_i 表示输入信号, w_{ik} 为神经元的突触权值, u_k 为输入信号线性组合器的输出, b_k 为外部偏置, $f(\cdot)$ 为激活函数, y_k 为神经元输出信号。

其中激活函数是一个神经元的重要组成部分, 一个网络解决问题的能力 and 功效除了与神经网络的结果有关, 很大程度上还由激活函数决定。其主要有以下三种形式:

(1) 阈值函数。也称作阶梯函数, 当函数的自变量 < 0 时, 函数的输出 $= 0$; 当函数的自变量 ≥ 0 时, 函数的输出值 $= 1$ 。其数学模型如下:

$$y = f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2-3)$$

(2) 线性型函数。如 $y = f(x) = kx$ (2-4)

(3) 非线性转移函数。该函数为实数域 R 到 $[0, 1]$ 的非连续函数。常见的有 Sigmoid 函数和双曲正切函数。其表达式分别如下:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (\text{单极性 Sigmoid 函数表达式}) \quad (2-5)$$

$$f(x) = \frac{2}{1 + e^{-x}} - 1 = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (\text{双曲正切表达式}) \quad (2-6)$$

2.1.3 神经网络的结构及工作方式

神经网络是以网络拓扑结构的方式, 把大量人工神经元联系起来, 构成群体并行式的计算结构。将一个神经元的输出送到另一个神经元作为输入信号称之为连接, 每个连接通路对应一个连接权系数, 相同神经元经过不同的连接方式得到不同特性的网络。按神经元联系方式的不同, 可将神经网络分为两大类: 即分层网络和连接型网络^[66]。

2.1.3.1 分层网络

将一个神经网络模型中的所有神经元按照功能分成若干层的网络成为分层网络。一般分为输入层、隐含层(中间层)和输出层。输入层是连接外部输入模式, 并由各输入单元传送给相连隐含层的各单元。隐含层是神经网络的内部处理单元。神经网络所

具有的模式变换能力(如模式分类、模式完善、特征抽取等)主要体现在隐含层单元的处理功能上。根据模式变换功能的不同,隐含层可以一层没有,也可以有多层。输出层是神经网络的输出模式层。

分层网络可以细分为三种互连方式:(1)单纯的前向网络。这种网络状态更新的流程是,先由输入层为网络导入输入模式,然后通过中间层依次变换模式,最终产生输出模式在输出层中。(2)具有反馈的前向网络。反馈的结构形式封闭环路,具有反馈的单元也称隐单元,其输出称为内部输出,而网络本身还是前馈型的。(3)层内互连的前向网络。同一层内单元的相互连接使它们彼此之间制约,限制同一层内能同时激活的单元个数,而从外部看来还是前向网络。一些自组织竞争网络就采用这种拓扑结构。

2.1.3.2 相互连接型网络

这种网络中,随机的两个单元之间都是有路径来联系的。这种互相连接的方式包括局部的连接,和整体连接两种类型。在局部连接的网络中,某些神经元之间是没有关系的。而在整体连接的网络中,所有的神经元输出,都有另外的神经元能和它联系。

在分层网络中的前向网络中,如果输入某一模式,网络就能快速产生相应的输出模式,并且稳定。但在互连式的网络中,输入某一指定的输入模式,结果不一定稳定。网络会处在一种一直变化输出的模式的情况中,有可能依旧产生相应的稳定的输出,可也可能进入一种周期性的振荡,或是进入混沌。

2.1.4 神经网络的学习

人工神经网络的学习(训练)过程是指神经网络因受到外界刺激而不断调整网络的内部自由参数,从而使自身以一种新的方式对外部环境作出反应的一个过程。能够从外部环境中学习并且在学习过程中不断提高自身性能是神经网络最有意义的性质,神经网络通过反复学习来达到对环境的了解。神经网络的学习方式可以分为以下三种。

(1)有导师学习。顾名思义,这种学习方式要先有模板来做“老师”。这里的“老师”就是一个含有输入值和输出值的样本数据集。先将输入样本数据导入到网络的输入层,把经过网络运行得到的实际输出与期望输出相比较得到误差,然后根据所得误差的实际情况来不断调整网络参数,使网络输出不断逼近期望的响应,直到实际输出与期望输出之差在允许的误差范围内。通常有这种学习方式的算法采用的是梯度下降法,但其容易存在局部极小值问题。故实际操作中一般需要特定方法对其进行优化,通常通过随机学习算法引入不稳定因子来解决这类问题,比较著名的随机学习算法有

模拟退火算法和遗传算法。

(2)无导师学习。和有师学习相比，这种学习方法没有输出样本数据集，只有输入样本集。在初始网络中，将输入样本数据集一次次地导入到网络中，以此方式来对神经网络进行拥有“记忆”，当之后，遇到曾经拥有的“记忆”时，连接权就会以某一个系数来增大，这样的过程一直循环，知道这种“记忆”能够使得相应的连接权增大到接近 1。这一循环的过程，使神经网络拥有了某种记忆，以至在今后形成了“条件反射”，每当相似的“记忆”重现的时候，便会产生相应的输出结构。典型的无师学习有 Hebb 算法和竞争学习算法。

(3)再励学习。也称强化学习，这种学习方法介于上述两者之间，外部环境对系统输出结果只给出评价而不是给出正确答案，学习系统通过强化那些受奖励的动作来改善自身性能。

2.2 B-P 神经网络

B-P 神经网络是一种多层前馈型神经网络，由于其可以实现从输入到输出的任意非线性映射并且具有良好的自组织自适应性和容错性，故自从 1986 年 Rumelhart 等人提出 B-P 算法以来，它就得到了广泛的应用。目前在人工神经网络的应用中，80%~90% 都采用 B-P 网络及其变化形式，它也体现了人工神经网络中最精华的部分^[67 68]。

2.2.1 B-P 神经网络模型结构

B-P 神经网络不仅包括输入层、输出层，还包括一层或者多层的隐含层，处于同层的各节点之间无连接，其神经元之间的传递一般为 S 型函数，输出量为 0 到 1 之间的连续量。B-P 网络的模型结构如图 2-2 所示。

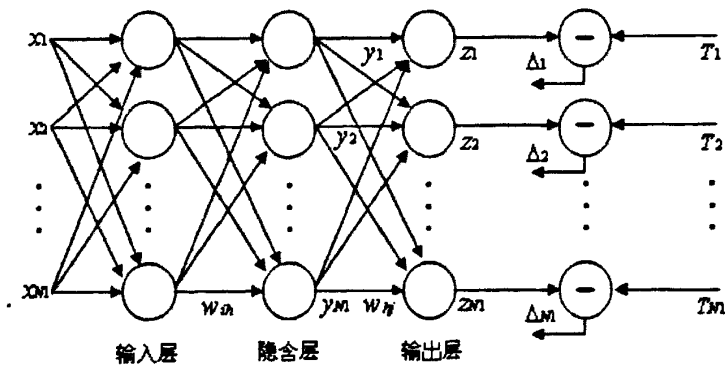


图 2-2 B-P 神经网络模型结构图

图中 B-P 网络的输入样本 $X = (x_1, x_2, \dots, x_{N1})$ 从输入层依次经过各隐含层节点, $Y = (y_1, y_2, \dots, y_{N1})$ 表示隐含层神经元的输出值, $Z = (z_1, z_2, \dots, z_{N1})$ 表示输出层神经元的输出值, w 代表神经元之间的连接权值, 箭头方向表示信息流动的方向。

B-P 网络按照有导师学习方式进行学习, 当一对导师样本 X 输入网络后, 其神经元的激活值将从各输入层经隐含层传向输出层, 在输出层输出相应的响应 Z 。而期望输出值为 T , 于是, 网络就按照减少 T 与 Z 之间误差 Δ 的原则, 从输出层经隐含层回输入层逐层修正连接权值 w , 故称之为误差反算。随着这种修正不断进行, 网络对输入样本的响应的正确率也不断提高, 最后达到允许的误差范围内。

2.2.2 B-P 神经网络的构建与算法

标准的 B-P 算法是基于梯度下降法来学习的, 他是通过调整权值和阈值, 使实际输出值与期望输出值的均方误差趋于最小而实现的。其标准算法的实现步骤如下:

为了描述方便, 先定义下列变量和向量: 输入向量 $X = (x_1, x_2, \dots, x_N)$; 隐含层的输入向量 $h_i = (hi_1, hi_2, \dots, hi_p)$; 隐含层的输出向量 $h_o = (ho_1, ho_2, \dots, ho_p)$; 输出层的输入向量 $y_i = (yi_1, yi_2, \dots, yi_q)$; 输出层的输出向量 $y_o = (yo_1, yo_2, \dots, yo_q)$; 期望的输出向量 $d = (d_1, d_2, \dots, d_q)$; 输入层与中间层的连接权值 w_{ih} ; 隐含层与输出层的连接权值 w_{ho} ; 隐含层各神经元的阈值 b_h ; 输出层各神经元的阈值 b_o ; 样本数据个数 $k = 1, 2, \dots, m$; 激活函数 $f(\cdot)$ 。

具体步骤: (1) 网络初始化。给 w_{ih} , w_{ho} , b_h 和 b_o 分别赋一个在 $(-1, 1)$ 之间的随机数, 设定误差函数 $e = \frac{1}{2} \sum_{o=1}^q (d_o(k) - y_o(k))^2$ (2-7), 给定计算精度值 ε 和最大学习次数 M 。

(2) 随机选取第 k 个输入样本 $x(k) = (x_1(k), x_2(k), \dots, x_n(k))$ 及对应的期望输出 $d(k) = (d_1(k), d_2(k), \dots, d_q(k))$ 。

(3) 计算隐含层各神经元的输入 $hi_h(k)$, 然后用 $hi_h(k)$ 和激活函数计算隐含层各神经元的输出 $ho_h(k)$:

$$hi_h(k) = \sum_i^n w_{ih} x_i(k) - b_h \quad h = 1, 2, \dots, p \quad (2-8)$$

$$h_{oh}(k) = f(h_{ih}(k)) \quad h = 1, 2, \dots, p \quad (2-9)$$

$$y_{io}(k) = \sum_h^p w_{ho} h_{oh}(k) - b_o \quad o = 1, 2, \dots, q \quad (2-10)$$

$$y_{oo}(k) = f(y_{io}(k)) \quad o = 1, 2, \dots, q \quad (2-11)$$

(4) 利用网络期望输出向量 $d(k) = (d_1(k), d_2(k), \dots, d_q(k))$ ，网络的实际输出 $y_{oo}(k)$ ，计算误差函数对输出层的各神经元的偏导数 $\delta_o(k)$ ：

$$\delta_o(k) = (d_o(k) - y_{oo}(k)) y_{oo}(k) (1 - y_{oo}(k)) \quad o = 1, 2, \dots, q \quad (2-12)$$

(5) 利用隐含层到输出层的连接权值 $w_{ho}(k)$ 、输出层的 $\delta_o(k)$ 和隐含层输出 $h_{oh}(k)$ 计算误差函数对隐含层各神经元的偏导数 $\delta_h(k)$ ：

$$\delta_h(k) = \left(\sum_{o=1}^q \delta_o(k) w_{ho} \right) h_{oh}(k) (1 - h_{oh}(k)) \quad (2-13)$$

(6) 利用输出层各神经元的 $\delta_o(k)$ 和隐含层各神经元的输出 $h_{oh}(k)$ 来修正连接权值 $w_{ho}(k)$ 和阈值 $b_o(k)$ ：

$$w_{ho}^{N+1}(k) = w_{ho}^N(k) + \eta \delta_o(k) h_{oh}(k) \quad (2-14)$$

$$b_o^{N+1}(k) = b_o^N(k) + \eta \delta_o(k) \quad (2-15)$$

式中 N —调整前； $N+1$ —调整后； η —学习率，在 $(0,1)$ 之间取值。

(7) 使用隐含层各神经元的 $\delta_h(k)$ 和输入层各神经元的输入 $x_i(k)$ 修正连接权和阈值：

$$w_{ih}^{N+1}(k) = w_{ih}^N(k) + \eta \delta_h(k) x_i(k) \quad (2-16)$$

$$b_h^{N+1}(k) = b_h^N(k) + \eta \delta_h(k) \quad (2-17)$$

(8) 计算全局误差 E

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_{oo}(k))^2 \quad (2-18)$$

(9) 判断网络误差是否满足要求 $E < \varepsilon$ 或学习次数大于设定的最大次数 M ，则结束

算法。否则，随机选取下一个学习样本及对应的期望输出，并且返回到第(3)步，进入下一轮的学习过程。

实际的使用中，还应该对已经训练好的网络进行测试，即用一组与训练样本不完全相同的测试样本数据输入到已经训练好的网络中，计算其得到结果是否在规定的精度范围内。

2.2.3 B-P 神经网络的局限性及优化

虽然 B-P 算法应用广泛，但是它还是存在着以下几点局限性：(1)收敛速度较慢(2)全局误差可能会有多个极小值点，而网络可能会收敛于其中的一个极小值点而不是全局极小值。(3)B-P 网络隐含层层数及它的隐含层节点数的选取较难确定。(4)网络的学习、记忆具有不稳定性。为了克服这些局限性，在实际应用中出现了很多 B-P 的优化算法。主要分为两大类，一种是对梯度下降法的优化，如动量 B-P 算法、弹性算法等。另一种是基于对数值理论优化的训练算法，如拟牛顿法、SCG 算法、L-M 算法等，其中拟牛顿法和 L-M 算法都不用计算二阶偏导及其逆矩阵，能大大简化计算量，故在实际应用中被广泛使用^[69]。

2.3 B-P 神经网络的软件实现

目前学者们对神经网络的研究多采用 Mathworks 公司推出的 MATLAB 产品组中的神经网络工具箱，MATLAB 语言简单快捷，编程效率高，在建模方面功能强大。也有学者采用 C、JAVA、FORTRAN 等语言编程实现。近年来，各软件公司纷纷推出各种功能强大的可视化的神经网络建模应用开发平台。这些平台扩展性强并且有着可视化的人机交互界面，可以有效的提高建模的效率。本文采用太普软件公司研发的太普数据挖掘套件(TIP DM Suite,TipDM)来构建用于大庆市 PM₁₀ 污染预测的 B-P 网络模型。

2.3.1 太普数据挖掘软件简介

太普数据挖掘套件是广州太普软件依托多年在人工智能领域研究的技术优势而自主研发的新一代数据挖掘平台。它使用 JAVA 语言开发，能从各种数据源获取数据，建立各种不同的数据挖掘模型(集成数十种预测算法和分析技术，亦可嵌入用户自身开发的算法)来进行数据挖掘工作。可完成包括对数据进行预处理，包括空值处理、降维处理、离散处理，因子分析、主成分分析、抽样、过滤等，创建、训练、评估模型，预测，修改模型参数，误差分析等一系列功能。

2.3.2 利用太普数据挖掘平台建模步骤

数据挖掘套件的流程如图 2-3 所示。

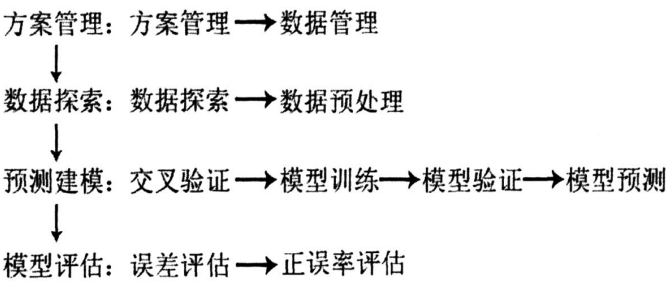


图 2-3 数据挖掘流程图

由图可以看到整个数据挖掘的流程。

第一步，进行方案管理，方案管理的界面如图 2-4。方案管理主要功能包括：打开方案、新建方案、帮助。在这里，方案创建成功后，就可以进行数据管理。通过数据管理，用户可以初步了解导入的数据体结构，如数据属性，各个属性的数据类型、数值范围等。对于已经导入的数据，也可以进行增删的操作。

在这里，系统提供两类数据加载方式：

- (1)数据文件加载，目前兼容格式包括 excel 文件和 Comma Separate Values 文件。
- (2)数据库导入，即通过后台任务方式导入第三方数据库数据到本系统中。

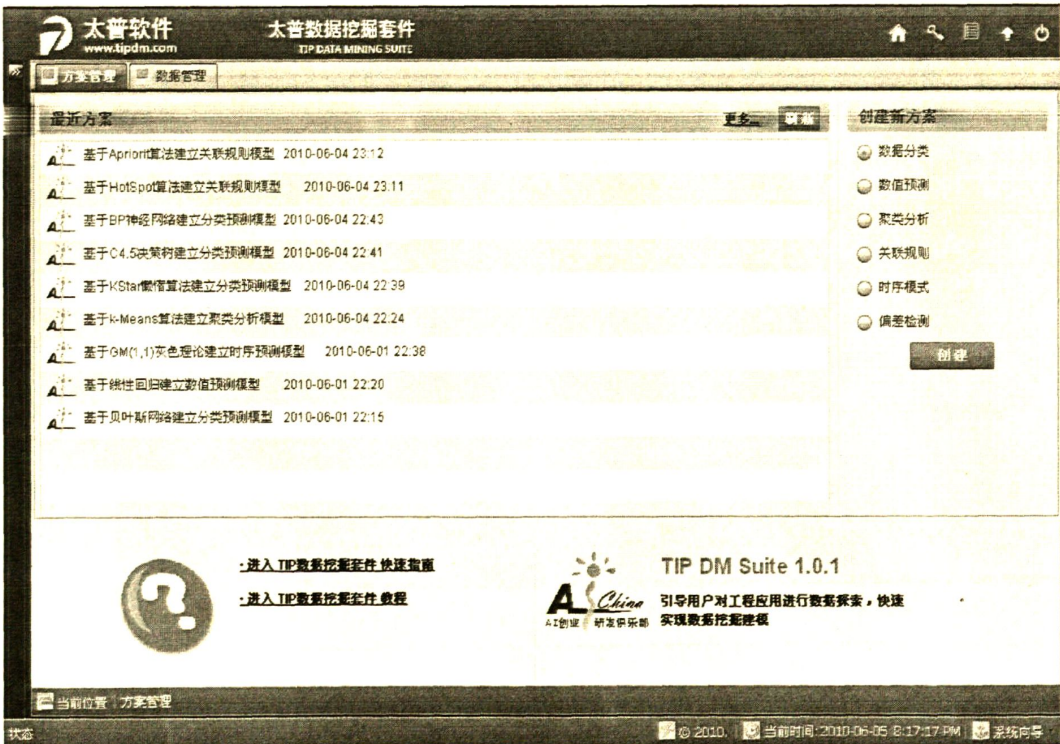


图 2-4 方案管理界面

评价预测模型的好坏，通常用相对绝对误差、平均绝对误差、根均方差、相对平方根误差等指标来衡量。如果采用了多个数据挖掘算法进行建模，则可以比较指标误差，寻求一个适合当前方案的最优算法。图 2-6 是多种算法误差综合对比分析界面，简单明了，便于比较。

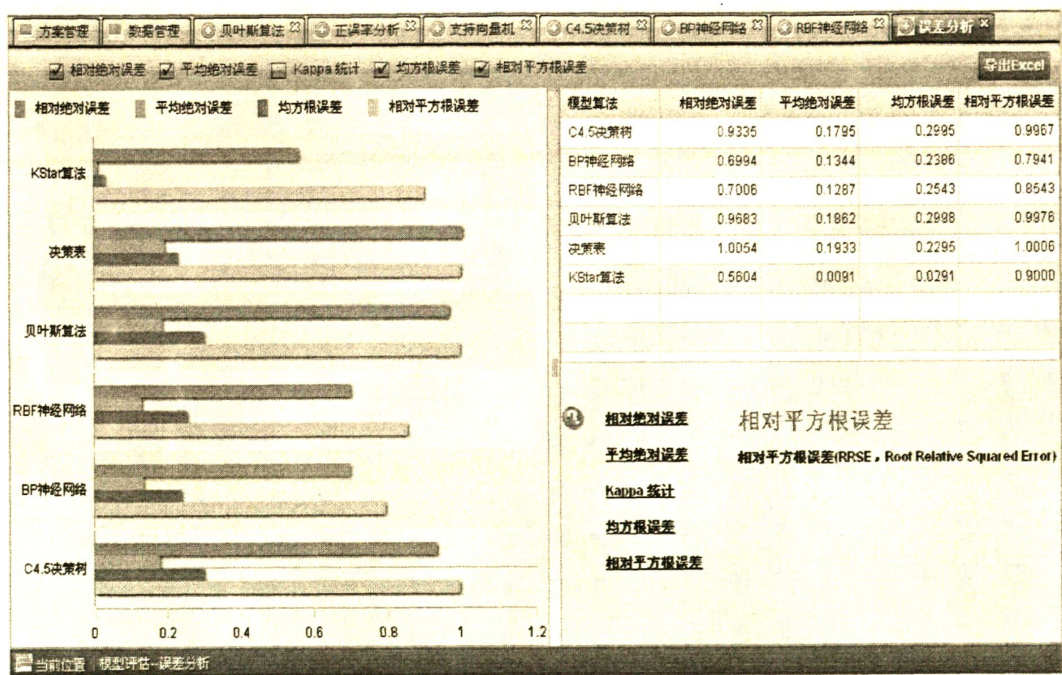


图 2-6 多种算法误差综合对比分析图

2. 4 数学建模原理简介

2. 4. 1 主成分分析方法

主成分分析又称主分量分析，它是利用降维的思想，在损失很少信息的前提下把多个指标转化为几个综合指标的多元统计方法。人们在某一事物进行研究时，为了更全面的反映出事物的特征和发展趋势，往往尽可能多的考虑与其相关的变量，这样就可能由于变量的增多而不可避免的造成共线性。而主成分分析法则可以通过对原变量相关矩阵或协方差矩阵内部结构关系的研究，利用原始变量的线性组合形成几个综合指标(主成分)，这样在保留原始变量主要信息的前提下就能起到降维简化的作用^[70]。

