

# 基于数学模型的PM<sub>2.5</sub>月尺度浓度预测研究

李鑫洋, 刘娟

天津理工大学环境科学与安全工程学院, 天津 300384

**摘要:** 精准预测PM<sub>2.5</sub>浓度能为大气污染治理提供科学依据。为了在明确PM<sub>2.5</sub>排放量的情况下使PM<sub>2.5</sub>浓度变化具有可获得性, 基于全国各省份2013—2017年PM<sub>2.5</sub>月均浓度和月排放量历史监测数据, 构建两者间的数学关系模型, 通过Pearson分析将全国各省份划分为两种类型(分布规律和分布分散), 并使用2018—2020年PM<sub>2.5</sub>月均浓度数据验证数学模型的精度。结果表明: 在PM<sub>2.5</sub>月排放量与月均浓度散点分布规律的19个省份中, 11个省份的决定系数( $R^2$ )介于0.60~0.90之间, 8个省份的 $R^2$ 介于0.50~0.60之间。分布规律省份数学模型精度验证结果显示, 14个省份的均方根误差(RMSE)介于6.00~16.00之间, 3个省份的RMSE介于16.00~20.00之间。在PM<sub>2.5</sub>月排放量与月均浓度散点分布分散的10个小组中, 5个小组的 $R^2$ 介于0.70~0.90之间, 4个小组的 $R^2$ 介于0.60~0.70之间。分布分散省份数学模型精度验证结果显示, 8个小组的RMSE介于6.00~16.00之间。因此, 通过拟合方法得出的数学模型对于PM<sub>2.5</sub>浓度预报具有一定的适用性, 且可以较准确地预测未来情景中的PM<sub>2.5</sub>浓度变化。

**关键词:** PM<sub>2.5</sub>; 拟合分析; 排放量; 决定系数; 浓度预测

**中图分类号:** X823 **文献标志码:** A **文章编号:** 1002-6002(2025)01-0180-11

**DOI:** 10.19316/j.issn.1002-6002.2025.01.19

## Prediction of PM<sub>2.5</sub> Monthly Scale Concentration Based on Mathematical Model

LI Xinyang, LIU Juan

School of Environmental Science and Safety Engineering, Tianjin University of Technology, Tianjin 300384, China

**Abstract:** Accurate prediction of PM<sub>2.5</sub> concentration can provide a scientific basis for air pollution management. In order to have accessibility to the concentration changes in the case of clear PM<sub>2.5</sub> emissions, based on the historical monitoring data of monthly average PM<sub>2.5</sub> concentration and monthly emissions of each province from 2013 to 2017, a mathematical relationship model was constructed, the provinces were classified into two types by the Pearson analysis of SPSS, and the accuracy of the mathematical model was verified using the monthly average PM<sub>2.5</sub> concentration from 2018 to 2020. The results showed that: there were 19 provinces in the category I of the scatter distribution of PM<sub>2.5</sub> emissions and monthly average concentrations, the coefficient of determination ( $R^2$ ) of 11 provinces ranged from 0.60 to 0.90, and the  $R^2$  of 8 provinces ranged from 0.50 to 0.60, and the root mean square error (RMSE) of 14 provinces ranged from 6.00 to 16.00 and the RMSE of 3 provinces ranged from 16.00 to 20.00 in the validation of the model's accuracy; and in the dispersed distribution of the category II group of provinces was divided into 10 groups, of which 5 groups had  $R^2$  between 0.70 and 0.90, 4 groups had  $R^2$  between 0.60 and 0.70, and 8 groups had RMSE between 6.00 and 16.00 at the time of model accuracy validation. It is therefore concluded that the mathematical model obtained by the fitting method has certain applicability to the prediction of PM<sub>2.5</sub> concentration and can better predict the changes in PM<sub>2.5</sub> concentrations in future scenarios.