主成分分析的基本原理与数学模型如下：

设对某一事物的研究设计 p 个变量，分别用 $X_1, X_2, \dots, X_p$ 表示，这 p 个变量构成的 p 维随机向量为 $X = (X_1, X_2, \dots, X_p)'$ 。设随机向量 X 的均值为 μ ，协方差矩阵为 Σ 。

对 X 进行线性变换，可以形成新的综合变量，用 Y 表示，即新的综合变量可以由

$$\left\{ \begin{array}{l} Y_1 = u_{11}X_1 + u_{12}X_2 + \dots + u_{1p}X_p \\ Y_2 = u_{21}X_1 + u_{22}X_2 + \dots + u_{2p}X_p \\ \dots\dots\dots \\ Y_p = u_{p1}X_1 + u_{p2}X_2 + \dots + u_{pp}X_p \end{array} \right. \quad (2-19)$$
$$1. u_i' u_i = 1, \text{ 即 } u_{i1}^2 + u_{i2}^2 + \dots + u_{ip}^2 = 1 \quad (i=1, 2, \dots, p)$$

3. Y_1 是 $X_1, X_2, \dots, X_p$ 的一切满足原则 1 的线性组合中方差最大者; Y_2 是与 Y_1 不相关的 $X_1, X_2, \dots, X_p$ 所有线性组合中方差最大者; Y_p 是与 $Y_1, Y_2, \dots, Y_{p-1}$ 都不相关的 $X_1, X_2, \dots, X_p$ 的所有线性组合中方差最大者。

主成分分析的步骤为：(1)标准化。在实际问题中，为了减少原始数据量纲、量级对数据分析的影响，一般需对原始数据进行标准化。(2)求样本矩阵的协方差矩阵 V 或相关矩阵 R 。对标准化后的矩阵 X^* 有： $R = X^{*T} X^* / (n-1)$ (3)求特征方程的非负实根。

有 $|R - \lambda I_p| = 0$ 或 $|V - \lambda I_p| = 0$ 解得 $\lambda_1 > \lambda_2 > \lambda_3 > \dots \lambda_p \geq 0$ 。(4) 计算主成分数。计算各主成分贡献率和累计贡献率。若前几个主成分的累计贡献率达到 85%，则选择这几个主成分来代表原来的多个变量。(5) 确定主成分。第 t 个主成分 Z_t 与原始指标 $X_1, X_2, \dots, X_p$ 的关系为 $Z_t = \beta_{t1}X_1 + \beta_{t2}X_2 + \dots + \beta_{tp}X_p$ 。其中 $\beta_t = (\beta_{t1}, \beta_{t2}, \dots, \beta_{tp})^T$ 是属于特征根 λ_t 的单位特征向量， X_p 是样本数据或样本数据标准化后的指标。本文采用 SPSS 软件来进行

主成分分析过程。

2.4.2 逐步回归方法

逐步回归的原理是将所选预报因子 X 按其对所预报 Y 的相关程度的大小逐个引入回归方程，在因子引入的过程中，对相关程度不大的因子逐渐的剔除，最后得出最优的回归方程。过程如下：

(1) 求多元回归方程系数的标准方程组。

设选入多元回归方程的因变量为 $X = (X_1, X_2, \dots, X_m)$ 共 m 个， m 值并不固定，在逐步回归的过程中会逐步剔除，设回归方程为(Z_t 为标准化之后的值)：

$$Z_t = \beta_0 + \beta_1 Z_{t1} + \beta_2 Z_{t2} + \dots + \beta_m Z_{tm} + \varepsilon_t \quad (2-20)$$

假设 $t=1, 2, \dots, n$ ，变量是由 n 个数值构成的时间序列，以上模型可写为矩阵方式。

$$X = \begin{bmatrix} 1 & \frac{X_{11}-\bar{X}_1}{S_1} & \frac{X_{12}-\bar{X}_2}{S_2} & \dots & \frac{X_{1m}-\bar{X}_m}{S_m} \\ 1 & \frac{X_{21}-\bar{X}_1}{S_1} & \frac{X_{22}-\bar{X}_2}{S_2} & \dots & \frac{X_{2m}-\bar{X}_m}{S_m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \frac{X_{n1}-\bar{X}_1}{S_1} & \frac{X_{n2}-\bar{X}_2}{S_2} & \dots & \frac{X_{nm}-\bar{X}_m}{S_m} \end{bmatrix}, \quad Y = \begin{bmatrix} \frac{y_1-\bar{y}}{S_y} \\ \frac{y_2-\bar{y}}{S_y} \\ \dots \\ \frac{y_n-\bar{y}}{S_y} \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & 0 & 0 & \dots & 0 \\ 0 & r_{11} & r_{12} & \dots & r_{1m} \\ 0 & r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & r_{m1} & r_{m2} & \dots & r_{mm} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix}, \quad X'Y = \begin{bmatrix} 0 \\ r_{1y} \\ r_{2y} \\ \dots \\ r_{my} \end{bmatrix}$$

$$r_{ij} = \frac{\sum_{i=1}^n (X_{ti} - \bar{X}_i)(X_{tj} - \bar{X}_j)}{S_i S_j}, \quad r_{iy} = \frac{\sum_{i=1}^n (X_{ti} - \bar{X}_i)(y_t - \bar{y})}{S_i S_y}$$

r_{ij} 为因子 X_i 与因子 X_j 之间的相关系数， r_{iy} 为因子 X_i 与 y 之间的相关系数， R 为因子 X_i 与 X_j 之间相关系数构成的矩阵。有

$$\begin{bmatrix} 1 & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_m \end{bmatrix} = \begin{bmatrix} 0 \\ r_{1y} \\ r_{2y} \\ \dots \\ r_{my} \end{bmatrix}$$

根据正规方程组原理, 得 $\beta_0=0$, 故回归模型可写为:

$$Z_t = \beta_1 Z_{t1} + \beta_2 Z_{t2} + \dots + \beta_m Z_{tm} + \varepsilon_t \quad (2-21)$$

(2) 计算因子对方程的方差贡献大小。

计算公式: $V_k = b_k^2 / C_{kk}$ (2-22), $V_k (k=1, 2, \dots, m)$ 是 k 个变量方差贡献, b_k 是对应

因子变量的元素, C_{kk} 为逆矩阵 $(X^T X)^{-1}$ 中的元素。

(3) 变量剔除的 F 检验。

计算公式如下:

$$V_k^{(L)} = \min V_k^{(L)}, t=1, 2, \dots, L-1 \quad (2-23)$$

$$F_2 = \frac{V_k^{(L)}}{r_{yy}^{(L)}} (n-L-1) \quad (2-24)$$

服从分子分母自由度为 L 的分布, 式中 $V_k^{(L)}$ 是变量方差中最小值, $r_{yy}^{(L)}$ 是引入因子后线性回归方程的偏回归平方和, 当 $F_2 \leq F_{\alpha}(1, n-L-1)$ 时剔除因子 $V_k^{(L)}$ 所对应的变量。

2.4.3 回归方程的显著性检验

在正态线性回归的条件下, 如果所有的因变量与自变量之间不存在线性关系, 即 $b_1 = b_2 = \dots = b_m = 0$, 则统计量

$$F = \frac{\frac{u}{m}}{\frac{Q}{n-m-1}} \quad (2-25)$$

为服从自由度 $(m, n-m-1)$ 的 F 分布, 式中 n 为样本容量, m 为引进方程的因子个数, u 和 Q 分别为方程的回归平方和与残差平方和, 可由样本的实测值 y_i , 估算值 $\hat{y}_i$ 和平均值 $\bar{y}_i$ 求得。公式分别为:

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2-26)$$

$$u = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i) \quad (2-27)$$

统计量 F 可以看做是因变量方差的两个部分值之比，分子是自变量与因变量的线性联系对因变量的方差贡献，分母为其它随机因素对因变量的方差贡献。显然， F 的值越大，回归效果越显著，给定置信度 α ，假设 $b_1 = b_2 = \dots = b_m = 0$ ，查表求出满足 $P(F \geq F_\alpha) = \alpha$ 时的 F 的置信限 F_α ，如果 $F \geq F_\alpha$ ，可拒绝原假设，则认为自变量全体与因变量是呈显著线性相关的。这时， m 个自变量中至少有一个 $b_i \neq 0$ ，但是不排除其中有若干个 $b_i \neq 0$ 。又因为 $u = R^2 S_{yy}$ ， $S_{yy} = u + Q$ ，其中 R 为复相关系数， S_{yy} 为样本总方差，故有：

$$F = \frac{\frac{u}{m}}{\frac{Q}{n-m-1}} \text{ 又可以改写为 } F = \frac{\frac{R^2}{m}}{\frac{1-R^2}{n-m-1}} \quad (2-28), \text{ 由此可得复相关系数的另一表达}$$

式 $R = \sqrt{\frac{mF}{mF + n - m - 1}} \quad (2-29)$ ，对应于 F_α ，可以求得相应的复相关系数的置信限 R_α ，

求得 $R_\alpha = \sqrt{\frac{mF_\alpha}{mF_\alpha + n - m - 1}} \quad (2-30)$ ，当求得的复相关系数 $R \geq R_\alpha$ 时，则拒绝

$b_1 = b_2 = \dots = b_m = 0$ 的假设。

第三章 大庆市 PM₁₀ 污染特征分析

3.1 大庆市空气质量状况

3.1.1 数据来源及分析方法

本论文所采用的 PM₁₀ 污染数据为大庆市环境监测中心站提供的 2008 年 1 月 1 日至 2010 年 6 月 30 日共 30 个月的逐日 PM₁₀、SO₂、NO₂ 实测资料。2005 年至 2009 年 SO₂、NO₂ 空气污染指数的年均实测资料。分析 PM₁₀ 污染特征主要是依据空气污染指数(API)的分级范围及空气质量级别, 数据分析主要基于概率论与数理统计。

大庆市共布设了五个大气环境质量监测点位, 分布于五个行政区, 分别是萨尔图区环境监测站、红岗区环境监测站、让胡路区环境监测站、龙凤区环境监测站、大同区环境监测站。其中大同区环境监测点位远离城区, 是大庆市的空气环境清洁对照点。五个监测点所处的居住区、商业交通居民混合区、农业区、一般工业区均为环境空气质量功能区的二类区。各采样点的位置如图 3-1 所示:

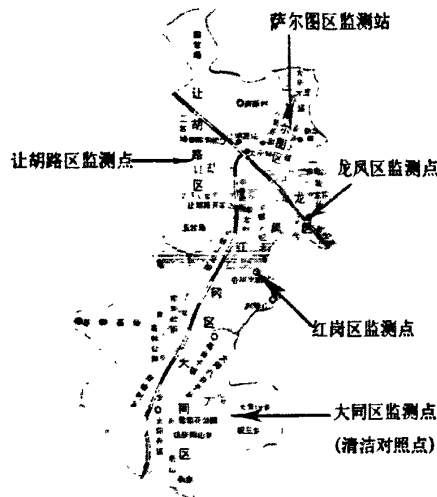


图 3-1 大庆市大气环境质量监测布点图

3.1.2 大庆市空气质量状况

评价大庆市的空气质量状况, 主要依据空气污染指数(API)分级及其浓度限值, 其具体概念及其对应的空气质量级别介绍如下:

(1) 空气污染指数分级浓度限值。空气污染指数是根据环境空气质量标准和各项污染物的生态环境效应及其对人类健康的影响, 来确定污染指数的分级数值及其相应的

污染物浓度限值。我国近年来开展了大中城市空气污染指数 (API) 日报工作，目前计入空气污染指数的污染物定为：PM₁₀、SO₂、NO₂、CO 和 O₃。空气污染指数的范围从 0 到 500，其中 50、100、200 分别对应于我国《环境空气质量标准》中的一、二、三级标准的污染物评价浓度值，500 则对应于对人体产生明显危害的污染水平^[7]。空气污染指数分级浓度限值见表 3-1，与之相应的空气质量级别以及对人体健康的影响见表 3-2：

表 3-1 国家环保部公布的空气污染指数对应的污染物浓度限值

空气污染指数	污染物浓度限值 (mg/m ³)				
API 值	PM ₁₀ (日均值)	SO ₂ (日均值)	NO ₂ (日均值)	CO (小时均值)	O ₃ (小时均值)
50	0.050	0.050	0.080	5	0.120
100	0.150	0.150	0.120	10	0.200
200	0.350	0.800	0.280	60	0.400
300	0.420	1.600	0.565	90	0.800
400	0.500	2.100	0.750	120	1.000
500	0.600	2.620	0.940	150	1.200

(2) 空气污染指数的计算方法。空气污染分指数 I_i，可由实测的浓度值按照分段线性方程计算。设 I 为某污染物的污染指数，C 为其浓度。则有：

$$I=\frac{I_{\text{大}}-I_{\text{小}}}{C_{\text{大}}-C_{\text{小}}}(C-C_{\text{小}})+I_{\text{小}}\quad (3-1)$$

上式中：C_大和 C_小：在污染物浓度限值表(表 3.1)中最接近 C 值的两个值，C_大为大于 C 的限值，C_小为小于 C 的限值。I_大和 I_小：在污染物浓度限值表中(表 3.1)最贴近 I 值的两个值，I_大为大于 I 的值，I_小为小于 I 的值。

API 的计算结果取整数，小数点后的数值需全部进位。各种污染物的污染分指数计算出来后，取最大者为该城市的空气污染指数。即 API=Max(I₁,I₂,I₃.....I_n)，该污染物即为该城市的大气首要污染物。仅当各污染物的污染分指数大于 50 时，才公布首要污染物。

(3) 空气污染指数相应的空气质量级别

通过以上对 API 的计算，得出城市大气各污染物的污染分指数，取最大值为该城市的 API。根据城市 API 的范围及与之相应的空气质量级别(表 3-2)，来判断该城市的空气质量级别，该城市大气污染对人体健康的影响以及该城市空气质量如何评价。

表 3-2 空气污染指数范围以及相应的空气质量级别

空气污染 指数 API	空气质 量级别	空气质 量描述	表征 颜色	对健康的 影响	对该空气质量的适用 范围
0~50	I	优秀	蓝色	可正常活动	自然保护区、风景名胜 区和其他需要特殊保 护的地区
51~100	II	良好	绿色		城镇规划中确定的居 住区、商业交通居民混 合区、文化区、一般工 业区和农村地区
101~200	III	轻度污 染	黄色	长期接触,易感人群症状轻度加 剧,健康人群出现刺激症状	特定工业区
201~300	IV	中度污 染	橙黄色	一定时间接触后,心脏病和肺病 患者症状显著加剧,运动耐受力 下降,健康人群中普遍出现症状	—
≥300	V	重度污 染	红色	健康人群明显强烈症状,降低运 动耐受力,提前出现某些疾病	—

自 2001 年 1 月 1 日开始通过新闻媒体向社会发布空气质量日报以来,大庆市空气质量状况总体优良,在全国大气质量重点监测的城市中始终位列前茅,近年来大庆市环境空气质量处于一级与二级的天数占全年天数的比例均超过 95%,2005 年至 2009 年全市年均 API 值分别为 54、56、51、51、49。按国家对环保模范城市大气指标的规定,全年 $API < 100$ (即空气质量达到二级标准,日均浓度限值 $< 0.150\text{mg/m}^3$) 的天数要大于全年天数的 80%(292 天)。综上所述,大庆市的大气环境质量状况远远超过国家对环保模范城市的要求。

大庆市环境空气质量监测的项目共有四项:SO₂、NO₂、PM₁₀、降尘。通过对历年数据进行的分析汇总得出大庆市的 SO₂ 与 NO₂ 的污染较轻,而在 API 超过 50 需报告首要污染物的日子,基本上每日的首要污染物都是 PM₁₀。表 3-3 给出了大庆市 2005-2009 年各污染物指数的实测等级。

表 3-3 各污染物污染指数实测等级

年度	二氧化硫(SO ₂)		二氧化氮(NO ₂)		可吸入颗粒物(PM ₁₀)	
	API 值	等级	API 值	等级	API 值	等级
2005	13	优	11	优	54	良
2006	13	优	9	优	56	良
2007	13	优	9	优	51	良
2008	13	优	11	优	51	良
2009	11	优	9	优	49	优

由表 3-3 可见,大庆市空气质量等级可用 PM₁₀ 污染等级来表述,空气质量年均

API 值即为 PM₁₀ 的污染分指数值。对大庆市近年来 PM₁₀ 污染等级频率分析，有表 3-4。

表 3-4 大庆市近年 PM₁₀ 污染等级频率分布表(单位:%)

时间	I 级	II 级	III 级	IV 级	V 级
2001	56.4	39.2	3.6	0.8	0
2002	56.7	36.7	6.3	0.3	0
2003	45.8	49.0	5.2	0	0
2004	48.9	47.5	3.3	0	0
2005	54.0	43.5	2.5	0	0
2008	63.9	33.3	2.5	0.27	0
2009	69.6	29.3	1.1	0	0
2010	71.8	25.7	2.5	0	0

从表 3-4 中看出，“十五”期间空气优良天数逐年增加，空气质量Ⅲ、Ⅳ级的天数逐年减少。近三年来大庆市空气质量Ⅲ、Ⅳ级的频率极低，而空气质量Ⅴ级的情况从未发生。空气质量处于Ⅰ、Ⅱ级别的天数也是逐年上升，这表明大庆市的空气质量总体状况优秀并趋于更好。

3.2 大庆市 PM₁₀ 污染特征分析

对 PM₁₀ 污染的特征分析主要包括对 PM₁₀ 的年变化趋势分析，对 PM₁₀ 季节变化、月变化以及采暖期与非采暖期的变化特征分析，还包括对一些特殊情况下的 PM₁₀ 污染分析(比如沙尘暴情况下的分析)。

3.2.1 大庆市 PM₁₀ 污染年变化特征

对 2001 年至 2009 年 PM₁₀ 实测数据进行分析，得到大庆市历年 PM₁₀ 浓度年均值变化趋势，如图 3-2 所示。

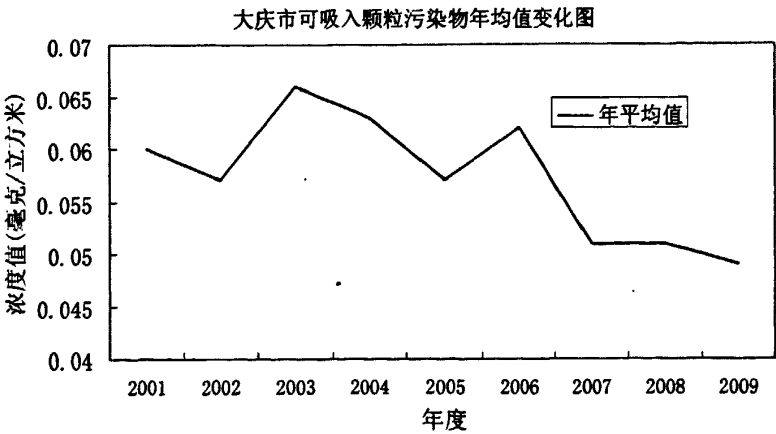


图 3-2 大庆市 PM₁₀ 浓度年均值变化图

由图可见, 2001 年至 2006 年大庆市 PM_{10} 浓度年均变化平稳, 分别为 0.060mg/m^3 、 0.057mg/m^3 、 0.066mg/m^3 、 0.063mg/m^3 、 0.057mg/m^3 、 0.062mg/m^3 , 始终在 0.060mg/m^3 上下波动。从 2006 年开始, PM_{10} 浓度有了明显的下降, 2007 年至 2009 年三年来, 浓度始终在 0.050mg/m^3 上下波动, 其主要原因是大庆市从 2005 年开展了以低空大气污染治理为重点的“蓝天工程”, 这项工程使得大庆市近年来 PM_{10} 污染状况大幅好转。

3.2.2 大庆市 PM_{10} 污染季节变化特征

对大庆市 PM_{10} 浓度分四季进行统计分析, 春季从 3 月至 5 月, 夏季从 6 月至 8 月, 秋季从 9 月至 11 月, 冬季从 12 月至次年 2 月, 得到大庆市 PM_{10} 按四季划分的等级频率表 3-5。

表 3.5 大庆市 2007 年冬至 2010 年夏 PM_{10} 四季污染等级频率分布表(单位:%)

时间	季节	I 级	II 级	III 级	IV 级	V 级
2007 年	冬	38.33	61.67	0	0	0
2008 年	春	47.83	43.48	7.6	1.09	0
	夏	82.60	17.40	0	0	0
	秋	79.35	17.39	2.17	0	0
2008 年	冬	44.44	55.56	0	0	0
	春	58.70	38.04	3.26	0	0
	夏	88.04	11.96	0	0	0
2009 年	秋	82.42	16.30	1.09	0	0
2010 年	冬	55.56	42.22	2.17	0	0
	春	67.39	27.17	5.40	0	0
	夏	80.43	19.57	0	0	0

从表中可以看出, 2007 年冬至 2010 年夏季的 PM_{10} 污染状况季节差异性很大, 冬季与春季 PM_{10} 污染较为严重, 而夏秋两季 PM_{10} 污染状况较轻。夏秋两季处于空气质量一级的天数平均为冬春两季天数的 1.6 倍。空气质量处于三级及以上的日子几乎全部在冬春两季, 夏秋两季空气质量处于三级及以上的天数极少。 PM_{10} 污染产生季节性差异、冬春两季空气质量较差的原因分析主要有以下几点:

(1) 每年春季为大庆市沙尘天气的高发期, 并且早春发生频率高于晚春, 此时春季强冷空气携外域沙尘入侵大庆市, 可在短时期内使全市 PM_{10} 污染短时间内恶化。并且

每年每年春季的 3 月、4 月大庆市仍旧处于采暖期(大庆市采暖期从每年 10 月 1 日至次年 4 月 1 日)，也会导致PM₁₀的污染加重。

(2)大庆市冬季的气候寒冷干燥，风速小，极易形成逆温天气。多年的观测资料显示，大庆市区冬季逆温的频率极高而且强度大，出现频率达 100%。逆温天气发生时，就像给城市罩上了一层被子，阻碍了空气的垂直运动，这导致大气输散污染物的能力差，污染物易在低空大气长时间滞留，使PM₁₀的污染趋势加重。

(3) 大庆市处于高寒地带，每年采暖期达六个月，由于大庆市是著名的石油能源城市，故采暖多使用渣油与天然气为燃料。可是近年来随着国际油价的上涨，大庆市采暖成本不断上涨，为节约成本，以煤代油气已成为必然趋势。近几年来大庆市冬季采暖的能源结构中燃煤所占比例逐年升高，大量燃煤导致烟尘排放量增多，因此在每年冬季由于燃煤需求量加大而极易形成典型的“煤烟”型污染。

(4)大庆市的供暖方式主要是采用区域性集中供暖，但在某些区尤其是城乡结合部和外来人员集聚居住区仍一定程度上存在着低矮散烧热源和小锅炉取暖的现象。由于小锅炉是燃煤散烧，没有除尘装置，很大程度上会导致PM₁₀污染的加重。

(5)冬季是民用电和工业用电的高峰期，电力资源的大量使用使得燃煤量大增，这在一定程度上也加重了PM₁₀的污染。

3.2.3 大庆市 PM₁₀污染月变化特征

对大庆市PM₁₀污染状况分月份进行统计分析，绘出PM₁₀污染指数各月均值变化曲线图 3-3。

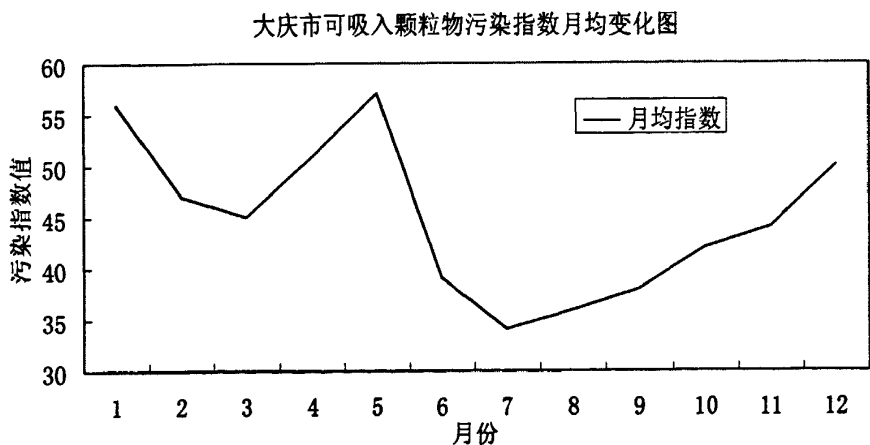


图 3-3 大庆市 PM₁₀污染指数月平均变化图

通过对图 3-3 的分析可知，大庆市PM₁₀月变化趋势显著，每年的 1 月、2 月、4 月、

5 月、11 月、12 月污染指数较高，3 月、6 月、7 月、8 月、9 月污染指数较低，近年来 PM₁₀ 污染指数月均值最低月为 7 月，最高月为 5 月。这与分析大庆市 PM₁₀ 污染的季节特征分析中得出的冬春两季污染较重，夏秋两季污染较轻的结论一致，其原因已在 3.2.2 中阐述。

3. 2. 4 大庆市 PM₁₀ 污染特殊情况分析

3. 2. 4. 1 对大庆市采暖期与非采暖期 PM₁₀ 污染特征的分析

对 2008 年 1 月至 2010 年 6 月的 PM₁₀ 污染等级分采暖期与非采暖期分别进行统计，得出大庆市采暖期与非采暖期 PM₁₀ 污染等级频率分布表，见表 3-6。

表 3-6 大庆市采暖期与非采暖期 PM₁₀ 污染等级频率分布表(单位:%)

时间	I 级	II 级	III 级	IV 级	V 级
2008 年非采暖期	73.22	23.49	2.73	0.55	0
2008 年采暖期	54.64	43.17	2.19	0	0
2009 年非采暖期	79.78	18.68	1.64	0	0
2009 年采暖期	58.79	40.66	0.55	0	0
2010 年非采暖期	74.86	22.40	2.73	0	0
2010 年采暖期	64.44	33.33	2.22	0	0

由表 3-6 可知，大庆市非采暖期与采暖期的 PM₁₀ 污染等级差别极大，非采暖期的空气质量主要以一级为主，2008、2009、2010 年空气质量一级的情况都占到非采暖期天数的 73%以上，是空气质量为二级天数的 3 倍以上，而在采暖期空气质量一级与二级的天数差别不大，这表明 PM₁₀ 污染等级与采暖期还是非采暖期关系非常大。其主要原因已在 3.2.2 中阐述。

3. 2. 4. 2 对大庆市沙尘天气时 PM₁₀ 污染特征分析

大庆市的沙尘来源主要是外部沙尘，但本地及周边沙源对大庆市的大气环境也有一定程度的干扰。大庆市的沙尘天气发生期较集中，春季为高发期，随着近年来大庆市加强了以防风固沙为重点的林业生态工程建设和以草原“三化”治理为重点的草原生态治理与恢复工程建设，西北风口的治理初见成效，大庆的沙尘天气的发生率也已经大大降低，但对大庆市春季沙尘的防治仍不可掉以轻心^[72]。下面对大庆市 2008 年 5 月 27 日发生的一次强沙尘天气过境中 PM₁₀ 的污染特征进行个例分析。

2008 年 5 月 27 日，受上游沙尘暴的影响，大庆市出现了强沙尘天气。对 2008 年

5 月 27 日至 5 月 31 日全市各区 PM₁₀ 的污染指数情况统计如下表 3-7。

表 3-7 2008 年一次强沙尘天气过程中全市各区 PM₁₀ 污染指数统计表

	2008.5.27	2008.5.28	2008.5.29	2008.5.30	2008.5.31
萨尔图区	52(Ⅱ级)	226(Ⅳ级)	168(Ⅲ级)	59(Ⅱ级)	28(Ⅰ级)
各 让胡路区	29(Ⅰ级)	279(Ⅳ级)	215(Ⅳ级)	67(Ⅱ级)	22(Ⅰ级)
区 龙凤区	20(Ⅰ级)	394(Ⅴ级)	97(Ⅱ级)	52(Ⅱ级)	32(Ⅰ级)
API 红岗区	48(Ⅰ级)	265(Ⅳ级)	156(Ⅲ级)	70(Ⅱ级)	28(Ⅰ级)
值 大同区	48(Ⅰ级)	301(Ⅴ级)	168(Ⅲ级)	65(Ⅱ级)	40(Ⅰ级)
全市	40(Ⅰ级)	296(Ⅳ级)	161(Ⅲ级)	62(Ⅱ级)	30(Ⅰ级)

从表 3-7 可以看出，5 月 26 日全市 API 为 43,5 月 27 日全市 API 为 40，空气质量等级均为一级。5 月 28 日，受强沙尘过境影响，全市 API 值迅速恶化到 296，接近五级重度污染。5 月 29、30 日，全市受浮尘影响，API 值分别为 161 和 62，空气质量等级分别为三级和二级。5 月 31 日开始随着沙尘暴影响的结束，全市空气质量才恢复至一级。5 月 28 日是沙尘污染的高峰期，这一天，全市 PM₁₀ 空气污染指数迅速恶化至往年 5 月污染指数均值的 5 倍以上，而同一天监测到的 SO₂ 和 NO₂ 的污染分指数分别为 16 与 12。这表明，强沙尘天气对 SO₂、NO₂ 的污染指数影响不大，而对 PM₁₀ 污染影响巨大。

由于全市各空气监测点的地理位置不同，此次沙尘暴对全市不同位置的影响不同，从上表可以看出，5 月 28 日全市 PM₁₀ 污染指数最高出现于龙凤区，高达 394。而受污染程度较轻的萨尔图区污染指数也有 226，这表明此次沙暴天气中沙尘路径径直略过大庆而并没有向东或者西偏，这使大庆市全境内空气质量短期内极度恶化，全市各区普遍受到此次沙尘天气的巨大影响。

突发的沙尘天气给大庆市的经济社会发展带来了严重的影响，主要表现在以下三个方面：其一，风沙产生的沙尘会加快石油、石化生产设备的磨损速度，并且会影响某些仪器的精度。同时，沙尘会导致浅层土壤盐碱化程度加剧，进而腐蚀地下的油气管线。其二，会对农业生产产生影响。沙化的土地会导致土地地力下降，农田减产。其三，会对人民的身体健康和生活质量产生影响。以 2001 年 4 月 7 日的特大沙尘暴为例，此次沙暴测定的室外含尘量最大为 500 mg/m³，超过国家标准的 20 倍，短期内导致医院呼吸系统疾病患者数量的剧增。对大庆市受沙尘天气影响而导致 PM₁₀ 污染的原

因分析如下:

(1)从气候来看,大庆市处于大陆性季风气候,降水少而风大,处于黑龙江省风力指数的高值区。尤其是4月和5月,平均风速为5.7 m/s,远远超过起沙的临近风速,而这段时间恰好降水稀少,大庆市每年3月至5月的降水仅有50mm左右,这样空气干燥、降水稀少、风力大,就使得干旱季节与风力高值季节在时间配置上吻合,大大增加了沙尘天气的隐患。

(2)大庆市处于黑龙江省土壤沙漠化的敏感区,迄今为止,全市仍有46万 hm^2 的沙化土地需要治理,分布于西北的杜尔伯特蒙古族自治县,肇源县、林甸县等地,市区也存在着大量的盐碱土、风沙土。同时由于大庆市远处的内蒙古自治区沙化严重,近处又毗邻敖古拉沙地,故几乎每年春季都会有强冷空气携带大量的西北部外域沙尘入侵大庆,这些本底因素导致大庆市对沙尘的防治任重道远。

(3)由于在石油开采储运过程中经常发生漏油现象,同时由于在不少采油区仍有偷油放油现象,使采油区植被被污染而死亡进而导致大量的沙化土地裸露在外。再加上乱砍乱伐,过度放牧等原因使大庆周边及市区植被覆盖率很低。这些因素综合起来使大庆沙尘天气的风险大增。

3.2.4.3 对大庆市过年期间(烟花爆竹燃放期间) PM_{10} 污染特征的分析

对大庆市2008、2009、2010三年除夕至初五这六天内的 PM_{10} 的污染指数进行比对分析。2008年除夕至初五这六天为2月7日至2月12日,全市 PM_{10} 的API值分别为86、59、61、62、86、32,均值为64,高于当月 PM_{10} 的API均值47。2009年为1月25日至1月30日,全市 PM_{10} 的API值分别为39、66、58、68、69、92,均值为65,高于当月的 PM_{10} 的API均值60。2010年为2月13日至2月18日,全市 PM_{10} 的API值分别为59、80、52、44、57、56,均值为58,高于当月的 PM_{10} 的API均值47。这三年过年期间全市 PM_{10} 的污染指数分别比当月均值偏高36.2%、8.3%、23.4%,这表明,过年期间全市 PM_{10} 污染明显重于平时,这与过年期间烟花爆竹的燃放关系很大。

3.3 大庆市 PM_{10} 污染的源解析及防治对策

3.3.1 大庆市 PM_{10} 污染的来源解析

大庆市 PM_{10} 污染来源呈现多样性,其来源一是能源结构中的燃煤导致的煤烟型污染,电力、化工、石油加工、供暖等以燃煤为主要燃料的行业每年都向大气中排放大

量的工业废气与粉尘,成为 PM_{10} 污染的点源。来源二是城市扬尘形成的面源,分为自然因素与人为因素两种。自然因素包括外域沙尘或城市周边沙尘的侵袭,或是由于风力过大而导致的城市中沙化土地的扬尘。人为因素一是城市建设过程中建筑工地施工过程产生的扬尘,其二是机动车行驶时气流扰动而造成的二次扬尘。来源三是机动车行驶过程中排放废气而形成的线源。汽车在行驶过程中排放的 NO_x 、烃类等污染物会衍生成颗粒物,同时行驶过程中因不完全燃烧而排放的黑烟更是一种对人体危害极大的可吸入颗粒物,城市中的 $PM_{2.5}$ 的主要来源就是机动车的尾气。其来源四就是一些局部的、不固定的因素。主要包括节假日或红白喜事燃放烟花爆竹,固体废弃物的焚烧,餐饮行业排放的油烟等。

3.3.2 大庆市 PM_{10} 污染的防治对策

由于大庆市 PM_{10} 污染来源途径的多样化,故对 PM_{10} 污染的防治应本着综合治理,统筹兼顾的原则,充分认识到这是一项长期而艰巨的民生系统工程,这样才能确保防治工作的深入开展,做到对各个污染源的全面覆盖。针对大庆市 PM_{10} 污染的防治问题,提出以下几点对策与建议^[73-74]:

(1)要加强对煤烟型污染的控制,拔除 PM_{10} 污染的点源。全市范围内应继续大力推进区域性的集中供热,合理规划全市供暖布局;坚决取缔低矮散烧热源和自烧燃煤小锅炉,尤其是要加大对城乡结合部与外来人员集聚居住区的监管力度;应加大对全市能源结构的改革调整,合理分配煤、燃油、天然气的比例,推广清洁能源的使用;针对工业企业的废气排放方面,各企业应继续加大对环保除尘设备的投入力度,研究引进节能清洁技术,提高燃料的燃烧利用率,推广使用型煤;坚决取缔“三高一低”的企业,加强对重点企业的监管和污染治理。

(2)要加强对城市扬尘的控制,减少 PM_{10} 污染的面源。应继续开展对大庆市西北风口的绿化工程,营造防风固沙林,加强草原“三化”综合治理;扩大城市绿地公园的面积,着力减少裸露沙地的面积。针对城市建筑工地的扬尘污染,应建立环保、建委、城市管理、公安等部门的协调联动模式,重点打造出控制城市工地扬尘的“防护网”,加强对主城区的施工工地和施工车辆的管理,按环保要求控制扬尘措施。对主要街道采用机械化洒水,增加机械化清洁、洒水车辆以减少道路扬尘。

(3)要加强对城市机动车辆尾气排放的控制,降低 PM_{10} 污染的线源。应限制高排放量的机动车在市区行驶时间和路程;鼓励发展小排量机动车和清洁能源汽车。应加

强道路建设，完善道路系统，提高交通运输效率以分流市区机动车流量。

3.4 本章小结

本章主要对大庆市空气质量状况和 PM_{10} 的污染特征从统计学角度进行了分析。得出结论如下：

(1) 该市空气质量状况总体优良并趋于更好，远优于国家对环保模范城市的要求。

(2) 通过对大庆市 PM_{10} 污染的年变化特征分析得出，从 2001 年至 2006 年大庆市 PM_{10} 浓度年均变化平稳，始终在 $0.060\text{mg}/\text{m}^3$ 上下波动。自 2006 年开始， PM_{10} 浓度明显的下降，表明近年来大庆市对 PM_{10} 污染的整治初见成效。

(3) 对大庆市 PM_{10} 污染季节变化特征分析得出，大庆市 PM_{10} 污染季节性差异显著，冬季与春季 PM_{10} 污染较为严重，而夏秋两季 PM_{10} 污染状况较轻。

(4) 对大庆市 PM_{10} 污染月变化特征分析得出，大庆市 PM_{10} 月变化趋势显著，每年的 1 月、2 月、4 月、5 月、11 月、12 月污染指数较高，3 月、6 月、7 月、8 月、9 月污染指数较低。

(5) 对大庆市 PM_{10} 污染特殊情况的分析得出，大庆市采暖季的 PM_{10} 污染较非采暖季严重，大庆市在沙尘暴过境期间和过年期间的 PM_{10} 污染明显较平时严重。

(6) 对大庆市 PM_{10} 污染的来源进行了解析并对大庆市 PM_{10} 污染的防治给出了具体有效的对策。

第四章 大庆市 PM₁₀ 污染与气象要素关系分析

表征大气状态的物理量在气象学中统称为气象要素,主要包括气温、气压、相对湿度、降水、能见度、风速、日照时数等。城市的空气质量与气象要素联系密切,因为这些重要气象要素对污染物的扩散、迁移、转化都有着巨大影响。有关研究表明,在不同的天气状况下,同一污染源排放造成的大气污染物浓度可相差几十倍乃至几百倍。因此,对城市 PM₁₀ 的污染气象学研究至关重要。

本章主要研究以下内容:(1)以大庆市气象局提供的 2006 年 1 月 1 日至 2010 年 6 月 30 日的逐日气象要素实测值为数据基础,对大庆市的气象要素特征分季节进行统计分析。(2)对 PM₁₀ 空气污染指数高值日的气象要素进行统计分析,找出污染高值日各气象因子的特征。(3)对 2008 年 1 月 1 日至 2010 年 6 月 30 日的 PM₁₀ 逐日污染指数值与同期的气象要素实测值分四季进行相关分析,找出与大庆市不同季节 PM₁₀ 污染情况相关性高的气象因子。

4.1 大庆市气候特征概述

大庆市地处黑龙江省松嫩平原中部,在纬度 45°46′~46°55′,东经 124°19′~125°12′ 之间。大庆市光照充足,降水偏少,冬季时间长而寒冷干燥,夏秋凉爽。全市年平均气温 4.2℃,最冷月平均气温 -18.5℃,极端最低气温 -39.2℃;最热月平均气温 23.3℃,极端最高气温 39.8℃,年均无霜期 143d;年均风速 3.8m/s,年大于 16 级风日数为 30d;年降水量为 427.5mm,年蒸发量为 1635mm;年日照时数为 2726h,年太阳总辐射量为 491.4KJ/cm²。

由于受大陆性气候和季风的共同影响,大庆市年气候变化多样,四季分明。冬季受大陆冷高压控制影响,盛行偏北风,寒冷少雪,热量严重匮乏;夏季受副热带海洋气团影响,盛行偏南风,夏季前期干热,后期降水集中且变率大,时有旱涝;春秋两季为过渡季节,春季冷暖多变,干旱多风,易发沙尘天气;秋季多寒潮,降温急剧。

4.2 大庆市气象要素特征统计分析

对大庆市历年气象要素特征进行统计分析,旨在研究近年大庆市气候和气象要素的变化趋势,为 PM₁₀ 空气污染预报提供依据,数据集中的逐日实测气象因子主要包括气压因子(08h 气压、日均气压、日气压最高值、水汽压),气温因子(08h 气温、日均气温、日最高气温),相对湿度因子(08h 相对湿度、日均相对湿度),能见度因子(08h 能见度、14h 能见度),日照时数因子,总云量因子,风速因子(08h 风速、14h 风速、20h 风

速、日均风速、日最大风速)等。

4.2.1 气压分析

气压是指大气压强，气压对污染物污染程度的影响主要表现在：低压控制时，气流从四周高压流入中心低压，产生气流辅合上升过程，这时大气一般处于不稳定或者中性状态，风速较大，利于污染物的转移、扩散和稀释。高压控制时，大气稳定度较高，风速较小，容易形成下沉逆温现象，从而阻止污染物在在垂直方向上扩散，如果高压移动缓慢，长期停留在某一区域，就会造成由于高压控制而形成的小风速和稳定层结，不利于污染物的转移、扩散和稀释。

对气压因子的分析包括 08h 气压、日均气压，大庆市气压因子的季节变化特征如下表 4-1 所示。

表 4-1 大庆市四季日均气压、08h 气压均值年分布表(单位:hPa)

	日均气压				08h 气压			
	春季	夏季	秋季	冬季	春季	夏季	秋季	冬季
2006	992.5	987.4	999.2	1004.8	993.3	988.0	999.8	1005.4
2007	993.2	987.5	999.8	1005.5	994.0	988.2	1000.7	1005.8
2008	992.6	988.9	997.1	1002.1	993.2	989.6	997.7	1003.4
2009	994.0	986.1	998.8	1003.7	994.7	986.8	999.6	1004.6
平均	993.1	987.5	998.7	1004.0	993.8	988.2	999.5	1004.8

从表中可知 2006 至 2009 年春季日均气压平均为 993.1hPa，夏季为 987.5 hPa，秋季为 998.7 hPa，冬季为 1004 hPa。四季日均气压的均值大小依次为冬季>秋季>春季>夏季。2006 年至 2009 年春季 08h 气压平均为 993.8 hPa，夏季为 988.2 hPa，秋季为 999.5 hPa，冬季为 1004.8 hPa。四季 08h 气压的均值大小依次为冬季>秋季>春季>夏季。从历年的统计数据来看，四季的 08h 气压略高于日均气压。由于秋冬两季的气压因子均值高于春夏两季，故秋冬两季污染物较之春夏两季不易扩散、稀释。

4.2.2 气温分析

地面观测气温是指距离地面 1.5m 高百叶箱中观测到的空气温度，气温与污染物扩散之间并没有物理上的直接联系，但地面气温可以通过影响大气稳定度进而影响污染物的扩散稀释过程。对气温因子的分析包括 08h 气温、日均气温、日最高气温。大庆市四季气温因子统计分析如下表 4-2。

表 4-2 大庆市四季气温因子均值逐年分布表(单位:℃)

	日均气温				08h 气温				日最高气温			
	春	夏	秋	冬	春	夏	秋	冬	春	夏	秋	冬
2006	5.9	21.9	5.8	-10.1	4.7	21.6	4.1	-14	11.7	25.7	12.4	-6.6
2007	6.2	23.4	6.6	-13.1	5.0	23.4	4.7	-16.1	13.3	28.4	13.2	-7.1
2008	8.5	22.8	6.2	-15.3	6.9	22.4	4.5	-18.7	12.7	26.4	12.2	-7.4
2009	7.1	20.5	4.1	-18.7	5.7	20.2	2.3	-21.5	14.3	27.5	11.7	-9.7
平均	6.9	22.2	5.7	-14.3	5.6	21.9	3.9	-17.6	13.0	27.0	12.4	-7.7

由上表可知，气温的季节差异性明显，夏季气温高，冬季气温低，春秋气温基本持平。从历年的气温统计数据来看，春季、秋季和冬季的 08h 气温都明显低于日均气温，而夏季的 08h 气温与日均气温基本持平。

4. 2. 3 湿度分析

大气中的水汽含量虽然不多，却是大气中极其活跃的成分，在天气与气候变化中扮演着极其重要的角色。大气中的水汽含量有多种测量方法，在日常生活中人们最为关注的是水汽压、绝对湿度和相对湿度三个因子。