**Keywords:** PM<sub>2.5</sub>; fitting analysis; emission; coefficient of determination; concentration prediction

近年来, 空气污染物浓度预测一直是学术界的研究热点, 其中PM<sub>2.5</sub>浓度预测备受关注。PM<sub>2.5</sub>是由空气动力学直径小于2.5 μm的颗粒物组成, 大多来源于燃料燃烧等人类活动<sup>[1-3]</sup>。PM<sub>2.5</sub>由于具有吸附性质, 很容易吸附有毒物质, 经过呼吸作用等进入肺部, 从而引起心血管疾病、

呼吸系统损伤甚至死亡, 已被国际癌症研究机构(IARC)确定为主要致癌物之一<sup>[4-7]</sup>。在我国, PM<sub>2.5</sub>监测指标在2012年被纳入环境空气质量标准(分期实施), 全国范围的PM<sub>2.5</sub>浓度监测在此之前是缺失的<sup>[8]</sup>。虽然近年来相关部门实施了一系列大气污染防治措施, 使得PM<sub>2.5</sub>浓度有所

收稿日期: 2023-08-12; 修订日期: 2024-03-27

第一作者简介: 李鑫洋(2000-), 女, 天津人, 在读硕士研究生。

通讯作者: 刘娟

下降,但在大部分地区仍是主要的大气污染物<sup>[9]</sup>。PM<sub>2.5</sub> 浓度是空气质量的关键指标之一,对其进行准确预测对于科学开展空气质量评估和环境管理政策制定具有重要意义。因此,对 PM<sub>2.5</sub> 浓度的有效预测需要引起关注,并应着重开发 PM<sub>2.5</sub> 浓度与潜在相关因素之间的相关性分析模型<sup>[10]</sup>。

PM<sub>2.5</sub> 浓度预测的主要手段包括物理模型、统计模型、机器学习等<sup>[11-12]</sup>。其中,物理模型采用物理/化学方法来对污染物的扩散、迁移、转化等进行模拟,常见模型包括社区多尺度空气质量(CMAQ)建模系统、天气研究和预报(WRF)模型、气象化学耦合(WRF-Chem)模型等。此类模型需要进行大量计算,存在计算过程复杂以及数据采集困难等问题<sup>[13-15]</sup>。统计模型利用人工智能模型和时间序列理论来描述历史数据与未来数据之间的统计关系,将空气质量与气象因素联系起来,常见的时间序列趋势分析方法包括线性回归等<sup>[16]</sup>。统计模型是基于数据的模型。相关研究发现,统计模型相比物理模型具有更好的预测性能<sup>[17]</sup>。机器学习具有较好的容错性和灵活性,主要包括人工神经网络(ANN)、支持向量机(SVM)、多层感知器(MLP)等。然而,由于网络结构相对简单,机器学习方法只能用以获取短时间序列变化,适合短期预测<sup>[18-21]</sup>。

影响 PM<sub>2.5</sub> 浓度的因素大致可分为两类,分别为自然因素和社会因素。自然因素中的风速、地形和降水等会影响 PM<sub>2.5</sub> 的扩散、聚集和转移,其中复杂的地形条件会给当地气温、风速、风向等带来改变,从而影响污染物的扩散。一般情况下,气象条件决定了 PM<sub>2.5</sub> 的扩散能力<sup>[22-23]</sup>。社会因素会影响 PM<sub>2.5</sub> 的排放,如人口变化、交通运输和工业生产情况等。相关研究通过对中国 300 余个重点城市的长期调研发现,城市化的推进通常伴随着经济增长和工业化加速,会在一定程度上增加空气污染程度<sup>[24-25]</sup>。

关于 PM<sub>2.5</sub> 排放量与浓度的关系,YU 等<sup>[26]</sup>基于足迹建模方法和排放清单对 2015 年华北平原 PM<sub>2.5</sub> 浓度进行了建模和浓度反演,并对 2016 年 PM<sub>2.5</sub> 浓度进行了试验预测。其研究结果显示,6 个城市的 PM<sub>2.5</sub> 浓度模拟值与观测值的相关系数在 0.40~0.56 之间,且该模型在 2016 年预测中保持了较高水准,分别有 92% 和 46% 的样本满足模型标准和建模目标。MA 等<sup>[27]</sup>比较了华北平原 3 份排放清单的初级颗粒物排放,并基于