水汽压(e)是大气压力中水汽的分压力，和气压一样用百帕(hPa)来度量。在一定温度下空气中水汽达到饱和时的分压力，称为饱和水汽压(E)。饱和水汽压随着气温的升高而迅速增加。绝对湿度(a)是指单位体积内湿空气中含有的水汽质量，也就是空气中的水汽密度。绝对湿度不容易测量，实际中使用较少。相对湿度(f)指空气中的水汽压(e)与同一温度下的饱和水汽压 E 之比，以百分数表示是 $f=e/E\%$ 。

湿度因子的大小会随着气温的变化而变化，当水汽压不变时，气温升高，饱和水汽压增大，相对湿度会减小，相反，气温降低，相对湿度会增大。大庆市湿度因子包括 08h 湿度(上午 8 时的相对湿度)，日均相对湿度和日均水汽压。历年分季节统计结果如下表 4-3。

表 4-3 大庆市四季湿度因子均值逐年分布表

	相对湿度日均(单位:%)				相对湿度 08h(单位:%)				日均水汽压(单位:hPa)			
	春	夏	秋	冬	春	夏	秋	冬	春	夏	秋	冬
2006	41.7	71.7	57.5	62.2	46.5	72.7	64.4	70.0	462.4	1802.6	622.1	118.5
2007	53.5	59.7	53.0	58.1	56.4	61.9	60.3	65.1	599.4	1796.4	578.4	92.7

表 4-3 大庆市四季湿度因子均值逐年分布表（续）

	相对湿度日均(单位:%)				相对湿度 08h(单位:%)				日均水汽压(单位:hPa)			
2008	50.3	65.1	54.7	66.0	57.1	67.1	60.7	69.6	587.7	1816.8	614.2	135.2
2009	45.4	72.5	60.7	68.9	51.3	73.9	67.1	74.2	451.6	1737.6	601.1	109.8
平均	47.7	67.3	56.5	63.8	52.8	68.9	63.1	69.8	525.5	1788.6	604.0	114.1

由上表可见，四季日均湿度的均值大小依次为夏季>冬季>秋季>春季，四季 08h 相对湿度的大小依次为冬季>夏季>秋季>春季，08h 相对湿度均值明显高于相对湿度日均值。这与大庆市的气候特征有关。大庆市冬夏两季降水较多，而春季干燥少雨并且风沙天气较多，湿度较小。

4. 2. 4 能见度分析

能见度指人们在正常的天气状况下，能够从天气背景中看到和辨认出的目标物的最大水平距离，单位一般用 km 表示。大庆市地面实测的能见度因子有 08h 能见度和 14h 能见度，分别表示 8:00 与 14:00 的大气能见度。

影响能见度的因素主要包括降水、沙尘等天气状况，湿度、风速等气象要素以及 PM₁₀ 的污染状况等。大气污染状况与能见度的联系密切，能见度的高低直接反映了空气质量的好坏，一般情况下，城市 PM₁₀ 污染指数高值时，空气中颗粒物对阳光的吸收与散射加强，能见度下降。与之相反，城市 PM₁₀ 污染指数低时，能见度较高。大庆市历年的 08h 能见度与 14h 能见度均值逐年统计如下表 4-4。

表 4-4 大庆市四季能见度因子均值逐年分布表(单位:Km)

	08h 能见度				14h 能见度			
	春	夏	秋	冬	春	夏	秋	冬
2006	16.7	17.8	17.4	16.0	19.7	20.1	20.5	19.3
2007	18.7	19.6	19.7	18.3	19.4	20.9	21.6	20.5
2008	17.8	18.8	18.4	17.4	19.6	19.8	20.3	19.1
2009	17.4	19.2	19.0	17.1	19.5	21.1	20.6	19.1
平均	17.6	18.9	18.6	17.2	19.5	20.5	20.8	19.4

由上表可以看出，历年 08h 能见度明显低于 14h 能见度。四季 08h 能见度均值大小依次为夏季>秋季>春季>冬季，14h 能见度均值大小依次为秋季>夏季>春季>冬季。由此可知夏秋两季能见度明显优于冬春两季,其主要原因是冬春两季 PM₁₀ 污染较夏秋两

季严重，而 PM₁₀ 污染能导致大气能见度的下降。

4.2.5 日照日合计分析

日照合计时数在本论文中指的是气象部门使用仪器测得的 24h 内白昼和曙暮光的时数，单位为小时(h)。日照时数长表明太阳辐射强，从而使地表增温的幅度较大，加热了低空大气，形成了递减逆温层结，大气处于不稳定状态，污染物扩散稀释的速率加快，降低了污染物的浓度。相反，若日照合计时数较短，则污染物浓度升高。下表 4-5 为大庆市四季日照时数均值的逐年分布表。

表 4-5 大庆市四季日照时数均值逐年分布表(单位:小时)

	日照时数			
	春季	夏季	秋季	冬季
2006	7.4	7.0	6.9	5.7
2007	7.0	8.5	7.1	6.2
2008	7.7	8.0	6.8	5.3
2009	7.3	8.4	7.0	5.6
平均	7.4	8.0	6.9	5.7

由表可知，春夏两季日照时数较长，分别为 7.4h 和 8.0h，冬季日照时数均值较其他三季低，仅为 5.7h，其原因一方面与北半球的气候特点有关，另一方面也与冬季 PM₁₀ 污染严重密不可分。

4.2.6 总云量分析

总云量是指天穹被全部云遮蔽的成数，我国将视野能见的天空分为十份，被云遮蔽了几成，总云量的数值就是几。云量在 3 以下为晴天，3-8 为少云或多云天气，超过 8 为阴天。

总云量因子与大气的运动情况联系密切，云对太阳辐射和地面辐射有反射作用，云层越厚，对辐射的反射作用越强。白天，云阻挡太阳辐射，夜晚，云层的存在使得地面的长波辐射被云层反射回地面，使得地面不容易冷却，而太阳辐射影响着大气层与地表的热状况、温度的变化，制约着大气的运动状态。大量云层的存在会使大气区域对流运动减弱、大气稳定度增加，这样不利于污染物的稀释扩散，易使污染物浓度增加。表 4-6 为大庆市四季总云量均值逐年分布表。

表 4-6 大庆市总云量均值逐年分布表(单位:成)

	总云量			
	春季	夏季	秋季	冬季
2006	5.1	6.3	3.7	2.9
2007	4.8	5.1	2.8	1.6
2008	4.7	5.3	3.3	2.4
2009	5.0	6.0	3.1	2.2
平均	4.9	5.7	3.2	2.3

由表可见，大庆市总云量的四季均值变化情况为:夏季>春季>秋季>冬季。

4. 2. 7 风速分析

风速是指空气水平运动的速度，单位为 m/s 或者 km/h。一般情况下气象部门测定的风速是指一定时间段内风速的平均值，如 2min 风速或者 10min 风速。风速因子与污染物的扩散、稀释联系密切，排入大气的污染物会随风扩散至其他地区。风速高值时，单位时间内会将污染物输送的距离更远且会使更多的清洁空气与污染物混合，从而降低污染物浓度。反之，风速低值时，对污染物的稀释扩散作用较弱，污染物浓度不易降低。但有时风速高时会导致城市地面灰尘扬起或沙尘天气的加剧，对大庆市而言，春季风速值较高，就极易导致沙尘天气的产生，从而使PM10污染指数增高。对大庆市风速因子的分析主要包括 2 分钟 08h、2 分钟风速 14h、2 分钟风速 20h、2 分钟风速日均，日最大风速五个。下表 4-7 为大庆市各风速因子四季均值逐年分布表。因为夏秋风速因子情况相似，故合并统计。

表 4-7 大庆市风速因子四季均值逐年分布表(单位:m/s)

	风速 08h			风速 14h			风速 20h			风速日均			风速日最大		
	春	夏秋	冬	春	夏秋	冬	春	夏秋	冬	春	夏秋	冬	春	夏秋	冬
2006	3.5	2.5	2.3	4.4	3.1	2.9	2.4	1.9	1.7	3.2	2.4	2.3	6.6	5.3	4.2
2007	2.9	2.6	2.0	3.9	3.2	2.9	2.1	1.8	1.7	2.8	2.4	2.1	6.0	5.2	4.1
2008	3.1	2.4	2.1	3.8	3.3	2.9	2.1	1.7	1.5	2.9	2.3	2.1	5.5	5.0	4.3
2009	3.3	2.7	1.8	4.1	3.2	2.7	2.2	1.8	1.6	3.1	2.4	1.9	6.5	5.1	4.0
平均	3.2	2.6	2.1	4.1	3.2	2.9	2.2	1.8	1.6	3.0	2.4	2.1	6.2	5.2	4.2

由上表可以看出，大庆市各风速因子均值差异显著，14h 风速均值为最高，20h 风

速均值最低。从四季风速因子划分来看,春季的风速各因子均值为最高,夏秋次之,冬季为最低。这符合大庆市春季多风的气候特点。

4.2.8 降水量分析

降水对大气污染物的清洁作用巨大,它是大气自清洁的一个重要过程。大庆市的降雨主要集中在 6-8 月,降雪主要集中在每年的 1、2、11、12 四个月。有关研究表明,降水可以有效的降低 PM_{10} 、 SO_2 、 NO_2 等污染物的浓度。一般情况下,降水持续时间越长,降水强度越大,则对污染物的去除程度越大。而且冬季降雪对污染物的去除能力大于夏季降雨对污染物的去除能力。

4.2.9 特殊情况分析

以上对影响大气污染的各主要气象因子进行了分析,此外还有一些气象参数对污染物浓度的影响巨大。如风向、大气稳定度、混合层特征等,这些因子对城市大气污染物的扩散稀释不容忽视,但是由于其过于偏重气象学专业,故本论文不予介绍。本文重点介绍一种重要的特殊天气情况,即逆温过程。

一般情况下,在低层大气中,通常气温是随着高度的增加而降低的。但有时在某些层次可能出现相反的情况,气温随高度的增加而升高,这种现象称为逆温。出现逆温的大气层称为逆温层。逆温的成因主要有以下几种:一是地面辐射冷却;二是空气平流冷却;三是空气下沉增温;四是空气的乱流混合;五是锋面上形成的逆温。

在逆温层中,暖而轻的空气位于冷而重的空气上面,形成一种极其稳定的空气层,像层棉被一样,笼罩在近地层的上空,严重阻碍着空气的对流,因此,近地层空气中的大气污染物无法向外向上扩散,只有飘浮在逆温层下面的空气层中形成云雾,近而降低了大气能见度。更严重的是,大气中的污染物无法及时稀释扩散,导致浓度增大。肖红,夏文前^[75]等研究了大庆市 1 月、4 月、6 月等月份逆温与大庆市 PM_{10} 浓度的关系,得出了逆温是制约大庆市大气污染尤其是冬季大气污染最主要的气象条件之一,逆温的持续性、强度、层高与大气污染物的稀释扩散成正比关系。大庆地区冬季逆温出现的频率高、厚度大、强度大,这也是大庆市冬季污染相对严重的主要原因之一。

4.3 大庆市 PM_{10} 的 API 高值日的污染气象学分析

由表 3-2 可知,当 API 大于 100 时,城市空气质量处于轻度污染状态,长期接触对人体健康有一定危害。故本文将大庆市 PM_{10} 的 API 值高于 100 的日子定为 API 高值

日，简称污染日。应着重对污染日做污染气象学分析，统计出这类高污染日的气象因子的变化特征，以便更好的找出各气象因子与污染指数之间的关系。对样本数据中大庆市PM₁₀的 API 大于 100 的日子的气象要素统计如下表 4-8。在 912 日的 PM₁₀ 实测指数与同期气象因子数据的样本集中，共找出 PM₁₀ 的 API>100 的日子 22 天。

表 4-8 PM₁₀污染指数>100 日对应气象因子统计表

日期	API 值	气压日 均	云 量	日 照	气 温	相对 湿度	能见度 08h	能见度 14h	风速日 均	日最大 风速
春季										
20080315	121	987.3	10.0	0.3	4.0	55	12	15	3.4	7.4
20080318	107	988	5.3	6.9	-0.2	42	15	20	5.3	8.1
20080414	125	987.8	6.3	9.4	11.8	29	17	21	3.2	8.3
20080419	112	992.1	3.3	9.8	21.9	27	18	20	3.9	8.5
20080420	110	995.5	8.7	3.0	19.9	22	16	16	4.4	7.9
20080430	108	977.3	5.7	3.9	15.1	64	16	16	5.4	10.6
20080528	296	979	9.3	5.1	13.4	64	16	12	1.3	9.9
20080529	161	984.2	6.7	2.5	11.7	87	15	15	2.0	9.1
20090507	152	989.3	6.4	6.2	19.9	25	12	15	3.1	9.0
20090519	116	984.6	5.5	13.1	16.3	33	12	12	1.8	9.5
20090526	106	995.1	6.1	9.8	21.5	25	14	14	3.9	7.2
20100408	100	982.4	5.3	0	7.4	54	12	15	6.9	11.3
20100409	173	992.4	6.6	9	0.9	41	15	15	4.8	9.1
20100507	152	985.6	6.4	11.6	9.9	68	12	12	3.4	9.5
20100519	116	973.7	8.8	5.5	14.1	86	15	15	6.2	9.4
20100521	106	987.1	8.6	9.4	20.6	66	16	16	2.5	9.7
春均值	135	986.3	6.8	6.6	13.0	49	15	16	3.8	9.0
秋季										
20081022	153	998.7	3.3	0	8.9	85	9	9	5.6	7
20081023	110	988.4	10.0	0	9.4	79	9	9	2.6	8.2
20091106	102	984.3	2.6	2.1	8.6	52	12	12	4.9	9.6
秋均值	122	990.5	5.3	0.7	9.0	72	10	10	4.4	8.3

表 4-8 PM₁₀污染指数>100 日对应气象因子统计表 (续)

日期	API 值	气压日 均	云 量	日 照	气温	相对湿 度	能见度 08h	能见度 14h	风速 日均	日最大 风速
冬季										
20090107	100	1012.7	6.1	5.4	-17.8	75	12	15	1.4	3.7
20100118	104	1007.7	8.2	0	-17.7	83	12	12	3.3	4.3
20100119	127	1000.2	6.7	0	-11.3	86	12	15	2.8	5.1
冬均值	110	1006.9	7	1.8	-15.6	81	12	14	2.5	4.4

由表可知,大庆市 2008 年 1 月 1 日至 2010 年 6 月 30 日这 912 天中出现的 PM₁₀ 的 API>100 的日子 73%都出现于春季,秋冬两季出现几率较小,而夏季则一天也没有出现。以下分季节对各污染日的气象要素进行分析。

对春季污染日的各气象因子均值与前文历年春季各气象因子均值比对后发现,春季污染日气压因子均值为 986.3hPa,小于历年春季日均气压因子均值 993.1hPa。平均气温为 13.3℃,高于历年春季均值 6.9℃。相对湿度为 49%,与历年均值的 47.7%基本持平。能见度因子上,历年春季的 08h 能见度均值与 14h 能见度均值分别为 17.6km 与 19.5km。高于污染日的 08h 能见度 15km 与 16km。日照日合计污染日均值为 6.6h,略低于历年春季均值的 7.4h。总云量因子方面污染日为 6.8,历年春季均值为 4.9,污染日的总云量高于历年均值。污染日的日均风速均值为 3.8m/s,日最大风速均值为 9.0m/s,大于历年春季均值的 3.0m/s 与 6.2m/s。由此可见,春季污染日在气象上具有气压较低、高气温、低能见度、低日照、高云量、高风速的特点。

对秋季污染日各气象因子均值与前文历年秋季各气象因子对比后发现,秋季污染日的气象因子具有日均气压因子略低、气温基本持平、相对湿度高、能见度偏低、日照偏低、总云量偏高和风速偏高的特点。

对冬季污染日各气象因子均值与前文历年冬季各气象因子对比后发现,冬季污染日的气象因子具有日均气压因子偏高、气温偏低、相对湿度偏高、能见度偏低、日照日合计值偏低、总云量偏高和风速基本持平的特点。

对各污染日前五日与后五日各气象因子均值进行统计,并与污染日当日的气象因子数据进行对比。因为数据统计量较大,故不列表写出。直接得出结论如下:春季与秋季各污染日气压低于前五日气压均值与后五日气压均值,污染日当天气压有明显的下降趋势,而后气压又上升,呈“V”形分布。冬季各污染日气压略高于前五日气压均值与后五日气压均值,污染日当天有明显上升趋势,而后又下降,呈“八”形分布。

气温因子方面, 污染日当日气温与前后五日气温均值对比变化趋势不明显。总云量方面, 在污染日当天总云量要高于前后五日均值, 以阴天为主。日照因子方面, 在污染日当天的日照时数要明显少于前后五日的日照日合计的均值。相对湿度方面, 春季污染日的相对湿度与前后五日相对湿度均值对比变化不明显, 秋冬两季污染日相对湿度要高于前后五日相对湿度的均值。能见度方面, 污染日当日的 08h 能见度与 14h 能见度明显低于前五日与后五日的能见度因子均值, 污染导致能见度下降。风速因子方面。各污染日的当日风速值高于前五日与后五日的风速均值, 理论上风速能使污染物加速稀释扩散, 而有时, 风速大可能会导致风沙天气进而使 PM_{10} 污染指数增高, 所以应进一步对各季节的风速因子进行分析。

4.4 大庆市各季 PM_{10} 污染与气象要素相关关系

4.4.1 相关分析的基本概念

前文已对大庆市的气象因子分季节进行了统计分析, 并且对大庆市污染指数高值日的气象特点进行了统计分析, 目的是找出 PM_{10} 污染趋势与各气象因子的联系。本节将用统计软件 SPSS 对 PM_{10} 的污染指数与各气象因子进行相关分析^[76], 应用统计学与数学原理分析 PM_{10} 污染指数与气象要素之间的关系。

描述变量之间线性的相关程度的强弱, 并用适当的统计指标表示出来的过程称为相关分析。描述变量间相关关系的方法有很多, 比较直接的有绘制散点图, 但是这种方法不很精确。为了更加准确的描述变量之间的线性相关程度, 可以通过计算相关系数来进行相关分析。在二元变量的相关分析过程中比较常用的几个相关系数有 Pearson 相关系数、Spearman 秩相关系数和 Kendall 相关系数。本文应用 Pearson 和 Spearman 相关系数, 其介绍如下:

(1) Pearson 相关系数。一般用于对定距变量的数据进行计算, 即分析两个连续性变量之间的关系, 其计算公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4-1)$$

(2) Spearman 秩相关系数。用于描述分类或等级变量之间、分类或等级变量与连续变量之间的关系。其计算公式如下:

$$r_s = 1 - \frac{6 \sum_{i=1}^n (R_i - Q_i)^2}{n(n^2 - 1)} \quad (4-2)$$

其中 R_i 代表 x_i 在 $(x_1, x_2, x_3...x_n)$ 中的秩, 所谓秩是指 x_i 在 $(x_1, x_2, x_3...x_n)$ 中按照一定准则的排列顺序。 Q_i 代表 y_i 在 $(y_1, y_2, y_3...y_n)$ 中的秩。

上述两种相关系数在实际应用计算中都要对其进行假设检验, 使用 t 检验方法检验其显著性水平以确定其相关程度。研究表明, 在正态分布假定下, Spearman 秩相关系数与 Pearson 相关系数在效率上是等价的, 而对于非正态分布或者分布不明的数据, 则采用 Spearman 秩相关系数更为合适。故本论文将两种相关系数的计算结果都列出, 综合分析以确定相关关系。其中两种相关系数的计算与检验过程由 SPSS 软件完成。

计算各季节 PM₁₀ 污染指数与气象因子的相关系数是为了找出不同季节中影响 PM₁₀ 污染趋势的相关性高的气象因子, 为后面分季节建模时选取预报因子打下基础。论文所采用的春季数据样本集共 276 天, 夏季数据样本集共 214 天, 秋季样本集共 182 天, 冬季样本集共 240 天, 时间跨度为 2008 至 2010 三年。

为了更好的研究影响 PM₁₀ 污染的各种因子, 在气象因子之外加入一个特征变量, 即前日 PM₁₀ 的 API 值, 有关研究表明, 前日 PM₁₀ 的 API 值对当日 PM₁₀ 的 API 值影响巨大, 因为正常情况下 PM₁₀ 污染是一个积聚的过程, 故将其与气象因子一同进行相关分析, 现将需要研究的各因子列表, 见下表 4-9。

表 4-9 各气象因子与特征因子表

X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
本站水汽 压	气压 08h	气压日均	气压日最 高	前日气压 日均	前日 08h 气压	总云量	日照日合 计
X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆
气温 08h	气温日均	日最高气 温	前日气温 日均	前日与当 日温差	能见度 08h	能见度 14h	相对湿度 08h
X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	X ₂₂	X ₂₃	
相对湿度 日均	风速 08h	风速 14h	风速 20h	风速日均 值	日风速最 大值	前日 PM ₁₀ API	

上表中 X₁、X₂、X₃、X₄、X₅、X₆ 为气压因子, X₇ 为云量因子, X₈ 为日照因子, X₉、X₁₀、X₁₁、X₁₂、X₁₃ 为气温因子, X₁₄、X₁₅ 为能见度因子, X₁₆、X₁₇ 为湿度因子, X₁₈、X₁₉、X₂₀、X₂₁、X₂₂ 为风速因子, X₂₃ 为特征因子。以下分四季对大庆市 PM₁₀ 的 API 与各因子进行相关分析。

4. 4. 2 春季各因子的相关分析

将春季PM₁₀的 API 与各影响因子做相关分析，得到结果如下表 4-10。

表 4-10 春季PM₁₀ 的 API 与各因子相关分析结果

因子	因子含义	Pearson 系 数	Spearman 系数	因子	因子含义	Pearson 系 数	Spearman 系数
X ₁	本站水汽压	0.224**	0.220**	X ₁₃	前日与当 日温差	0.055*	-0.001
X ₂	气压 08h	-0.340**	-0.298**	X ₁₄	能见度 08h	-0.482**	-0.337**
X ₃	气压日均	-0.329**	-0.302**	X ₁₅	能见度 14h	-0.446**	-0.340**
X ₄	气压日最 高	-0.314**	-0.273**	X ₁₆	湿度 08h	-0.043	-0.164
X ₅	前日气压 日均	-0.261**	-0.117	X ₁₇	相对湿度 日均	-0.059	-0.117
X ₆	前日气压 08h	-0.256**	-0.115	X ₁₈	风速 08h	0.156	0.331**
X ₇	总云量	0.231*	0.183	X ₁₉	风速 14h	0.029	0.126
X ₈	日照日合 计	-0.136*	-0.192**	X ₂₀	风速 20h	0.273**	0.410**
X ₉	气温 08h	0.270**	0.295**	X ₂₁	风速日均	0.251**	0.325**
X ₁₀	气温日均	0.257**	0.298**	X ₂₂	日风速最 大	0.418**	0.364**
X ₁₁	日最高气 温	0.252**	0.309**	X ₂₃	前日 API	0.385**	0.475**
X ₁₂	前日气温 日均	0.272**	0.305**				

表中标注“**”的代表双侧检验的显著性结果的 Sig 小于 0.01，成强的相关性。标注“*”的代表双侧检验的显著性结果的 Sig 大于 0.01 而小于 0.05，相关性较强。未标注的则表示相关性弱。

对上表的两种相关系数综合分析可知，与春季PM₁₀的 API 值相关性较高的因子有 X₁-本站水汽压、X₂-气压 08h、X₃-气压日均、X₄-气压日最高、X₈-日照日合计、X₉-气温 08h、X₁₀-气温日均、X₁₁-日最高气温、X₁₂-前日气温日均、X₁₄-能见度 08h、X₁₅-能见度 14h、X₂₀-风速 20h、X₂₁-风速日均、X₂₂-日最大风速、X₂₃-前日 PM₁₀ API 值。其中，X₂、X₃、X₄、X₈、X₁₄、X₁₅ 呈负相关性，其他因子显现正相关性。通过相关分析，剔除了 X₅、X₆、X₇、X₁₃、X₁₆、X₁₇、X₁₈、X₁₉，因子由原来的 23 个变为 15 个，这些因子即作为进入春季逐步回归方程和主成分分析的备选因子。可以看出，与春季PM₁₀的

API 值相关性较高的因子有气压因子、气温因子、日照因子、能见度因子、风速因子和 前日 API 值因子，与云量和湿度因子相关性较小，其主要原因是春季多风沙、多寒潮，气候干燥降水少，天气形势多变。

4.4.3 夏季各因子的相关分析

将夏季PM₁₀的 API 与各影响因子做相关分析，得到结果如下表 4-11。

表 4-11 夏季PM₁₀的 API 与各因子相关分析结果

因子	因子含义	Pearson 系数	Spearman 系数	因子	因子含义	Pearson 系数	Spearman 系数
X ₁	本站水汽压	0.268**	0.246**	X ₁₃	前日与当日温差	0.099	0.079
X ₂	气压 08h	0.141*	0.170*	X ₁₄	能见度 08h	-0.320**	-0.263**
X ₃	气压日均	0.137*	0.154*	X ₁₅	能见度 14h	-0.298**	-0.277**
X ₄	气压日最高	0.155*	0.181**	X ₁₆	湿度 08h	-0.031	-0.044
X ₅	前日气压日均	0.310**	0.333**	X ₁₇	相对湿度日均	-0.160*	-0.135*
X ₆	前日气压 08h	0.308**	0.342**	X ₁₈	风速 08h	0.114	0.059
X ₇	总云量	0.265*	0.246*	X ₁₉	风速 14h	0.178	0.203
X ₈	日照日合计	-0.089	-0.065	X ₂₀	风速 20h	0.105	0.196
X ₉	气温 08h	0.398**	0.378**	X ₂₁	风速日均	0.110	0.071
X ₁₀	气温日均	0.489**	0.484**	X ₂₂	日风速最大	0.107	0.143*
X ₁₁	日最高气温	0.448**	0.453**	X ₂₃	前日 API	0.531**	0.516**
X ₁₂	前日气温日均	0.543**	0.550**				

对上表的两种相关系数综合分析可知，与夏季PM₁₀的 API 值相关性较高的因子有 X₁-本站水汽压、X₄-气压日最高、X₅-前日气压日均、X₆-前日气压 08h、X₇-总云量、X₉-气温 08h、X₁₀-气温日均、X₁₁-日最高气温、X₁₂-前日气温日均、X₁₄-能见度 08h，X₁₅-能见度 14h、X₁₇-相对湿度日均、X₂₃-前日 PM₁₀ API 值。其中 X₁₄、X₁₅、X₁₇ 呈负相关性，其他因子呈正相关性。通过相关分析，剔除了 X₂-气压 08h、X₃-气压日均、X₈-日照日合计、X₁₃-温差、X₁₆-湿度 08h、X₁₈-风速 08h、X₁₉-风速 14h、X₂₀-风速 20h、X₂₁-风速日均、X₂₂-日最大风速。因子由原来的 23 个变为个 13 个，这些因子即作为进

入夏季逐步回归方程和主成分分析的备选因子。可见夏季备选因子与春季备选因子的区别在于:(1)夏季气压因子由强负相关变为强正相关。(2)与云量的相关性增大。(3)与日均相对湿度相关性增大,这说明了降水对PM₁₀的 API 值的影响。(4)夏季风速因子对PM₁₀的 API 值影响不大,这也符合大庆市夏秋少风的气候特点。

4. 4. 4 秋季各因子的相关分析

将秋季PM₁₀的 API 与各影响因子做相关分析,得到结果如下表 4-12。

表 4-12 秋季PM₁₀的 API 与各因子相关分析结果

因子	因子含义	Pearson 系 数	Spearman 系数	因子	因子含义	Pearson 系 数	Spearman 系数
X ₁	本站水汽压	0.029	-0.080	X ₁₃	前日与当日温差	-0.230**	-0.332**
X ₂	气压 08h	-0.011	0.070	X ₁₄	能见度 08h	-0.450**	-0.216**
X ₃	气压日均	0.027	0.081	X ₁₅	能见度 14h	-0.456**	-0.157*
X ₄	气压日最高	0.053	0.157*	X ₁₆	湿度 08h	0.255*	0.240*
X ₅	前日气压日均	-0.006	0.333**	X ₁₇	相对湿度日均	0.209**	0.107
X ₆	前日气压 08h	0.224**	0.351**	X ₁₈	风速 08h	-0.145	-0.187
X ₇	总云量	0.210*	0.194	X ₁₉	风速 14h	-0.151	-0.131
X ₈	日照日合计	-0.394**	-0.388**	X ₂₀	风速 20h	0.084	0.111
X ₉	气温 08h	-0.018	-0.115	X ₂₁	风速日均	0.271**	0.131
X ₁₀	气温日均	-0.013	-0.108	X ₂₂	日风速最大	0.192**	0.101
X ₁₁	日最高气温	0.002	-0.087	X ₂₃	前日 API	0.547**	0.439**
X ₁₂	前日气温日均	-0.073	-0.178*				

对上表的两种相关系数综合分析可知,与秋季PM₁₀的 API 值相关性较高的因子有 X₆-前日气压 08h、X₇-总云量、X₈-日照日合计、X₁₃-前日与当日温差、X₁₄-能见度 08h、X₁₅-能见度 14h、X₁₆-相对湿度 08h、X₁₇-相对湿度日均、X₂₁-风速日均、X₂₃-前日PM₁₀的 API 值。剔除了水汽压、气压 08h、气压日均值、气压日最高值、前日气压日均值、气温 08h、气温日均、日最高气温、前日气温日均、风速 08h、风速 14h、风速 20h、日最大风速等 13 个因子,剩下 10 个因子即作为进入秋季逐步回归方程和主成分分析

的备选因子。可见秋季备选因子与之前春夏季的区别有:(1)秋季PM₁₀的 API 值与气压因子相关性并不高, 仅仅与前日气压 08h 相关性较大。(2)与日照因子相关性很强, 说明进入秋季, 日照时间逐渐变短, 不利于PM₁₀污染的稀释扩散。(3)与气温各因子相关性差, 但是与气温变化因子相关性较好。(4)与风速因子相关性差。

4.4.5 冬季各因子的相关分析

将冬季PM₁₀的 API 与各影响因子做相关分析, 得到结果如下表 4-13。

表 4-13 冬季PM₁₀的 API 与各因子相关分析结果

因子	因子含义	Pearson 系数	Spearman 系数	因子	因子含义	Pearson 系数	Spearman 系数
X ₁	本站水汽压	0.149 [*]	0.149 [*]	X ₁₃	前日与当日温差	-0.140 [*]	-0.103
X ₂	气压 08h	0.051	0.060	X ₁₄	能见度 08h	-0.435 ^{**}	-0.367 ^{**}
X ₃	气压日均	0.061	0.068	X ₁₅	能见度 14h	-0.380 ^{**}	-0.383 ^{**}
X ₄	气压日最高	0.067	0.073	X ₁₆	湿度 08h	0.434 ^{**}	0.459 ^{**}
X ₅	前日气压日均	0.172 ^{**}	0.171 ^{**}	X ₁₇	相对湿度日均	0.409 ^{**}	0.414 ^{**}
X ₆	前日气压 08h	0.161 [*]	0.153 [*]	X ₁₈	风速 08h	-0.166	-0.252 [*]
X ₇	总云量	0.054	0.095	X ₁₉	风速 14h	-0.226 [*]	-0.221 [*]
X ₈	日照日合计	-0.387 ^{**}	-0.355 ^{**}	X ₂₀	风速 20h	0.017	0.057
X ₉	气温 08h	-0.016	-0.005	X ₂₁	风速日均	-0.149 ^{**}	-0.209 ^{**}
X ₁₀	气温日均	-0.013	-0.003	X ₂₂	日风速最大	-0.216 ^{**}	-0.264 ^{**}
X ₁₁	日最高气温	0.003	0.029	X ₂₃	前日 API	0.520 ^{**}	0.524 ^{**}
X ₁₂	前日气温日均	-0.098	-0.081				

对上表的两种相关系数综合分析可知, 与冬季PM₁₀的 API 值相关性较高的因子有 X₁-本站水汽压、X₅-前日气压日均值、X₆-前日气压 08h、X₈-日照日合计、X₁₃-温差、X₁₄-能见度 08h、X₁₅-能见度 14h、X₁₆-相对湿度 08h、X₁₇-相对湿度日均、X₁₉-风速 14h、X₂₁-风速日均值、X₂₂-日风速最大值、X₂₃-前日PM₁₀的 API 值。剔除了气压 08h、气压日均、气压日最高、总云量、气温 08h、气温日均、日最高气温、前日气温日均、风速 08h、风速 20h 等 10 个, 剩下 13 个因子即作为进入冬季逐步回归方程和主成分分析的

备选因子。对比冬季备选因子与之前春夏秋冬四季备选因子可知冬季的备选因子有如下特点:(1)冬季 PM_{10} 的 API 值与当日的各气压因子相关性不大,而与前日 08h 气压与前日日均气压相关性较大。(2)进入冬季,日照时间变短,对 PM_{10} 污染影响很大,呈强负相关性,说明不利于污染的扩散稀释。(3)与各气温因子相关性不大,与温差有些许相关性。(4)与相对湿度呈很强相关性,这与大庆市的气候特点有关,大庆市冬季相对湿度很大,多降雪过程,而少量降雪不但不能对 PM_{10} 污染起到清洁作用,反而会吸附空气中的 PM_{10} 而使污染加重,所以相对湿度与 PM_{10} 污染指数呈强的正相关性。(5)与风速因子的相关性由原来的正相关变为负相关。这是由于大庆市冬季积雪覆盖,故风速大时也不能将地表的沙尘吹起,反而起到了加速污染物扩散稀释的作用。

4.5 本章小结

(1)以大庆市近年来的气象要素逐日实测值为数据基础,对大庆市的气压因子、气温因子、相对湿度因子、能见度因子、风速因子等共计 18 个气象因子进行了统计学分析,找出了不同季节各气象因子的变化规律,初步探讨了不同季节各气象因子与 PM_{10} 污染情况之间的关系。

(2)对大庆市各季节 PM_{10} 污染高值日的气象要素值进行了统计学分析。找出了不同季节 PM_{10} 污染高值日各气象因子的特征及变化规律,进一步探讨了四季各气象要素与 PM_{10} 污染之间的关系。

(3)应用 SPSS 软件对大庆市四季 PM_{10} 的 API 值与各气象因子进行了相关分析,找出了与四季 PM_{10} 的 API 值相关性较高的气象因子。并综合统计分析的结果最后确定了与大庆市春季 PM_{10} 污染相关性较强的气象因子 15 个,与大庆市夏季 PM_{10} 污染相关性较强的气象因子 13 个,与大庆市秋季和冬季 PM_{10} 污染相关性较强的气象因子分别为 10 个和 13 个。这些因子将作为建立 PM_{10} 污染预报模型的备选因子。

第五章 PM_{10} 污染数学预报模型的建立

5.1 预报因子的选取

通过第四章对大庆市 PM_{10} 污染与各影响因子的相关分析可以得出, 进入大庆市春季 PM_{10} 污染预报模型的主要影响因子有气压 08h、气压日均、日照日合计、气温日均、日最高气温、能见度 14h、风速日均、日最大风速、前日 PM_{10} 的 API 值等 15 个。进入夏季模型的备选因子有前日气压日均、总云量、气温 08h、气温日均、日最高气温、前日气温日均、能见度 08h、能见度 14h、前日 PM_{10} API 值等 13 个。进入秋季模型的备选因子有前日气压 08h、前日与当日温差、能见度 08h、能见度 14h、风速日均、前日 PM_{10} 的 API 值等 10 个。进入冬季的备选因子有前日气压日均值、日照日合计、温差、能见度 08h、能见度 14h、相对湿度 08h、相对湿度日均、风速日均值、前日 PM_{10} 的 API 值等 13 个。

这些预报因子的选择能够较好的反映出大庆市 PM_{10} 污染的情况, 但是由于有些因子在实际建模的时候并没有发挥实际作用, 反而会给建模过程带来问题。具体表现在以下两点: (1) 备选因子间通常或多或少存在共线性, 这种高度的重叠与相关会给建模过程带来障碍。如在建立多元回归方程的时候, 变量间的高度共线性会使所建回归方程参数不准确甚至建模失败。(2) 在建模过程中如果涉及到过多的因子会导致计算量的增大。这就需要在建模之前对这些备选因子进行筛选, 应用逐步回归方法就可以很好的解决这个问题, 其基本原理已在第二章中介绍。逐步回归分析, 首先要建立因变量 y 与自变量 x 之间的总回归方程, 再对总方程及每个自变量进行假设检验, 总方程不显著时, 表明该多元回归方程不成立; 而当某个自变量对 y 的方程不显著, 则把它剔除, 重新建立不包含该变量的新方程, 最后, 筛选出最显著的影响因子而建立最优化回归方程, 以达到精选因子的目的。

5.2 四季多元回归预报模型的建立与检验

依据 2.4.2 节阐述的逐步回归方法对备选因子进行精选并对大庆市 PM_{10} 污染的 API 值分四季建立多元回归预报模型, 对模型进行显著性检验和预报准确率检验并且绘制出实测值与预报值的拟合曲线。基于的数据长度为春季 276 天、夏季 214 天、秋季 182 天、冬季 240 天。

5.2.1 春季预报模型的建立及检验

通过精选因子的过程得出春季最优方程情况下的主要影响因子有 X_1 -气温 08h、 X_2 -能见度 08h、 X_3 -能见度 14h、 X_4 -日最大风速、 X_5 -前日 PM_{10} 的 API 值五个。所建立的标准化的回归方程如下：

$$Y(\text{春})=0.106 \cdot X_1-0.209 \cdot X_2-0.196 \cdot X_3+0.213 \cdot X_4+0.188 \cdot X_5-2.9E-015$$

模型建立好之后，需要对其进行检验，具体包括两种检验。其一是对整个方程的 F 检验，用以检验所建立回归方程的显著性，具体方法已在 2.5.3 节中阐述。其二是对所建预报方程预报的准确率检验，以确定模型的可用性。对春季回归方程的 F 检验结果见表 5-1。

表 5-1 春季回归方程的 F 检验表

季节	自由度(m,n-m-1)	F _{0.05} 值	F _{0.05} 计算值	相关系数
春季	(5,270)	2.240	32.833	0.367

由上表可知，春季的 $F_{0.05}$ 计算值=32.833> $F_{0.05}$ 值=2.240，可知方程的回归效果高度显著，可以作为预报方程。

对春季回归方程的预报准确率检验包括以下两个方面：

(1) 预报的 API 数值准确率的检验, 公式为：

预报 API 数值准确率=1-(预测值-实测值)/实测值 (5-1)

(2) 预报污染级别正确性评分和预报精确度评分。按照《气象部门空气质量预报考核和管理暂行办法》规定，分别用以下两公式对预报级别正确性评分和预报精确评分进行计算。

预报级别的正确性评分：

$$f = \begin{cases} 100(\text{预报等级与日报相同}) \\ 50 (\text{预报等级与日报不同}) \end{cases} \quad (5-2)$$

预报的精确度评分：

$$H = \left(1 - \frac{|y_i - x_i|}{\text{Max}(y_i, x_i)}\right) * 100 \quad (5-3)$$

式中 y_i 为 API 的预报值， x_i 为 API 的实测值

通过对春季 276 日的实测值与预报值的对比，得出所建立的春季回归模型对大庆市春季 PM_{10} 的 API 数值预报的平均准确率为 73.9%。样本数据中的所有实测值与预测

值的拟合曲线如下图 5-1。

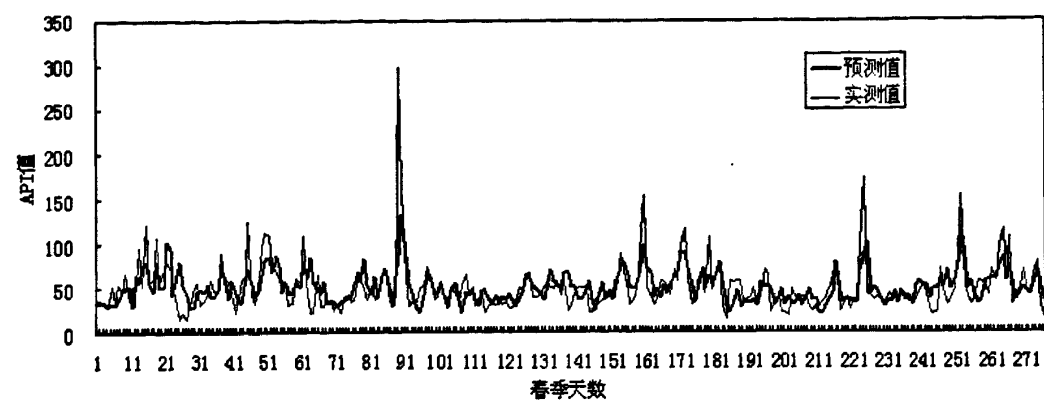


图 5-1 春季PM₁₀的 API 值实测值与预测值拟合曲线图

由图可见，预测值与实测值拟合程度较好，大体趋势拟合准确，但是对突发情况如 2008 年 5 月 27 日的沙尘暴天气在数值预测方面较差。计算得该春季模型的平均预报级别正确性评分为 84.8 分，预报的精确度评分为 76.8 分。可以满足实际的预报工作需要。

5. 2. 2 夏季预报模型的建立及检验

通过精选因子的过程得出进入夏季最优方程的因子有 X_1 -前日气压日均值、 X_2 -前日气温日均值、 X_3 -能见度 08h、 X_4 -日最高气温、 X_5 -前日 PM₁₀ 的 API 值。所建立的夏季的标准化的回归方程如下：

$$Y(\text{夏})=0.181 \cdot X_1+0.179 \cdot X_2-0.206 \cdot X_3+0.195 \cdot X_4+0.309 \cdot X_5-9.1E-015$$

对夏季回归方程的 F 检验结果见表 5-2。

表 5-2 夏季回归方程的 F 检验表

季节	自由度(m,n-m-1)	F _{0.05} 值	F _{0.05} 计算值	相关系数
夏季	(5,208)	2.260	36.992	0.471

由上表可知，夏季的 F_{0.05} 计算值=36.992> F_{0.05} 值=2.260，可知所建立的夏季回归方程回归效果显著，可作为预报方程。

对夏季回归方程的预报准确率的检验如下，通过对夏季样本中 214 日的实测值与预测值对比，得出所建立的夏季预报模型对大庆市夏季 PM₁₀ 的 API 数值预报的平均准确率为 77.6%。样本数据中的所有实测值与预测值的拟合曲线如下图 5-2 所示。

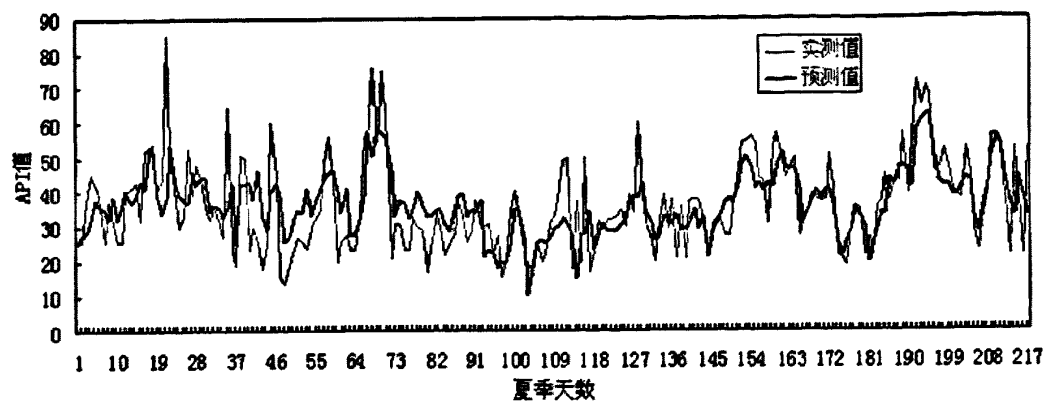


图 5-2 夏季PM₁₀的 API 值实测值与预测值拟合曲线图

由图 5-2 可见，夏季的预测值与实测值拟合程度较好，大体趋势拟合较为准确。个别波峰与波谷数值上拟合相对较差。计算本模型的平均预报级别正确性评分为 93.3 分，预报模型的精确度评分为 82.3 分，可以满足实际预报工作的需要。