WRF-Chem 模型对 PM<sub>2.5</sub> 浓度变化进行了再现。其研究结果显示,模拟结果的空间分布效果良好,大部分地区的误差小于 20 μg/m<sup>3</sup>。CHUAI 等<sup>[28]</sup>在 PM<sub>2.5</sub> 排放量和浓度的空间关系分析中,将各省份 PM<sub>2.5</sub> 排放量和浓度的平均值作为变量,对 1998—2015 年数据进行了线性回归分析,结果显示相关系数在 0.68~0.79 之间,其中 1999 年的相关系数最低。ZHANG 等<sup>[29]</sup>模拟了 PM<sub>2.5</sub> 浓度与大气污染物(SO<sub>2</sub>、NO<sub>x</sub> 等)排放量之间的统计关系,以及城市层面的风和邻近转移等相关因素的影响,发现风力较小时,PM<sub>2.5</sub> 浓度的模拟值与监测值存在较大偏差,大多数城市的模拟值低于监测值。

进入 21 世纪以来,随着经济的快速发展,我国大气环境污染问题凸显,重污染天气频发。研究显示,相关大气污染问题主要是受污染物排放和气象条件等因素的综合影响<sup>[30]</sup>。因此,本文旨在通过研究 PM<sub>2.5</sub> 排放量与浓度之间的关系,使得在已知 PM<sub>2.5</sub> 排放量的情况下对 PM<sub>2.5</sub> 浓度开展预测更具有可操作性,在获得 PM<sub>2.5</sub> 排放量的情况下根据省份数学拟合模型就可以估算出 PM<sub>2.5</sub> 浓度信息,这将有助于分析不同地理特征下的由 PM<sub>2.5</sub> 排放量变化引起的 PM<sub>2.5</sub> 浓度变化。

## 1 数据与方法

### 1.1 数据来源与处理

#### 1.1.1 数据来源

本文研究范围为全国各省份(由于数据缺失,不包括新疆、西藏、台湾、香港、澳门)。本研究使用的数据来源于各省(区、市)生态环境厅(局)发布的环境空气质量月报,以及空气质量历史数据查询网([www.aqistudy.cn/historydata](http://www.aqistudy.cn/historydata))收录的历史数据。PM<sub>2.5</sub> 浓度取各省(区、市)月均值,单位为 μg/m<sup>3</sup>,时间范围为 2013—2020 年。PM<sub>2.5</sub> 月排放量数据来源于清华大学编制的中国多尺度排放清单,单位为 10<sup>4</sup> t<sup>[31-33]</sup>。

#### 1.1.2 数据预处理

进行模型验证时,需要对数据作归一化处理,以消除不同维度数据的差别,而后再进行拟合分析。

$$M = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中: $M$  为变量  $x$  经归一化处理后的值, $x$  为变量原始值, $x_{\min}$  为变量最小值, $x_{\max}$  为变量最大值。

归一化后的浓度数值大小在 $[0,1]$ 区间内。

## 1.2 研究方法

研究显示,  $PM_{2.5}$  浓度与天气因素之间的关系呈现复杂的非线性特征。虽然空气质量与污染物排放量关系密切<sup>[34]</sup>, 但从排放源排放出的  $PM_{2.5}$  会经过各种因素的叠加作用, 最终表现出浓度数值的变化。本文讨论了基于  $PM_{2.5}$  排放量的  $PM_{2.5}$  浓度预测方法, 使用线性拟合和多项式拟合的方法探讨了两者之间的关系。其中, 2013—2017年各省份  $PM_{2.5}$  月均浓度数据用于数学模型的建立, 2018—2020年各省份  $PM_{2.5}$  月均浓度数据用于模型精确度评估。

### 1.2.1 分类划分省份

在建立各省份数学模型之前, 需要判断其数据是否满足拟合的标准, 即是否可以得出数学模型。首先, 利用 Pearson 相关系数进行相关性分析。若该省份排放量和平均浓度的散点分布满足相关性条件, 则可以直接进行拟合。若不满足, 则需要对散点进行划分, 使之满足相关性条件后, 再进行建模。因此, 会形成两种类别的省份分组: 一种是  $PM_{2.5}$  月排放量与月均浓度散点分布规律的省份, 可以直接进行线性拟合; 另一种是散点分布分散的省份, 需要对散点进行划分, 使其满足特定的分布规律后, 再进行数据拟合, 进而得到对应的数学模型。

通过 Pearson 相关系数来判断省份  $PM_{2.5}$  排放量与平均浓度的线性相关程度。相关性判别主要基于两个指标, 分别为  $P$  值和相关系数。其中:  $P$  值用以表征两个变量是否显著相关。若  $P < 0.05$ , 则认为两个变量显著相关; 否则两者不相关。在满足相关的基础上, 相关系数用来度量变量之间的相关性, 其值介于  $-1 \sim 1$  之间。相关系数绝对值越大, 表示变量之间的相关性越强; 反之, 相关性越弱。在本文中, 该值越接近 1, 说明排放量与平均浓度的正相关性越强; 越接近 0, 说明两者之间彼此独立, 没有相关性。

### 1.2.2 建立拟合数学模型

利用 2013—2017 年各省份  $PM_{2.5}$  月均浓度数据与中国多尺度排放清单模型中的排放量数据, 建立两者之间的数学模型, 从而对未来的  $PM_{2.5}$  浓度进行预测<sup>[35]</sup>。其中, 使用的统计方法包括线性拟合和多项式拟合。计算得出的决定系数 ( $R^2$ ) 用来衡量两者之间的线性关系强度, 取值范围为  $0 \sim 1$ 。自变量组合对应的  $R^2$  值越高, 表明

其对  $PM_{2.5}$  浓度变化的解释越准确<sup>[36]</sup>。

### 1.2.2.1 线性拟合模型

对散点分布规律的省份进行线性拟合, 线性拟合模型函数为

$$y = kx + b \quad (2)$$

式中:  $x$  为自变量, 在本文中指排放量,  $10^4 t$ ;  $y$  为因变量, 在本文中指  $PM_{2.5}$  月均浓度,  $\mu g/m^3$ ;  $k$  和  $b$  分别表示斜率和截距。

### 1.2.2.2 多项式拟合模型

对线性拟合关系不显著的省份进行组合, 按照趋势划分后, 采用多项式拟合方法进行分析。多项式拟合模型函数为

$$y = I + B_1 \times x + B_2 \times x^2 \quad (3)$$

式中:  $I$  为截距,  $B_1$  和  $B_2$  表示系数。

### 1.2.3 模型精确度验证

对数学模型的精确度进行评估, 若性能良好, 则此模型可用于  $PM_{2.5}$  浓度预测。在模型评价中, 使用各省份 2018—2020 年  $PM_{2.5}$  月均浓度数据, 选取均方根误差 (RMSE) 和  $R^2$ <sup>[37]</sup> 对数学模型的性能进行评价。其中: RMSE 用于评价真实值与预测值之间的偏差, 反映模型对较大误差的敏感程度;  $R^2$  表示因变量的变异中能被自变量解释的比例, 是衡量拟合优度的指标<sup>[38-39]</sup>。RMSE 越接近 0, 表明模型精度越高;  $R^2$  越接近 1, 表明模型的预测能力越好。两指标计算公式分别见公式 (4) 和公式 (5)。

$$RMSE = \sqrt{\frac{1}{N} \sum_{m=1}^N (Y_m - \hat{Y}_m)^2} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{m=1}^N (Y_m - \hat{Y}_m)^2}{\sum_{m=1}^N (Y_m - \bar{Y})^2} \quad (5)$$

式中:  $Y_m$  为第  $m$  个样本的  $PM_{2.5}$  实际浓度,  $\mu g/m^3$ ;  $\hat{Y}_m$  为第  $m$  个样本的  $PM_{2.5}$  预测浓度,  $\mu g/m^3$ ;  $\bar{Y}$  为预测浓度的平均值,  $\mu g/m^3$ ;  $N$  为总样本数。

## 2 结果与讨论

### 2.1 分布规律省份拟合结果

为了达到良好的预测效果, 选择 0.70 为截点对相关系数进行划分。结果显示, 满足  $P < 0.05$  且相关系数大于 0.70 的省份共有 19 个 (图 1)。青海、江西和福建等省份虽然  $P < 0.05$ , 但相关系

数小于0.70(约为0.40),表明其散点分布规律性不强。广东、海南等省份  $P>0.05$ ,表明其PM<sub>2.5</sub>排放量与平均浓度不相关。综上,将剩余10个省份判定为散点分布分散省份。

将散点分布规律的19个省份的2013—2017年PM<sub>2.5</sub>月排放量与月均浓度数据进行线性拟合(图2)。拟合结果显示,有8个省份的 $R^2$ 介于0.70~0.90之间,另有8个省份的 $R^2$ 介于0.60~0.70之间,表明此模型的拟合效果良好。以上19个省份散点分布规律的原因大致可以归结为:其排放源排放出的PM<sub>2.5</sub>受到的气候、地形等因素的作用是相似的。