5.2.3 秋季预报模型的建立及检验

通过精选因子的过程得出入选秋季最优方程的因子有 X₁-前日 API 值、X₂-前日与当日气温差、X₃-能见度 14h、X₄-前日气压 08h、X₅-风速日均等 5 个。所建立的回归方程如下：

$$Y(\text{秋})=0.413 \cdot X_1-0.312 \cdot X_2-0.310 \cdot X_3+0.313 \cdot X_4+0.231 \cdot X_5-0.066$$

对秋季回归方程的 F 检验结果见表 5-3。

表 5-3 秋季回归方程的 F 检验表

季节	自由度(m,n-m-1)	F _{0.05} 值	F _{0.05} 计算值	相关系数
秋季	(5,176)	2.270	22.078	0.565

由上表可知，秋季的 F_{0.05} 计算值=22.078> F_{0.05} 值=2.270，可知方程的回归效果显著，可以作为预报方程。

对秋季回归方程的预报准确率的检验如下，通过对秋季样本中 182 日的实测值与预测值对比，得出所建立的秋季预报模型对大庆市秋季PM₁₀的 API 数值预报的平均准确率为 77%。样本数据中的所有实测值与预测值的拟合曲线如图 5-3 所示。

由图 5-3 可以看出，秋季模型的实际值与预测值之间拟合度很高，对某些 API 高值日的预测也表现良好。经计算本模型的平均预报级别正确性评分为 91.8 分，预报模型的精确度评分为 80.7 分，可以满足实际预报工作的需要。

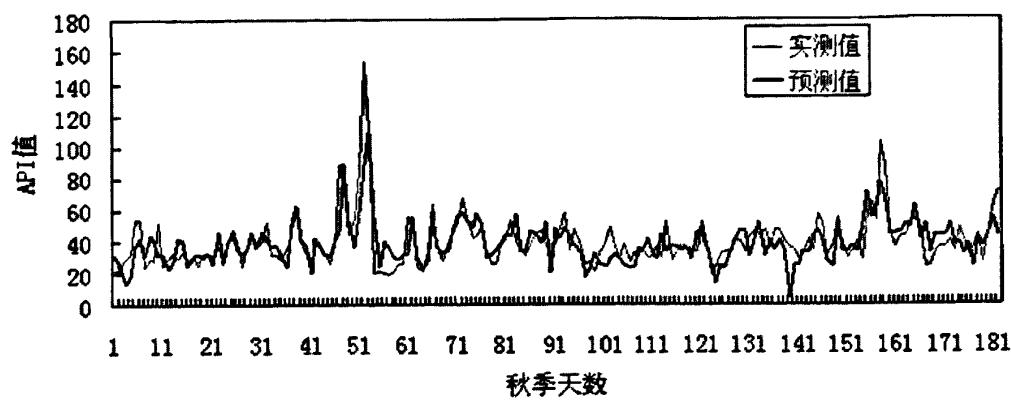


图 5-3 秋季PM₁₀的 API 值实测值与预测值拟合曲线图

5. 2. 4 冬季预报模型的建立及检验

通过精选因子的过程得出进入冬季最优方程的因子有 X_1 -前日气压日均值、 X_2 -日照日合计、 X_3 -前日与当日气温差、 X_4 -能见度 08h、 X_5 -相对湿度 08h、 X_6 -前日 PM₁₀ 的 API 值。建立的冬季回归方程如下：

$$Y(\text{冬})=0.101 * X_1-0.130 * X_2-0.152 * X_3-0.253 * X_4+0.196 * X_5+0.407 * X_6-5.3E-015$$

对冬季的回归模型的 F 检验结果见表 5-4。

表 5-4 冬季回归方程的 F 检验表

季节	自由度(m,n-m-1)	F _{0.05} 值	F _{0.05} 计算值	相关系数
冬季	(6, 234)	2.260	38.137	0.495

由上表可知，冬季方程的 F_{0.05} 计算值=38.137> F_{0.05} 值=2.260，由此可知方程的回归效果显著，可作为预报方程。

对冬季回归方程的预报准确率的检验如下，通过对冬季样本中 240 日的实测值与预测值对比，得出所建立的冬季预报模型对大庆市冬季 PM₁₀ 的 API 数值预报的平均准确率为 79.7%。样本数据中的所有实测值与预测值的拟合曲线如图 5-4 所示。

由图 5-4 可知，该模型对冬季 PM₁₀ 的 API 的预测值与实测值的拟合度很好，趋势拟合准确。对 API 高值日和低值日的预测都较为准确。经计算本模型的平均预报级别正确性评分为 87.7 分，预报模型的精确度评分为 83.3 分，可以满足实际预报工作的需要。

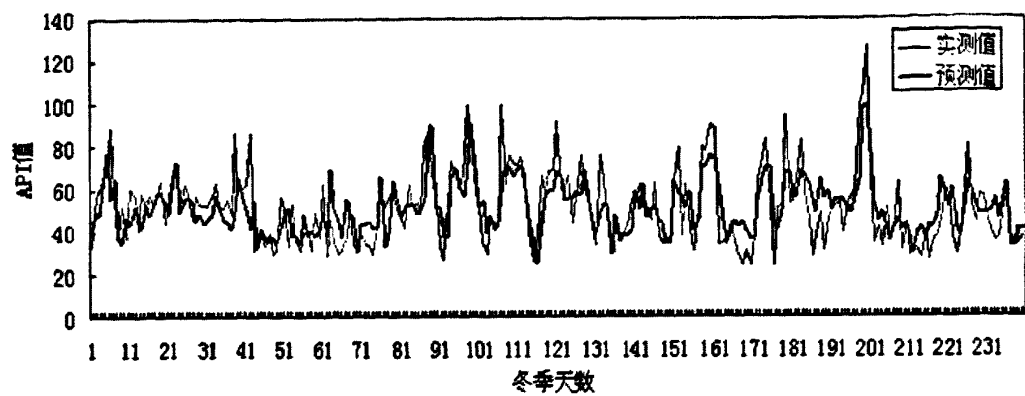


图 5-4 冬季PM₁₀的 API 值实测值与预测值拟合曲线图

5. 2. 5 四季模型对比

通过以上建模分析可以看出，春季预报的 API 平均准确率为 73.9%，模型的预报级别正确性评分和模型精确度评分分别为 84.8 分和 76.8 分。夏季预报的 API 平均准确率为 77.6%，模型预报级别正确性评分和模型精确度评分分别为 93.3 和 82.3 分。秋季模型预报的 API 平均准确率为 77%，模型预报级别正确性评分和模型精确度评分分别为 91.8 分和 80.7 分。冬季预报的 API 平均准确率为 79.7%，模型的预报级别正确性评分和模型精确度评分分别为 87.7 分和 83.3 分。故在预报准确率方面冬季模型>夏季模型>秋季模型>春季模型。其原因为冬季的天气形势比较稳定，而春天天气多变并且易发沙尘暴等突发的天气情况，故所建春季模型难以对突发形势作出预测。对于夏季模型，由于本论文没有考虑降水因子，故对有降水当天的 API 预测的误差较大，若能加入降水因子，则模型会更加完善。

5. 3 本章小结

(1)应用逐步回归方法对备选因子进行了精选，选出了进入春季回归模型的预报因子 5 个、进入夏季模型的预报因子 5 个，进入秋季和冬季模型的预报因子分别为 5 个和 6 个。

(2)应用以上预报因子分四季建立了大庆市 PM₁₀污染的标准化回归模型

$$Y(\text{春})=0.106 \cdot X_1-0.209 \cdot X_2-0.196 \cdot X_3+0.213 \cdot X_4+0.188 \cdot X_5-2.9\text{E-}015$$
$$Y(\text{夏})=0.181 \cdot X_1+0.179 \cdot X_2-0.206 \cdot X_3+0.195 \cdot X_4+0.309 \cdot X_5-9.1\text{E-}015$$
$$Y(\text{秋})=0.413 \cdot X_1-0.312 \cdot X_2-0.310 \cdot X_3+0.313 \cdot X_4+0.231 \cdot X_5-0.066$$
$$Y(\text{冬})=0.101 \cdot X_1-0.130 \cdot X_2-0.152 \cdot X_3-0.253 \cdot X_4+0.196 \cdot X_5+0.407 \cdot X_6-5.3\text{E-}015$$

(3) 对四季回归预报模型进行了预报准确率的检验，并进行了预报级别正确性评分和模型精确度评分。得出了模型预报准确率分别为 73.9%、77.6%、77%、79.7%，平均预报准确率为 77.1%。预报级别正确性评分分别为 84.8、93.3、91.8、87.7。模型精确度评分分别 76.8、82.3、80.7、83.3。所建立模型效果理想，能达到实际预报工作的需要。

第六章 PM₁₀ 污染神经网络预报模型的建立

第五章建立了基于多元回归分析的大庆市四季PM₁₀污染预报的数学模型,取得了平均 77%的预报准确率。而实际建模中各相关因子与PM₁₀的 API 值之间有时并不存在明显的线性关系,而更趋向与非线性问题。人工神经网络的工作原理是基于对人脑非线性思维的模拟,因此,其更适用于解决非线性问题。本论文要解决的是对PM₁₀污染的预报,需利用大量的历史数据对网络进行训练,从而使神经网络能够掌握从输入空间向输出空间的映射关系,这属于有导师的学习方式,而在有导师学习方式的神经网络中,B-P 神经网络非常经典且应用广泛。故本章将建立大庆市四季PM₁₀污染预报的B-P 神经网络模型并对其进行预报准确率检验,同时与第五章建立的多元回归模型进行对比分析。

6.1 模型建立的理论依据

关于人工神经网络和 B-P 神经网络的基本原理已在第二章中有过详细介绍。B-P 神经网络设计过程主要有已下三个方面过程:

(1)训练样本的组织:包括输入层样本的选择与预处理过程。

(2)B-P 网络拓扑结构的设计:包括输入层和输出层节点数的选取,隐含层数的确定、隐含层节点数的选取、各层神经元节点之间传递函数的选取等。

(3)B-P 神经网络主要参数的确定:包括最大循环次数、动量常数、学习速率等参数的确定。

6.2 数据准备及预报因子的精选

6.2.1 数据的准备及预处理

首先要建立 B-P 建模过程的样本数据库,其中包括与大庆市PM₁₀相关程度高的各气象因子和特征因子的逐日资料以及与之对应的同期PM₁₀的逐日 API 值,数据库中每条记录为当日的各气象因子值、特征因子值和PM₁₀的 API 值。数据库的时间跨度为2008年1月1日至2010年6月30日,分四季建成四个样本数据集。其中春季样本集包括276条记录,夏季样本集包括214条记录,秋季样本集包括182条记录,冬季样本集包括240条记录。

由于样本数据库中涉及到多个气象因子与特征因子,这些数据的数量级、量纲有很大的差异,这样在开始进行数据训练之前需要对输入样本进行归一化处理。具体可

分为把数据归一化到[0,1]区间, 或把数据归一化到[-1,1]区间, 或把数据归一化为零均值和单位方差三种。本文采用的方法是把数据归一化到[-1,1]区间, 具体方法如下。

设 $x_{\max}$ 和 $x_{\min}$ 分别为原始数据中的最大值和最小值, x_i 为实际数据, $\hat{x}_i$ 为归一化后的数据, 归一化公式为:

$$\hat{x}_i = \frac{x_i - 0.5(x_{\max} + x_{\min})}{0.5(x_{\max} - x_{\min})} \quad (6-1)$$

使神经网络输入输出数据位于[-1,1]区间, 最后使用下式把归一化后的数值重新换算回真实值。

$$x_i = 0.5(x_{\max} - x_{\min})\hat{x}_i + 0.5(x_{\max} + x_{\min}) \quad (6-2)$$

6.2.2 建模因子的初选

在进行 B-P 神经网络建模之前, 需要确定进入 B-P 网络输入层的因子。同第五章建立多元回归模型时相同, B-P 网络建模的因子选取的过程也分为因子初选与因子精选的过程, 前文 4.4 节中已通过相关分析确定了进入建模过程的初选因子如下, 进入春季模型的备选因子有 X_1 -本站水汽压、 X_2 -气压 08h、 X_3 -气压日均、 X_4 -气压日最高、 X_5 -日照日合计、 X_6 -气温 08h、 X_7 -气温日均、 X_8 -日最高气温、 X_9 -前日气温日均、 X_{10} -能见度 08h、 X_{11} -能见度 14h、 X_{12} -风速 20h、 X_{13} -风速日均、 X_{14} -日最大风速、 X_{15} -前日 PM₁₀ 的 API 值等 15 个。进入夏季模型的备选因子有 X_1 -本站水汽压、 X_2 -气压日最高、 X_3 -前日气压日均、 X_4 -前日气压 08h、 X_5 -总云量、 X_6 -气温 08h、 X_7 -气温日均、 X_8 -日最高气温、 X_9 -前日气温日均、 X_{10} -能见度 08h、 X_{11} -能见度 14h、 X_{12} -相对湿度日均、 X_{13} -前日 PM₁₀ API 值等 13 个。进入秋季模型的备选因子有 X_1 -前日气压 08h、 X_2 -总云量、 X_3 -日照日合计、 X_4 -前日与当日温差、 X_5 -能见度 08h、 X_6 -能见度 14h、 X_7 -相对湿度 08h、 X_8 -相对湿度日均、 X_9 -风速日均、 X_{10} -前日 PM₁₀ 的 API 值等 10 个。进入冬季的备选因子有 X_1 -本站水汽压、 X_2 -前日气压日均值、 X_3 -前日气压 08h、 X_4 -日照日合计、 X_5 -温差、 X_6 -能见度 08h、 X_7 -能见度 14h、 X_8 -相对湿度 08h、 X_9 -相对湿度日均、 X_{10} -风速 14h、 X_{11} -风速日均值、 X_{12} -日风速最大值、 X_{13} -前日 PM₁₀ 的 API 值等 13 个。以上因子将作为主成分分析法精选因子过程的备选因子。

6.2.3 基于主成分分析法的因子精选

主成分分析是研究如何通过原来变量的少数几个线性组合来解释原来变量绝大多

数信息的一种多元统计方法，具体原理在第二章中已有详细阐述。由于参与建模过程的初选因子之间有一定的相关性，这样就必然存在着起支配作用的共同因素，因此，通过研究原始变量相关矩阵或协方差矩阵的内部结构关系，可以利用原始变量的线性组合形成几个综合性的指标(即主成分)，这样就可以在保留原始变量大部分信息的同时对备选因子进行降维减噪，而达到精选因子的目的。

本论文应用 SPSS 软件进行主成分分析过程，对大庆市PM₁₀污染预报建模过程分四季进行因子精选。

对大庆市春季的 15 个备选因子进行主成分分析结果如下表 6-1。

表6-1 春季因子解释原有变量总方差情况

因子编号	初始因子解			因子提取后的解		
	特征值	方差贡献率	累计方差贡献率	特征值	方差贡献率	累计方差贡献率
1	6.239	41.591	41.591	6.239	41.591	41.591
2	3.194	21.296	62.886	3.194	21.296	62.886
3	1.703	11.354	74.241	1.703	11.354	74.241
4	1.207	8.044	82.285	1.207	8.044	82.285
5	0.846	5.638	87.923	0.846	5.638	87.923
6	0.488	3.251	91.174	0.488	3.251	91.174

由上表可知，前5个主成分的累计贡献率就已经达到了87.923%，可以认为原变量丢失信息较少，主成分分析效果较好。春季5个主成分对应的特征向量如下表6-2所示。

表6-2 春季主成分对应特征向量

变量	F1	F2	F3	F4	F5
X ₁	0.320	0.039	-0.221	0.223	-0.120
X ₂	-0.323	0.140	0.265	0.343	-0.075
X ₃	-0.333	0.109	0.225	0.358	-0.085
X ₄	-0.330	0.063	0.289	0.316	-0.140
X ₅	-0.105	0.434	0.114	-0.300	0.260
X ₆	0.323	0.283	0.024	0.188	-0.127
X ₇	0.322	0.302	0.073	0.138	-0.087
X ₈	0.284	0.339	0.172	0.113	-0.047
X ₉	0.329	0.208	0.008	0.272	-0.037
X ₁₀	-0.149	0.411	0.103	-0.228	0.136
X ₁₁	-0.119	0.427	-0.068	-0.255	0.197

表 6-2 春季主成分对应特征向量 (续)

变量	F1	F2	F3	F4	F5
X ₁₂	0.142	-0.111	0.552	-0.125	-0.172
X ₁₃	0.178	-0.185	0.509	-0.250	0.023
X ₁₄	0.255	-0.162	0.330	-0.154	0.127
X ₁₅	0.107	-0.147	0.094	0.393	0.869

由上表特征向量值可依次写出5个主成分的表达式:

$$F_1=0.320 * X_1-0.323 * X_2-0.333 * X_3-0.330 * X_4-0.105 * X_5+0.323 * X_6+0.322 * X_7+0.284 * X_8+0.329 * X_9-0.149 * X_{10}-0.119 * X_{11}+0.142 * X_{12}+0.178 * X_{13}+0.255 * X_{14}+0.107 * X_{15}$$

$$F_2=0.039 * X_1+0.140 * X_2+0.109 * X_3+0.063 * X_4+0.434 * X_5+0.283 * X_6+0.302 * X_7+0.339 * X_8+0.208 * X_9+0.411 * X_{10}+0.427 * X_{11}-0.111 * X_{12}-0.185 * X_{13}-0.162 * X_{14}-0.147 * X_{15}$$

$$F_3=-0.221 * X_1+0.265 * X_2+0.225 * X_3+0.289 * X_4+0.114 * X_5+0.024 * X_6+0.073 * X_7+0.172 * X_8+0.008 * X_9+0.103 * X_{10}-0.068 * X_{11}+0.552 * X_{12}+0.509 * X_{13}+0.330 * X_{14}+0.094 * X_{15}$$

$$F_4=0.223 * X_1+0.343 * X_2+0.358 * X_3+0.316 * X_4-0.300 * X_5+0.188 * X_6+0.138 * X_7+0.113 * X_8+0.272 * X_9-0.228 * X_{10}-0.255 * X_{11}-0.125 * X_{12}-0.250 * X_{13}-0.154 * X_{14}+0.393 * X_{15}$$

$$F_5=-0.120 * X_1-0.075 * X_2-0.085 * X_3-0.140 * X_4+0.260 * X_5-0.127 * X_6-0.087 * X_7-0.047 * X_8-0.037 * X_9+0.136 * X_{10}+0.197 * X_{11}-0.172 * X_{12}+0.023 * X_{13}+0.127 * X_{14}+0.869 * X_{15}$$

故春季选择这5个主成分构造新的矩阵作为B-P神经网络的新的输入变量。由此可见,主成分分析很好的实现了输入层因子的降维处理,将原来的15个备选因子降维至5个新的输入变量,大大减少了建模的输入数据。

对大庆市夏季13个备选因子进行主成分分析结果如表6-3。

由表6-3可知,前5个主成分的累计贡献率就达到了90.37%,已经能够很好的解释原变量,主成分分析效果良好。

表6-3 夏季因子解释原有变量总方差情况

因子编号	初始因子解			因子提取后的解		
	特征值	方差贡献率	累计方差贡献率	特征值	方差贡献率	累计方差贡献率
1	4.416	33.972	33.972	4.416	33.972	33.972
2	3.827	29.440	63.412	3.827	29.440	63.412
3	1.929	14.840	78.252	1.929	14.840	78.252
4	0.914	7.028	85.280	0.914	7.028	85.280
5	0.662	5.094	90.373	0.662	5.094	90.373
6	0.324	2.490	92.864	0.324	2.490	92.864

夏季各主成分的特征向量表略，此处只给出依据特征变量写出的夏季各主成分表达式如下。

$F_1=-0.245 * X_1+0.196 * X_2+0.187 * X_3+0.186 * X_4-0.377 * X_5+0.225 * X_6+0.244 * X_7+0.338 * X_8+0.090 * X_9+0.354 * X_{10}+0.371 * X_{11}-0.425 * X_{12}-0.121 * X_{13}$

$F_2=0.348 * X_1-0.320 * X_2-0.296 * X_3-0.266 * X_4+0.040 * X_5+0.395 * X_6+0.417 * X_7+0.319 * X_8+0.404 * X_9-0.080 * X_{10}-0.019 * X_{11}+0.043 * X_{12}+0.116 * X_{13}$

$F_3=0.155 * X_1+0.311 * X_2+0.494 * X_3+0.506 * X_4+0.254 * X_5+0.184 * X_6+0.127 * X_7+0.109 * X_8+0.241 * X_9-0.226 * X_{10}-0.239 * X_{11}+0.099 * X_{12}+0.275 * X_{13}$

$F_4=0.340 * X_1+0.097 * X_2+0.117 * X_3+0.088 * X_4+0.050 * X_5+0.143 * X_6+0.087 * X_7+0.016 * X_8-0.100 * X_9+0.007 * X_{10}+0.007 * X_{11}+0.270 * X_{12}-0.860 * X_{13}$

$F_5=0.354 * X_1-0.296 * X_2+0.146 * X_3+0.193 * X_4+0.278 * X_5-0.097 * X_6-0.065 * X_7-0.102 * X_8+0.042 * X_9+0.568 * X_{10}+0.485 * X_{11}+0.300 * X_{12}+0.197 * X_{13}$

故夏季选择这5个主成分构造的矩阵作为B-P神经网络的新的输入变量。由此可见，主成分分析对夏季13个备选因子也很好的实现了降维处理，降至5个输入层节点。

对大庆市秋季10个备选因子进行主成分分析结果如下表6-4。

表6-4 秋季因子解释原有变量总方差情况

因子编号	初始因子解			因子提取后的解		
	特征值	方差贡献率	累计方差贡献率	特征值	方差贡献率	累计方差贡献率
1	3.900	39.004	39.004	3.900	39.004	39.004
2	1.654	16.543	55.547	1.654	16.543	55.547
3	1.198	11.984	67.531	1.198	11.984	67.531

表 6-4 秋季因子解释原有变量总方差情况（续）

因子编号	初始因子解			因子提取后的解		
	特征值	方差贡献率	累计方差贡献率	特征值	方差贡献率	累计方差贡献率
4	0.829	8.287	75.818	0.829	8.287	75.818
5	0.722	7.220	83.039	0.722	7.220	83.039
6	0.577	5.770	88.808	0.577	5.770	88.808
7	0.443	4.427	93.235	0.443	4.427	93.235

由上表可知，前6个主成分的累计贡献率就达到了88.808%，已经能够很好的解释原变量，主成分分析效果良好。秋季各主成分的特征向量表略，此处只给出依据特征变量写出的秋季各主成分表达式如下。

$$F_1=0.158 * X_1+0.373 * X_2-0.422 * X_3-0.094 * X_4-0.421 * X_5-0.379 * X_6+0.310 * X_7+0.369 * X_8+0.137 * X_9+0.276 * X_{10}$$

$$F_2=-0.457 * X_1-0.041 * X_2-0.012 * X_3+0.596 * X_4-0.033 * X_5-0.149 * X_6-0.218 * X_7-0.083 * X_8+0.468 * X_9+0.369 * X_{10}$$

$$F_3=-0.445 * X_1-0.132 * X_2+0.133 * X_3+0.082 * X_4+0.109 * X_5+0.248 * X_6+0.591 * X_7+0.523 * X_8+0.095 * X_9-0.237 * X_{10}$$

$$F_4=0.193 * X_1-0.377 * X_2+0.221 * X_3+0.323 * X_4+0.036 * X_5-0.200 * X_6+0.171 * X_7+0.179 * X_8-0.578 * X_9+0.481 * X_{10}$$

$$F_5=0.314 * X_1-0.558 * X_2+0.291 * X_3-0.305 * X_4-0.169 * X_5-0.088 * X_6+0.160 * X_7-0.047 * X_8+0.574 * X_9+0.134 * X_{10}$$

$$F_6=0.592 * X_1-0.014 * X_2-0.345 * X_3+0.499 * X_4+0.261 * X_5+0.367 * X_6+0.143 * X_7+0.037 * X_8+0.219 * X_9-0.095 * X_{10}$$

故秋季选择这6个主成分构造的矩阵作为B-P神经网络的新的输入变量。由此可见，主成分分析对秋季10个备选因子很好的实现了降维处理，降至6个输入层节点。

对大庆市冬季13个备选因子进行主成分分析结果如表6-5。

由表6-5可知，前6个主成分的累计贡献率就达到了87.182%，已经能够很好的解释原变量，主成分分析效果良好。

表6-5 冬季因子解释原有变量总方差情况

因子编号	初始因子解			因子提取后的解		
	特征值	方差贡献率	累计方差贡献率	特征值	方差贡献率	累计方差贡献率
1	3.672	28.243	28.243	3.672	28.243	28.243
2	3.311	25.469	53.712	3.311	25.469	53.712
3	1.636	12.585	66.297	1.636	12.585	66.297
4	1.109	8.533	74.830	1.109	8.533	74.830
5	0.863	6.636	81.466	0.863	6.636	81.466
6	0.743	5.715	87.182	0.743	5.715	87.182
7	0.502	3.863	91.045	0.502	3.863	91.045

冬季各主成分的特征向量表略，此处只给出依据特征变量写出的冬季各主成分表达式如下。

$$F_1=-0.341 * X_1+0.366 * X_2+0.373 * X_3+0.339 * X_4-0.145 * X_5+0.267 * X_6+0.353 * X_7-0.311 * X_8-0.323 * X_9-0.088 * X_{10}-0.170 * X_{11}-0.204 * X_{12}-0.008 * X_{13}$$

$$F_2=0.101 * X_1-0.217 * X_2-0.214 * X_3+0.153 * X_4+0.025 * X_5+0.250 * X_6+0.158 * X_7-0.323 * X_8-0.355 * X_9+0.400 * X_{10}+0.440 * X_{11}+0.407 * X_{12}-0.188 * X_{13}$$

$$F_3=-0.226 * X_1-0.358 * X_2-0.377 * X_3+0.229 * X_4+0.585 * X_5+0.303 * X_6+0.127 * X_7+0.007 * X_8+0.023 * X_9-0.149 * X_{10}-0.128 * X_{11}-0.256 * X_{12}+0.267 * X_{13}$$

$$F_4=-0.437 * X_1+0.173 * X_2+0.183 * X_3-0.182 * X_4+0.277 * X_5-0.234 * X_6-0.405 * X_7-0.197 * X_8-0.138 * X_9+0.296 * X_{10}+0.228 * X_{11}+0.050 * X_{12}+0.466 * X_{13}$$

$$F_5=0.269 * X_1+0.009 * X_2+0.018 * X_3+0.280 * X_4-0.299 * X_5+0.003 * X_6+0.305 * X_7+0.219 * X_8+0.143 * X_9+0.245 * X_{10}+0.089 * X_{11}+0.037 * X_{12}+0.727 * X_{13}$$

$$F_6=-0.203 * X_1+0.218 * X_2+0.207 * X_3-0.190 * X_4+0.173 * X_5+0.528 * X_6+0.109 * X_7+0.430 * X_8+0.370 * X_9+0.371 * X_{10}+0.126 * X_{11}+0.050 * X_{12}-0.182 * X_{13}$$

故冬季选择这6个主成分构造的矩阵作为B-P神经网络的新的输入变量。由此可见，主成分分析将冬季13个备选因子降至6个，降维效果良好。

6.3 B-P 神经网络预报模型的设计

B-P 神经网络模型的设计主要包括对网络拓扑结构设计及网络参数确定两部分,其中拓扑结构的设计过程主要包括隐含层数的确定、输入层与输出层节点的选取、隐含层节点数的选取、各层神经元节点间传递函数的选取等。对神经网络主要参数的确定

主要包括对最大训练次数、动量常数、学习速率等参数的确定。

6.3.1 网络层数的确定

在神经网络建模过程中,输入层与输出层是必备的,故网络层数的确定实际上就是确定隐含层数,隐含层数的确定直接关系到网络性能的优劣。关于 B-P 神经网络隐含层的确定,许多学者进行了大量的理论研究。Robert Hecht Nielson^[77]研究得出,对于任意闭区间上的任何连续函数,都可以用 1 个隐含层的 BP 神经网络来逼近,即一个三层的 B-P 神经网络可以完成任意的 N 维到 M 维的映射。Lippmann^[78]研究表明,只要两个隐含层,就能解决任何形式的分类问题。后来 Robert 等人进一步指出,有一个隐含层的网络,只要节点足够多,就可以以任意精度逼近一个非线性函数。在实际应用中,虽然多隐含层的 B-P 网络有着更快的训练速度,但在误差反算的过程中就会由于隐含层的增加而变的复杂起来,使训练时间变长。同时,由于隐含层的增加,也大大增加了网络陷入局部极小值的概率。因此,本模型的隐含层数选为 1 个。即网络层数共三层,分别为一个输入层、一个隐含层、一个输出层。

6.3.2 输入输出层节点的确定

输入层的节点数由所建模型需要解决的实际问题本身而决定的。对于本论文来说,四个季节分别建立的 B-P 模型的输入层节点是不同的,每个季节模型的输入层节点数即为通过主成分分析精选因子得出的主成分数,分别为春季节点数 5 个,夏季节点数 5 个,秋季节点数 6 个,冬季节点数 6 个。各主成分因子构造的矩阵即为 B-P 神经网络的输入样本。建模的主要目的是为了预报大庆市 PM_{10} 的 API 值,故模型输出层节点数为 1,即为 PM_{10} 的 API 值。

6.3.3 隐含层节点数的确定

B-P 人工神经网络建模时,隐含层节点数的确定至关重要,若隐含层节点选取过多,会导致网络训练时间过长。而隐含层节点数选择过少,会导致网络的容错性差,识别未经训练的样本数据的能力偏低,网络的训练精度不够。因此,在选择隐含层节点数时,应综合考虑,这往往也是 B-P 建模过程中的一个难题,迄今还没有一种统一的方法。本文使用试错法来确定最优的隐含层节点数,首先从较小的节点数起进行网络的学习并且检验网络性能,然后增加节点数重复学习检验,直至找到网络最优时的节点数。通过不断对比试错,本文最终确定的隐含层节点数为春季模型 8 个,夏季模

型 9 个, 秋季模型 9 个, 冬季模型 9 个。即对四季分别构造了结构为 5-8-1、5-9-1、6-9-1、6-9-1 的 B-P 神经网络模型。

6.3.4 各层传递函数的确定

由于 B-P 神经网络的非线性逼近能力是通过 S 型传递函数体现出的, 故输入层与隐含层之间、隐含层与输出层之间的传递函数均选用 Sigmoid 传递函数, 具体的表达式见本文章节 2.1.2。

6.3.5 B-P 神经网络主要参数的确定

B-P 神经网络的主要参数主要包括训练次数、学习速率、学习动量常数等。其中, 学习速率是网络权重更新的加权因子。学习速率设得太小, 网络收敛速度慢, 而且会收敛于局部最小点, 学习速率设得太大, 可能会导致网络不稳定。在此模型中学习速率经过反复验算设定为 0.01。学习动量常数是权重更新量与上次权重更新量之间的关系, 此模型中经过反复调整验算设为 0.9。训练次数对于不同模型来说不同, 此处设定一最大训练次数 5000。

6.4 基于 TIPDM 的 B-P 网络四季预报模型的建立与检验

本论文采用 TIPDM 数据挖掘软件来建立大庆市 PM₁₀ 污染的四季预报模型, 具体操作流程已在第二章中阐述。所基于的四季样本数据集分别为春季样本集 276 条记录, 夏季样本集 214 条记录, 秋季样本集 182 条记录, 冬季样本集 240 条记录。将这些样本数据分为训练样本和测试样本, 按照 2:1 的比例来划分, 即春季模型中 184 条训练数据, 92 条测试数据。夏季模型中 142 条训练数据, 72 条测试数据。秋季模型中 122 条训练数据, 60 条测试数据。冬季模型中 160 条训练数据, 80 条测试数据。将各季的训练数据的主成分矩阵作为神经网络的输入数据导入到 TIPDM 软件中。软件会自动将数据归一化至 [-1,1] 之间。以下分别对四季进行 B-P 神经网络的建模并对模型进行预报准确率检验。

6.4.1 春季预报模型的建立与检验

对春季建立 5-8-1 的 B-P 网络模型, 得到的模型参数如下表 6-6。

表 6-6 春季 B-P 模型参数表

季节	相关系数	平均绝对误差	均方根误差	相对绝对误差	相对平方根误差	样本总数
春季	0.901	10.5877	13.1547	47.9421%	35.8227%	184

由表可知，输入的主成分因子与 API 值相关性很高，所建 B-P 模型稳定。对春季 B-P 模型的预报准确率的检验如下(检验方法及公式同第五章)，通过对春季测试样本 92 日的实测值与预测值对比，得出所建立的春季 B-P 预报模型对大庆市春季PM₁₀的 API 数值预报的平均准确率为 77.9%。样本数据中测试数据的预测值与 API 实测值的拟合曲线如下图 6-1 所示。

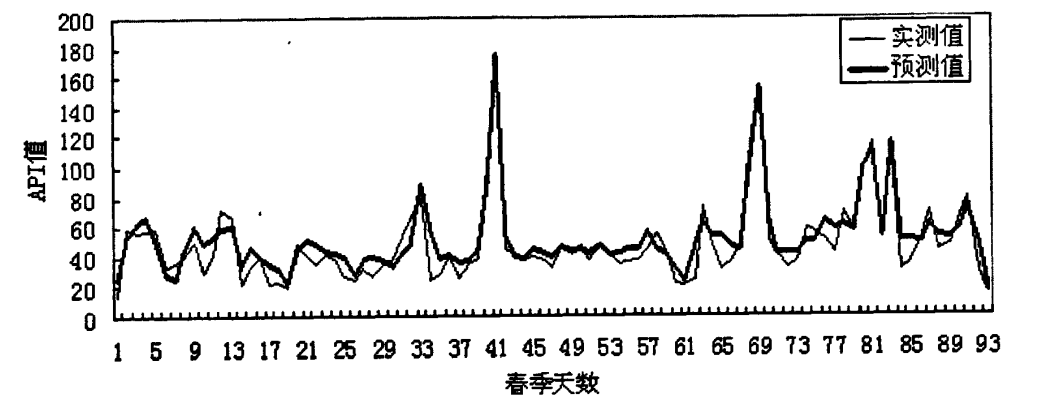


图 6-1 春季 B-P 模型PM₁₀的 API 值实测值与预测值拟合曲线图

由图可见，对春季 92 个测试数据的 B-P 网络预测值与 API 实测值的拟合度很高。尤其是对两个波峰的拟合效果非常好。经计算本模型的平均预报级别评分为 89.8 分，预报的精确度评分为 83.6 分，可以满足实际预报的需要。

6. 4. 2 夏季预报模型的建立与检验

对夏季建立 5-9-1 的 B-P 网络模型，得到的模型参数如下表 6-7。

表 6-7 夏季 B-P 模型参数表

季节	相关系数	平均绝对误差	均方根误差	相对绝对误差	相对平方根误差	样本总数
夏季	0.839	7.3344	8.9143	40.7276%	38.1255%	142

由表可知，输入的主成分因子与 API 值相关性很高，所建 B-P 模型稳定。对夏季 B-P 模型的预报准确率的检验如下，通过对夏季测试样本中 72 日的实测值与预测值对比，得出所建立的夏季 B-P 预报模型对大庆市夏季PM₁₀的 API 数值预报的平均准确率为 83.2%。样本数据中测试数据的预测值与 API 实测值的拟合曲线如图 6-2 所示。

由图可见，对夏季 72 个测试数据的 B-P 网络预测值与 API 实测值的拟合度很高，对波峰波谷拟合良好。经计算本模型的平均预报级别评分为 93.2 分，预报的精确度评分为 86.6 分，可以满足实际预报的需要。

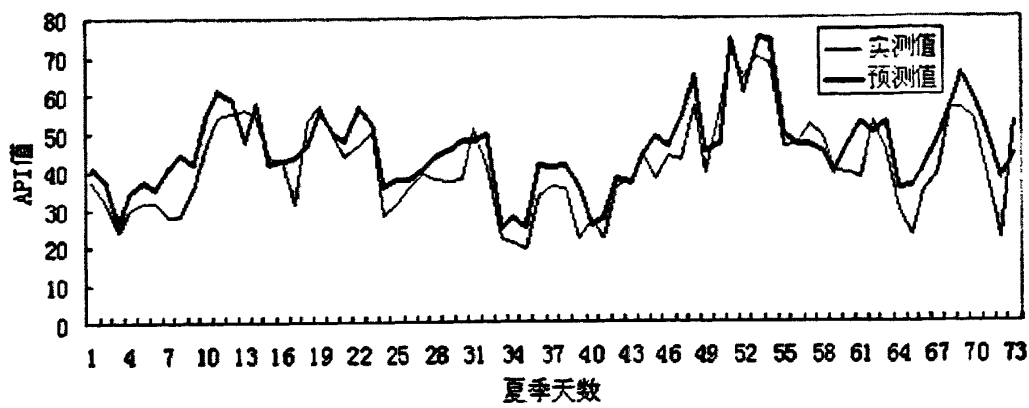


图 6-2 夏季 B-P 模型PM₁₀ 的 API 值实测值与预测值拟合曲线图

6. 4. 3 秋季预报模型的建立与检验

对秋季建立 6-9-1 的 B-P 网络模型，得到的模型参数如下表 6-8。

表 6-8 秋季 B-P 模型参数表

季节	相关系数	平均绝对误差	均方根误差	相对绝对误差	相对平方根误差	样本总数
秋季	0.8541	13.4457	15.1247	63.6244%	31.5153%	122

由表可知，输入的主成分因子与 API 值相关性很高，所建 B-P 模型稳定。对秋季 B-P 模型的预报准确率的检验如下，通过对秋季测试样本中 60 日的实测值与预测值对比，得出所建立的秋季 B-P 预报模型对大庆市秋季PM₁₀ 的 API 数值预报的平均准确率为 75.9%。样本数据中测试数据的预测值与 API 实测值的拟合曲线如下图 6-3 所示。

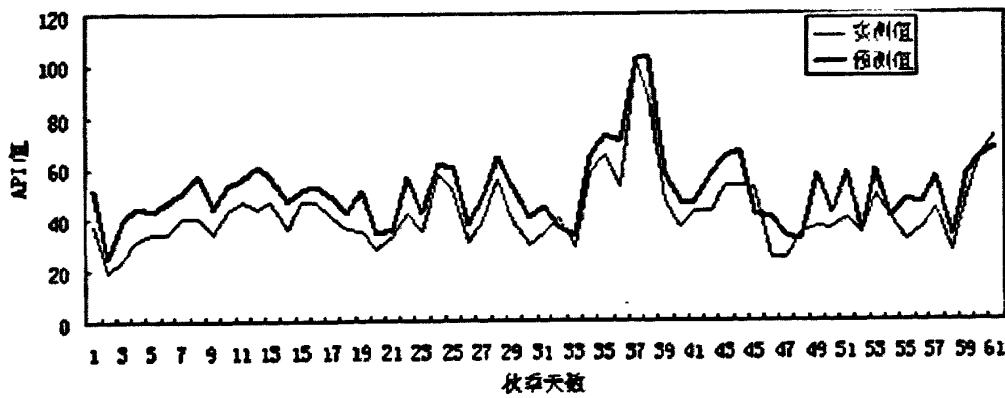


图 6-3 秋季 B-P 模型PM₁₀ 的 API 值实测值与预测值拟合曲线图

由图可见，对秋季 60 个测试数据的 B-P 网络预测值与 API 实测值的拟合度较高，对波峰波谷拟合良好，只是前 20 天预测数值偏低。经计算本模型的平均预报级别评分为 83.6 分，预报的精确度评分为 81.5 分，可以满足实际预报的需要。

6. 4. 4 冬季预报模型的建立与检验

对冬季建立 6-9-1 的 B-P 网络模型，得到的模型参数如下表 6-9。

表 6-9 冬季 B-P 模型参数表

季节	相关系数	平均绝对误差	均方根误差	相对绝对误差	相对平方根误差	样本总数
冬季	0.8529	8.8075	10.6591	34.9399%	31.4263%	160

由表可知，输入的主成分因子与 API 值相关性很高，所建 B-P 模型稳定。对冬季 B-P 模型的预报准确率的检验如下，通过对冬季测试样本中 80 日的实测值与预测值对比，得出所建立的冬季 B-P 预报模型对大庆市冬季 PM₁₀ 的 API 数值预报的平均准确率为 81.6%。样本数据中测试数据的预测值与 API 实测值的拟合曲线如下图 6-4 所示。

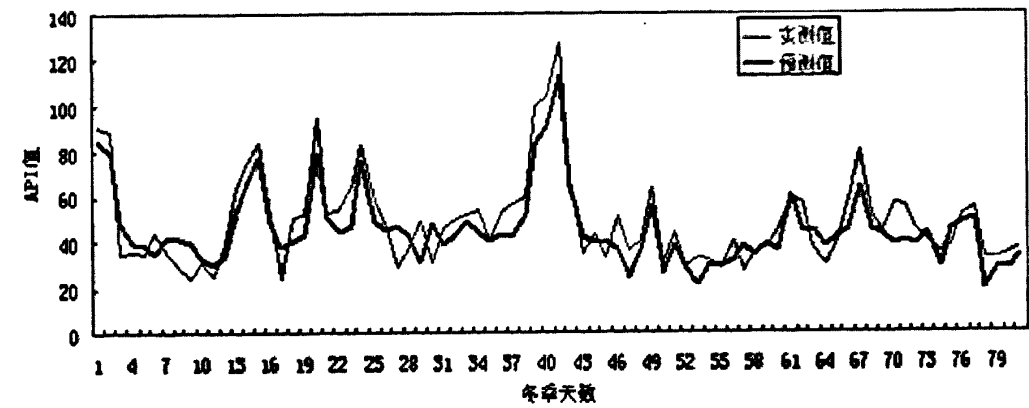


图 6-4 冬季 B-P 模型 PM₁₀ 的 API 值实测值与预测值拟合曲线图

由图可见，对冬季 80 个测试数据的 B-P 网络预测值与 API 实测值的拟合度较高，对波峰波谷拟合非常好。经计算本模型的平均预报级别评分为 89.5 分，预报的精确度评分为 84.6 分，可以满足实际预报的需要。

6. 4. 5 四季模型对比

通过对大庆市 PM₁₀ 的 API 值四季 B-P 神经网络预报模型的分析得出，春季模型预报的 API 值平均准确率为 77.9%，模型的平均预报级别评分为 89.8 分，预报的精确度评分为 83.6 分。夏季模型预报的平均准确率为 83.2%，模型的平均预报级别评分为 93.2 分，预报的精确度评分为 86.6 分。秋季模型预报的平均准确率为 75.9%，模型的平均预报级别评分为 83.6 分，预报的精确度评分为 81.5 分。冬季模型预报的平均准确率为 81.6%，模型的平均预报级别评分为 89.5 分，预报的精确度评分为 84.6 分。在预报准确率方面夏季模型>冬季模型>春季模型>秋季模型。主要原因是夏冬两季天气形式

较为稳定，故冬夏两季模型预报准确率较高。春季天气形式多变，多沙尘天气，导致春季模型预报率较低。而秋季 B-P 模型预报率最低，分析主要原因可能是秋季的训练数据过少导致所建模型不够稳定，若是基于大量实测数据的建模，相信预报准确率会大幅提升。

6.5 两种污染预报模型的比较

本文分别建立了基于逐步回归的大庆市四季 PM₁₀ 污染预报的数学模型和基于 B-P 人工神经网络的大庆市四季 PM₁₀ 污染预报模型，以下对两种模型的预报准确率、模型预报级别评分及精确度评分进行对比，见下表 6-10。

表 6-10 两种模型的对比评价表

模型	季节	预报准确率	预报级别评分	精确度评分
多元回归数学预报模型	春季	73.9%	84.8	76.8
	夏季	77.6%	93.3	82.3
	秋季	77%	91.8	80.7
	冬季	79.7%	87.7	83.3
B-P 神经网络预报模型	春季	77.9%	89.8	83.6
	夏季	83.2%	93.2	86.6
	秋季	75.9%	83.6	81.5
	冬季	81.6%	89.5	84.6