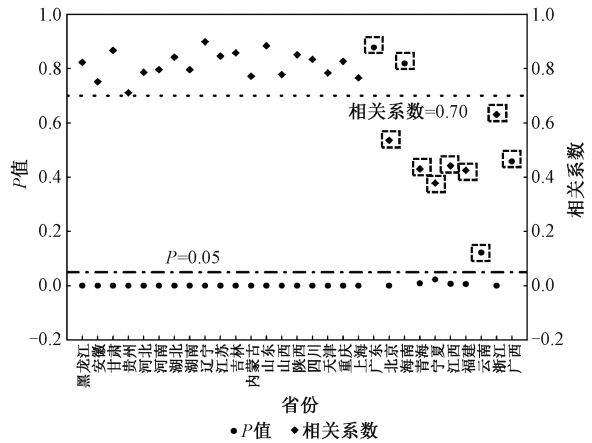


图1 省份划分结果

Fig. 1 Results of provinces division

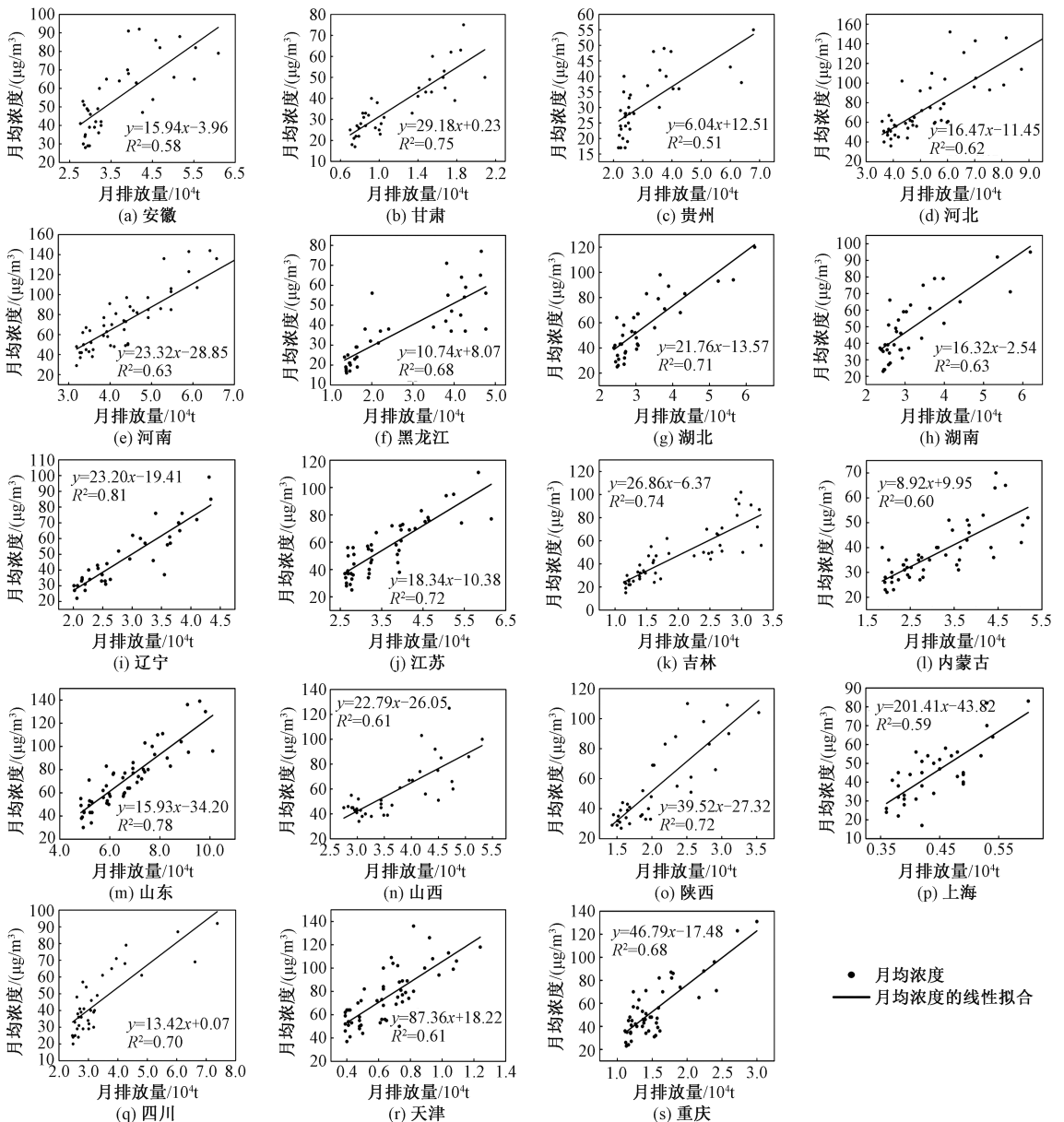


图2 分布规律省份2013—2017年PM<sub>2.5</sub>浓度的线性拟合

Fig. 2 The linear fit of PM<sub>2.5</sub> from 2013 to 2017 in provinces with distribution patterns

### 2.2 分布分散省份拟合结果

经 Pearson 相关性分析,最终得到 10 个散点分布分散的省份,分别为广东、北京、海南、青海、宁夏、江西、福建、云南、浙江和广西。上述省份可能由于在某一时间范围内受气象、地形等因素影响程度差别较大,所以呈不规则表现。对于分布分散省份,为了便于分析,按照月份进行划分,将在相同月份具有大致相同趋势的省份组合,然后分别进行分析。将每个省份(共 10 个)1—12 月的散点合并到一起,共分为 12 个大组。因为大组中的散点分布

较为分散,需要按照其中的分布规律进行划分。

由于涉及的月份较多,以 5 月(图 3)和 10 月(图 4)为例进行分析。5 月的散点组合共可划分成 5 个小组,各小组均呈规律分布且可以进行拟合。可以看出,5 个小组的  $R^2$  介于 0.56~0.99 之间(表 1),拟合效果良好。10 月的散点组合共可划分成 6 个小组,6 个小组的  $R^2$  介于 0.88~0.96 之间(表 1),拟合效果良好。对于划分后的小组,根据其散点分布特征选择恰当的拟合方法。其余月份的散点划分采用相同方式。

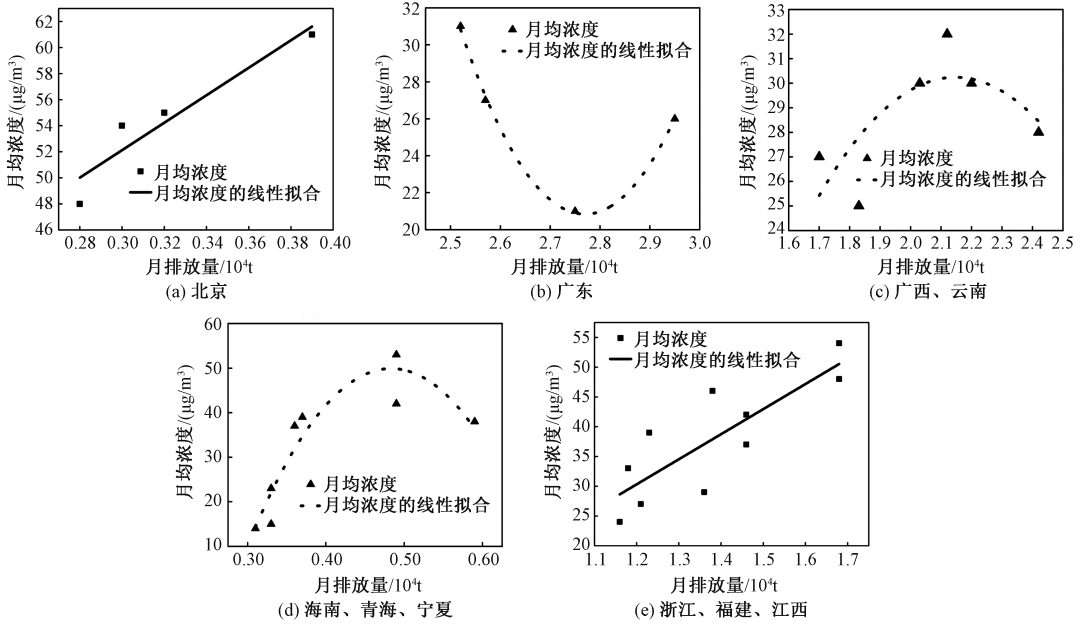


图 3 5 月分组省份拟合结果

Fig. 3 The fitting results of divided provinces in May

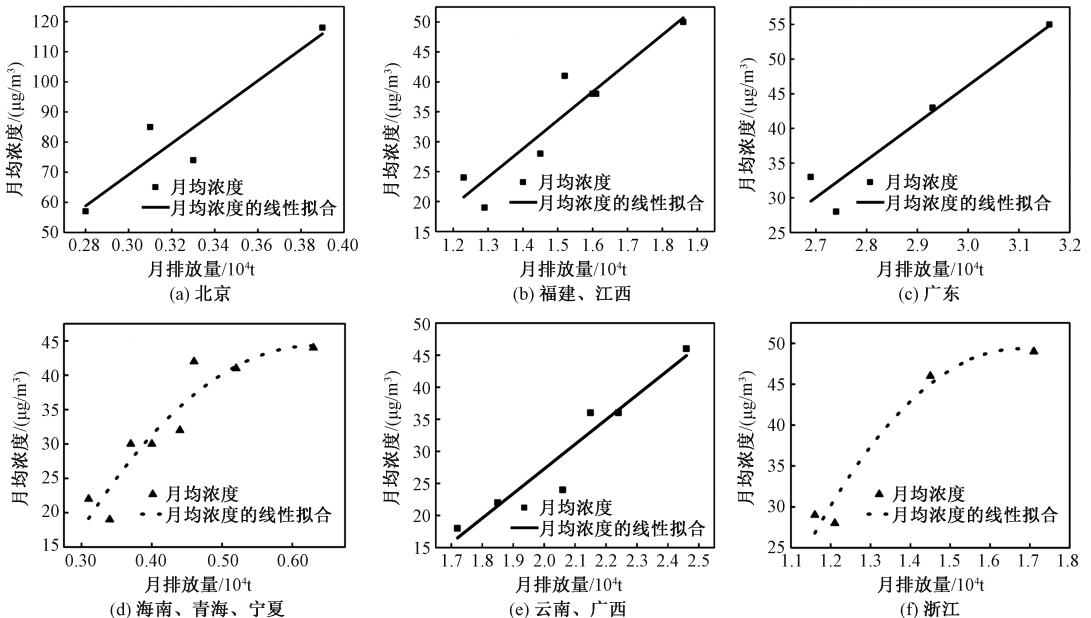


图 4 10 月分组省份拟合结果

Fig. 4 The fitting results of divided provinces in October

表 1 1—12 月划分省份的 PM<sub>2.5</sub> 月排放量和月均浓度拟合方程  
 Table 1 PM<sub>2.5</sub> monthly emission and average concentration fitting equations by province from January to December