从以上对比分析表可以看出，春季、夏季、冬季所建的两个模型中，B-P 神经网络预报模型在预报准确率、级别预报评分和精确度评分上均优于多元回归数学预报模型。在对 PM₁₀ 的 API 数值预报的准确率上分别春夏冬三季分别提升了 6%、5.6%、1.9%，尤其是对历年春季两次强沙尘天气中 PM₁₀ 的 API 值的预报十分精准，而且四季预报的平均准确率也达到了 80%。但是 B-P 神经网络预报秋季模型的预报准确率要较多元回归数学模型低，其主要原因可能是由于数据样本量的局限性，导致秋季 B-P 网络预报准确率偏低，由此看来，对大气污染的预报还是要建立在海量数据的基础上。但总体看，B-P 神经网络所建立的污染预报模型要比传统的统计方法拥有更好的预测能力，这也反映出了 B-P 人工神经网络在处理污染预报这类有着较强非线性变化特点的问题时优势明显。

6.6 本章小结

(1)应用主成分分析法对备选因子进行了因子精选过程,分别从春季、夏季、秋季、冬季的相关系数矩阵中提取出了 5 个、5 个、6 个、6 个主成分,这些主成分对原变量的解释都达到了 85%以上。总体上,原变量的损失信息较少,主成分分析效果良好,起到了很好的降维去噪作用。精选出的这些主成分矩阵即作为构建 B-P 神经网络预测模型的输入数据集。

(2)通过对 B-P 神经网络层数、输入输出节点、隐含层节点、各层传递函数及网络主要参数的确定,应用 TIPDM 软件实现了大庆市 PM₁₀ 污染的 B-P 神经网络预报模型的建立。对四季分别建立了结构为 5-8-1、5-9-1、6-9-1、6-9-1 的神经网络预报模型。

(3)对四季的 B-P 神经网络预报模型分别进行了预报准确率的检验,并进行了预报级别正确性评分和模型精确度评分。得出了模型预报准确率分别为 77.9%、83.2%、75.9%、81.6%,平均预报准确率为 80%。预报级别正确性评分分别为 89.8、93.2、83.6、89.5。模型精确度评分分别 83.6、86.6、81.5、84.6。所建立预报模型效果理想,能达到实际预报工作的需要。

(4)对所建立的四季回归预报模型和 B-P 神经网络预报模型进行了对比,结果表明 B-P 神经网络预报模型在预报准确率、级别预报评分和精确度评分上均优于回归预报模型。这反映出了 B-P 人工神经网络在处理污染预报这类有着较强非线性变化特点的问题时优势明显。

结论与建议

结论

本文以大庆市历年 PM_{10} 的 API 逐日实测值及同期的气象资料为数据基础, 对大庆市 PM_{10} 污染分年、季节、月进行了污染特征及趋势分析, 找出了 PM_{10} 污染变化规律及成因。并通过对大庆市历年气象因子数据库的统计分析, 探究了各气象因子与 PM_{10} 污染之间的关系, 找出了与 PM_{10} 污染相关性高的气象因子。同时分季节建立了基于逐步回归方法的多元回归数学模型和基于 B-P 人工神经网络模型两种预报模型, 并分别应用两种模型对大庆市四季 PM_{10} 污染的 API 值进行了历史拟合及预测, 得到了较好的效果。全文得出以下结论:

(1) 大庆市空气质量总体优良, PM_{10} 一直为该市首要污染物, 故空气质量等级可用 PM_{10} 的污染等级来表述, 及时开展大庆市 PM_{10} 污染的防治和预报工作十分重要。

(2) 从大庆市 PM_{10} 污染年变化趋势来看, 自 2001 年至 2006 年, PM_{10} 浓度始终在 $0.060\text{mg}/\text{m}^3$ 上下波动, 06 年以后, PM_{10} 浓度明显下降。从季节变化趋势看, 春冬两季 PM_{10} 污染较重, 夏秋两季污染状况较轻。从月变化趋势上看, 每年的 1、2、4、5、11、12 月污染指数较高, 3、6、7、8、9 月较低。对一些特殊情况分析可知, 大庆市采暖季的 PM_{10} 污染较非采暖季严重, 沙尘过境期间和过年期间 PM_{10} 污染明显较平时严重。

(3) 通过对大庆市历年各气象因子数据的统计分析可知, 气压因子方面, 大庆市四季日均气压和 08h 气压的均值大小依次为冬季>秋季>春季>夏季。气温因子方面, 夏季气温高, 冬季气温低, 春秋气温基本持平。湿度因子方面, 春秋干燥, 夏冬两季湿度较大。能见度因子方面, 四季 08h 能见度均值大小依次为夏季>秋季>春季>冬季, 14h 能见度均值大小依次为秋季>夏季>春季>冬季。日照因子方面, 春夏两季日照时数较长, 冬季较短。云量因子方面, 四季变化趋势为夏季>春季>秋季>冬季。风速因子方面, 春季的风速各因子均值最高, 夏秋次之, 冬季最低。

通过对大庆市历年 PM_{10} 的 API>100 的污染高值日气象因子分析得出, 春季高污染日在气象上具有低气压、高气温、低能见度、低日照、高云量、高风速的特点。夏季无高污染日。秋季高污染日具有日均气压略低、气温基本持平、相对湿度高、能见度低、日照低、总云量高和风速偏高的特点。冬季污染日具有日均气压高、气温偏低、相对湿度偏高、能见度偏低、日照日合计值偏低、总云量偏高和风速基本持平的特点。

(4)对大庆市 PM_{10} 污染的 API 值与各气象因子与特征因子分季节进行相关分析, 初选出与春季 PM_{10} 污染的 API 值相关度较高的因子 15 个, 夏季 13 个, 秋季 10 个, 冬季 13 个。这些因子将作为逐步回归建模与主成分分析过程的备选因子。

(5)对大庆市 PM_{10} 污染分四季建立了多元回归预报模型, 应用逐步回归的方法对各个季节备选因子做了精选, 得出回归方程分别为:

$$Y(\text{春})=0.106 * X_1-0.209 * X_2-0.196 * X_3+0.213 * X_4+0.188 * X_5-2.9E-015$$

$$Y(\text{夏})=0.181 * X_1+0.179 * X_2-0.206 * X_3+0.195 * X_4+0.309 * X_5-9.1E-015$$

$$Y(\text{秋})=0.413 * X_1-0.312 * X_2-0.310 * X_3+0.313 * X_4+0.231 * X_5-0.066$$

$$Y(\text{冬})=0.101 * X_1-0.130 * X_2-0.152 * X_3-0.253 * X_4+0.196 * X_5+0.407 * X_6-5.3E-015$$

对多元回归方程分别进行了 F 检验与预报准确率检验, F 检验结果表明各方程回归效果显著。对各方程的与预报准确率进行检验得出四季模型预报的 API 平均准确率分别为 73.9%、77.6%、77%、79.7%, 平均为 77.1%, 能够达到实际预报工作需要。

(6)通过主成分分析法对四季备选因子进行了精选, 分别选出了春季 5 个, 夏季 5 个, 秋季 6 个, 冬季 6 个主成分作为 B-P 神经网络建模的输入层节点, 表明主成分分析法对备选因子降维去噪效果良好。

(7)应用 TIPDM 软件对四季分别构造了 B-P 神经网络预报模型。得出四季模型预报的 API 值平均准确率分别为 77.9%、83.2%、75.9%、81.6%, 平均准确率为 80%, 能够达到实际预报工作的需要。

(8)对所建立两种模型的预报准确率、模型预报级别评分及精确度评分进行对比, 得出春、夏、冬三季的 B-P 神经网络模型这三项指标均优于回归预报模型, 证明了 B-P 神经网络在处理污染预报这类有着较强非线性变化特点的问题时优势明显。

建议

(1)可继续完善 B-P 神经网络的输入因子。在今后的研究中, 可尝试加入大气稳定性、逆温层数据等探空气象资料及污染源的资料, 将有利于提高预报准确率。

(2)本文只是建立了多元回归模型与 B-P 神经网络模型两个预报模型, 今后的研究中, 可以尝试建立更多模型来进行对比。比如可以尝试应用 RBF 神经网络和因子分析法来建立预报模型。

(3)本文应用 TIPDM 数据挖掘软件来建立 B-P 神经网络预报模型, 其是应用 Levenberg-Marquardt 动量项法(即 L-M 法)来实现的。以后的研究中, 可以尝试对多种

算法进行对比，如动量 B-P 算法、弹性 B-P 算法、BFGS 拟牛顿法、SCG 算法等。找出最适合的建模算法，提高网络训练的精度进而提高模型预报的准确率。

(4)本文应用神经网络预报方法与统计方法给出了预报模型，实际应用中，可以结合数值预报与潜势预报等多种预报方法。并可尝试将多种预报方法进行加权，集成预报，以提高预报的准确率。

致 谢

三年的研究生生活即将结束，在此论文完成之际，首先，我要衷心感谢恩师杨巧艳副教授对本人的精心指导。三年来，恩师的谆谆教导、悉心栽培使我时时如沐春风、倍感温馨。杨老师严于律己、宽以待人的高尚风范，朴实无华、平易近人的人格魅力，渊博的学识和严谨治学的态度必将使我受益终生。

诚挚感谢地环学院刘丹老师与李启斌老师一直给予无私的指导与帮助，同时感谢各位同学在生活与学习上给予我的帮助与支持。

感谢大庆市环保局与大庆市气象局的有关领导为本文提供了研究数据，并在某些问题上给予了中肯的意见。

感谢我的父母和亲朋好友，他们默默的支持和细心的关怀，使我能够不断前行，顺利完成学业。

最后，向在论文审议、评阅、答辩过程中给予指导的专家、教授表示衷心的感谢。

参考文献

- [1] Schwartz. J. Particular air pollution and chronic respiratory diseases[J]. Environ. Res. 1993, 62(1):7-13.
- [2] 魏复胜, 胡伟, 滕恩江, 等. 空气污染与儿童呼吸系统患病率的相关分析[J]. 中国环境科学, 2000, 20, (3):220-224.
- [3] 郝吉明, 马广大等. 大气污染控制工程(第二版)[M]. 北京:高等教育出版社, 2002.
- [4] 吴 扬, TSP源解析的回归计算及总量控制分析[D]. 兰州大学硕士论文, 2006.
- [5] Zanobetti A, Schwartz, Gold D. Are there sensitive subgroups for the effects of air borne particles[J]. Environ Health Perspect, 2000, 108(9), 841-845.
- [6] Lippmann M, Ito K, Nadas A, et al. Association of particulate matter components with mortality and morbidity in urban populations[J]. Res Rep. Health Eff Inst, 2000, (95), i~iv:1-84.
- [7] 秦世广, 杨杰等. 黑碳气溶胶及其在气候变化研究中的意义[J]. 气象, 2001, 27(11): 3-7.
- [8] 中国环境状况公报(2007、2008、2009). 国家环保部网站. <http://www.sepa.gov.cn>.
- [9] 中国环境状况概况(2007、2008、2009). 国家环保部网站. <http://www.sepa.gov.cn>.
- [10] 张美根, 韩志伟, 雷孝恩. 城市空气污染预报方法简述[J]. 气候与环境研究, 2001, 6(1):113-118.
- [11] 李宗恺等. 空气污染气象学原理及应用[M]. 北京:气象出版社, 1985.
- [12] 雷孝恩等. 大气污染数值预报基础和模式[M]. 北京:气象出版社, 1998.
- [13] 徐祥德, 汤绪等编著. 城市化环境气象学引论[M]. 北京:气象出版社, 2002.
- [14] 方志刚. 城市空气污染预报特点及方法[J]. 辽宁城乡环境科技, 2000, 20(5):19-21.
- [15] 陈柳, 马广大, 纪海维. 城市大气污染预报模式的研究进展. 西安科技学院学报, 2003, 23(4):411-414.
- [16] 殷达中, 刘万军. 海外大气预报的状况及方法[J]. 气象科技, 1997, (1):19-23.
- [17] Yilmaz Yildirim, Nuhi Demircioglu, Mehmet Kobya, Mahmut Bayramoglu. A mathematical modeling of sulphur dioxide pollution in Erzurum City[J]. Environmental Pollution, 2002, 118:411-417.

-
- [18] Ota M. On some statistical relations between oxidant concentration and low altitude meteorological elements in the coastal region of Hiroshima[J]. *Tenki*, 1976, 23: 487-494.
- [19] Ota M. Preliminary statistical investigation for the prediction of the daily maximum oxidant concentration in Hiroshima[J]. *Society of Japan*, 1983, 61(1):142-150.
- [20] Jorquera, Hector et al. Forecasting ozone daily maximum levels at Santiago, Chile. *Atmospheric Environment*. 1998, 32(20):3415-3424.
- [21] Petros Koutrakis, Sonja N Sax, Jeremy A Sarnat et al. Analysis of PM_{10} , $PM_{2.5}$ and $PM_{2.5-10}$ Concentrations in Santiago, Chile, from 1989-2001. *Journal of the Air & Waste Management Association*, 2005, 55(3):342-351.
- [22] Van Der Wal. J. T et al. How contribute emission of PM_{10} and meteorology to concentrations of fine particles in the Netherlands. *Aerosol Science*, 27, Suppl 1, Sep, 1996 Pergamon press Inc:5681-5682.
- [23] Gong Yuh Lin. Oxidant prediction by discriminant analysis in the South coast air basin of California[J]. *Atmospheric Environment*, 1982, 16(1):135-143.
- [24] Martin Branis, PM_{10} and black smoke in a small settlement: case study from the Czech Republic[J]. *Atmospheric Environment* 37, 2003:83-92.
- [25] Boznar M, Lesjak M, Mlakar P. A neural network-based method for the short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain[J]. *Atmospheric Environment, Part B, Urban Atmosphere*, 1993, 27(2):221-230.
- [26] Yi. J, Prybutok R. A neural network model forecasting for prediction of daily maximum ozone concentration in industrialized urban area[J]. *Environmental Pollution* 1996, 92(3):349-357.
- [27] Perez P, Trier A, Reyes J. Prediction of $PM_{2.5}$ concentrations several hours in advance using neural networks in Santiago, Chile[J]. *Atmos Environ*, 2000(34): 1189-1196.
- [28] P. Viotti, G. Liuti and P. Di Genova. Applications of an artificial neural network (ANN) to the city of Perugia. *Ecological Modelling*[J]. *Atmospheric urban pollution*,
-

- 2002, 148:27-46.
- [29] Sang Hyun Sohn, etc. Prediction of Ozone Formation Based on Neural Network [J]. *Envir. Engrg. ASCE*, 2000, 8:688-696.
- [30] 徐大海, 朱蓉. 大气平流扩散的箱格预报模式与污染潜势指数预报[J]. *应用气象学报*, 2000, 11(1):1-12.
- [31] 王迎春. 北京市空气质量业务预报[A]. *全国城市空气污染预报及防治会议论文集*, 2001, (8):1-12.
- [32] 王川, 刘子臣, 孟炜. 西安空气污染气象条件预报服务系统[J]. *陕西气象*, 2002(3):19-20.
- [33] 杨民, 杨文科, 王庆梅等. 兰州市冬季空气质量潜势预报[J]. *甘肃气象*, 2003, 21(1):24-27.
- [34] 刘实, 王宁, 朱其文等. 长春市空气污染潜势预报的统计模型研究[J]. *气象*, 2001, 28(1):8-12.
- [35] 雷孝恩, 张美根, 韩志伟等. 大气污染数值预报基础和模式[M]. 北京:气象出版社, 1998.
- [36] 徐大海, 朱蓉. 城市空气污染预报模式的推广应用[J]. *Chinese Academy of Meteorological Sciences Annual Report*, 1999, (00).
- [37] 朱蓉, 徐大海. 新一代 CAPPS 系统的开发与推广应用[J]. *Chinese Academy of Meteorological Sciences Annual Report*, 2002, (00).
- [38] 朱蓉, 徐大海, 周朝东等. CAPPS 在国家气象中心多城市污染指数数值预报业务系统中的应用[J]. *应用气象学报*, 2002, 13(SI):204-213.
- [39] 韩志伟, 杜世勇, 雷孝恩等. 城市空气污染数值预报模式系统及其应用[J]. *中国环境科学*, 2002, 22(3):202-206.
- [40] 董亚龙, 吴宁. CAPPS 模式在铜川空气质量预报中的效果检验[J]. *陕西气象*, 2004, (5):13-16.
- [41] 王淑英. 北京地区 PM_{10} 污染特征及其与气象条件的相关分析, 第八届全国大气环境学术会议论文集[A], 2000. 10:13-18.
- [42] 黄磊, 王赐震等. 大气污染时间序列分析预报方法研究[J]. *青岛海洋大学学报*, 2001, 31(1):14-20.
- [43] 柴微涛, 宋述军. 成都市城区空气污染指数的时间序列分析[J]. *成都理工大学学报*

- (自然科学版), 2007, 34(4):485-488.
- [44] 郭尚媛. 西安市采暖季 PM_{10} 污染预报研究[D]. 西安建筑科技大学论文, 2005.
- [45] 周变红. 基于因子分析法的宝鸡市 PM_{10} 污染四季预报模型研究[D]. 西安建筑科技大学硕士论文, 2007.
- [46] 李祚泳, 邓新民. 环境污染预测的人工神经网络模型[J]. 成都气象学院, 1997, 12(4):279-283.
- [47] 王俭, 郑龙熙等. 基于 BP 模型的大气污染预报方法的研究[J]. 环境科学研究, 2002, 15(5):62-64.
- [48] 王俭, 郑龙熙等. BP 模型的改进及其在大气污染预报中的应用[J]. 城市环境与城市生态, 2002, 15(5):17-19.
- [49] 王灿星, 祁国伟等. BP 网络用于大气中颗粒物(TSP)预测的研究[J]. 仪器仪表学报, 2003, 24(4):529-530.
- [50] 邓伟妮. 基于 BP 神经网络的西安市 PM_{10} 污染预报及其 MATLAB 实现[D]. 西安科技大学硕士论文, 2008.
- [51] 王兰霞. 基于 RBF 神经网络的 PM_{10} 污染预测研究[D]. 西安建筑科技大学硕士论文, 2008.
- [52] 王芳, 程水源等. 遗传算法优化神经网络用于大气污染预报[J]. 北京工业大学学报, 2009, 35(9):1230-1234.
- [53] 雷蕾. 人工神经网络在大气污染预报中的应用研究[D]. 北京工业大学硕士论文, 2007.
- [54] 周丹, 魏云慧. 大庆市空气质量分析及空气污染控制对策. 大庆社会科学, 2005, (2):51.
- [55] 张德丰等. Matlab 神经网络应用设计[M]. 北京:机械工业出版社, 2009.
- [56] 李祚泳. 城市综合环境质量评价的 BP 网络模型[J]. 系统工程, 1995, (13):9-13.
- [57] 白润才, 殷伯良, 孙庆宏. BP 神经网络模型在城市环境质量评价中的应用[J]. 辽宁工程技术大学学报(自然科学版), 2001, (3):373-375.
- [58] 李祚泳, 邓新民. 基于 BP 网络的大气环境质量评价[J]. 干旱环境监测. 1997, (1):62-63.
- [59] 郭劲松, 龙腾锐, 霍国友等. 四种水质综合评价方法的比较[J]. 重庆建筑大学学报

- 报, 2000, (4):6-12.
- [60] 郭宗楼. 径向基函数网络模型在水质评价中的应用. 浙江大学学报(农业与生命科学版), 2001, (3):335-338.
- [61] 倪深海, 白玉慧. BP 神经网络模型在地下水水质评价中的应用[J]. 系统工程理论与实践, 2000, (8):124-127.
- [62] 汤丽妮, 张礼清, 王卓. 人工神经网络在生态环境质量评价中的应用[J]. 四川环境, 2003, 22(3):69-72.
- [63] 李洪义, 史舟, 沙晋明等. 基于人工神经网络的生态质量遥感评价[J]. 应用生态学报, 2006, 17(8):1475-1480.
- [64] 郭宗楼, 马帆. 河流水质预测的因素状态网络模型研究[J]. 环境科学研究, 2001, 14(1):49-52.
- [65] 田景文, 高美娟. 人工神经网络算法研究及应用[M]. 北京:北京理工大学出版社, 2006.
- [66] 张良均等. 神经网络实用教程. 北京:机械工业出版社[M], 2008.
- [67] 韩立群. 人工神经网络理论设计与应用[M]. 北京:化学工业出版社, 2002.
- [68] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京:电子工业出版社, 2005.
- [69] 董长虹. Matlab 神经网络与应用[M]. 北京, 国防工业出版社, 2007.
- [70] 何晓群. 多元统计分析[D]. 中国人民大学出版社, 2008.
- [71] 国家环保部网站:<http://www.zhb.gov.cn/>.
- [72] 关莉, 李守春等. 解析大庆市“十五”期间环境空气中可吸入颗粒物污染现状[J]. 环境研究与监测, 2006, 19(4):48-50.
- [73] 苏春东, 候丽等. 大庆市环境空气质量控制初探[J]. 中国资源综合利用. 2008, 26(7):21-22.
- [74] 周丹, 魏云慧. 大庆市空气质量分析及空气污染控制对策[J]. 大庆社会科学, 2005, (2):51.
- [75] 肖红, 夏文前等. 从大庆地区逆温特征看大庆大气污染的防治措施[J]. 黑龙江气象, 2006, (3):28-29.
- [76] 罗应婷, 杨钰娟. SPSS 统计分析从实践到精通(第二版)[D]. 北京:电子工业出版社

社, 2010.

- [77] Hecht Nielson R. Counter Propagation Networks. Applied Optics. 1987, 26(12): 4979-4984
- [78] Lippmann R. P. An introduction to computer with neuralnets. IEEE ASSP Magazine, 1987(4):4-22.
-

攻读硕士学位期间发表的论文

发表论文:

王顶, 杨巧艳. 大庆市春冬两季 PM_{10} 污染预报模型的构建. [J]西南交通大学学报增刊. (已录)