月份	省份分组	方程	R <sup>2</sup>	月份	省份分组	方程	R <sup>2</sup>
1	福建、浙江	$y = 86.15x - 66.97$	0.98	7	北京	$y = 127.61 - 654.17x + 1448.23x^2$	0.81
	广西	$y = 8368.00 - 7773.19x + 1811.59x^2$	1.00		海南、宁夏	$y = 92.67x - 18.63$	0.94
	海南、青海、宁夏、北京	$y = 97.82x + 8.24$	0.72		浙江、江西	$y = 15.93x + 6.56$	0.88
	广东	$y = 3826.81 - 3278.84x + 709.66x^2$	0.92		云南、广西	$y = 15.91x - 11.01$	0.83
	江西	$y = 2737.59 - 2984.25x + 827.92x^2$	1.00		福建	$y = -494.40 + 807.68x - 315.48x^2$	1.00
	云南	$y = 22.22x - 24.33$	1.00		广东	$y = -457.53 + 347.22x - 62.52x^2$	0.95
2	福建、云南、广西	$y = 211.14 - 250.62x + 88.40x^2$	0.91	8	青海	$y = 178.95x - 32.37$	0.99
	江西	$y = -18.58x + 89.84$	0.96		北京	$y = 225.00x - 24.00$	0.93
	海南、青海、宁夏、北京	$y = 126.38x - 5.23$	0.72		海南、青海	$y = 307.48x - 87.99$	0.85
	广东	$y = -1533.11 + 1404.84x - 311.30x^2$	0.67		云南、广西	$y = 126.61 - 127.56x + 37.18x^2$	0.89
	浙江	$y = -71.69 + 179.16x - 62.56x^2$	0.88		浙江、江西	$y = 23.32x - 3.30$	0.84
3	北京	$y = -437.44 + 2243.21x - 2351.39x^2$	0.56	9	福建	$y = 232.82 - 323.46x + 123.01x^2$	0.99
	福建、云南、广西	$y = 43.98 - 15.80x + 6.78x^2$	0.53		广东	$y = -1609.50 + 1194.31x - 217.51x^2$	0.83
	青海、宁夏、海南	$y = -86.94 + 502.55x - 462.76x^2$	0.77		宁夏	$y = 439.64 - 1504.55x + 1363.64x^2$	1.00
	广东	$y = 939.60 - 695.92x + 133.87x^2$	0.74		浙江、江西	$y = -125.26 + 203.92x - 62.63x^2$	0.89
	江西	$y = -1738.77 + 1886.00x - 494.74x^2$	1.00		福建、广东	$y = 4.47x + 18.10$	0.87
	浙江	$y = 324.76 - 380.50x + 129.19x^2$	0.55		海南、青海、宁夏	$y = -94.79 + 483.63x - 460.74x^2$	0.71
4	北京	$y = 328.57x - 31.36$	0.93	10	广西	$y = -3602.96 + 3268.63x - 732.20x^2$	1.00
	福建、广西	$y = 9.60x + 19.30$	0.86		云南	$y = -700.24 + 772.07x - 205.79x^2$	1.00
	海南、青海、宁夏	$y = -74.62 + 430.93x - 397.29x^2$	0.74		北京	$y = 496.68 - 2759.46x + 4270.79x^2$	0.96
	云南	$y = -155.05 + 206.39x - 56.93x^2$	1.00		福建、江西	$y = 47.55x - 37.73$	0.88
	广东	$y = 538.35 - 411.18x + 83.41x^2$	0.99		海南、青海、宁夏	$y = -53.00 + 308.66x - 244.79x^2$	0.88
	江西	$y = 684.36 - 857.95x + 284.09x^2$	1.00		云南、广西	$y = 38.40x - 49.54$	0.92
5	浙江	$y = -19.90 + 70.54x - 16.81x^2$	0.96	11	北京	$y = 519.69x - 86.70$	0.88
	浙江、福建、江西	$y = 42.10x - 20.20$	0.69		广东	$y = 53.88x - 115.42$	0.93
	广西、云南	$y = -80.81 + 103.44x - 24.09x^2$	0.56		浙江	$y = -188.84 + 284.27x - 84.82x^2$	0.96
	海南、青海、宁夏	$y = -227.06 + 1146.15x - 1185.80x^2$	0.87		北京	$y = -2066.22 + 5931.89x - 4012.36x^2$	0.98
	广东	$y = 1248.55 - 886.44x + 160.01x^2$	0.99		广东	$y = 1156.97 - 754.09x + 126.74x^2$	0.99
6	北京	$y = 105.45x + 20.49$	0.85	12	海南、青海、宁夏	$y = 126.47x - 25.14$	0.88
	北京	$y = -643.10 + 4182.60x - 6180.94x^2$	0.89		江西、广西	$y = 145.43 - 78.08x + 13.45x^2$	0.68
	福建、江西	$y = 27.12x - 13.74$	0.75		福建	$y = 521.22 - 711.10x + 254.41x^2$	0.96
	广东	$y = 1062.20 - 762.20x + 138.75x^2$	0.94		云南	$y = 337.54 - 315.87x + 78.92x^2$	1.00
	海南、青海、宁夏	$y = -125.40 + 651.93x - 650.61x^2$	0.75		浙江	$y = 232.55 - 269.50x + 94.89x^2$	0.99
	浙江	$y = 29.86x - 7.58$	0.98		北京、青海、宁夏、海南	$y = 71.57 - 315.40x + 526.97x^2$	0.63
	广西	$y = 380.96 - 314.71x + 68.37x^2$	1.00	福建	$y = 762.53 - 1013.96x + 350.19x^2$	0.88	
	云南	$y = 955.57 - 1080.49x + 310.02x^2$	1.00	江西、广西	$y = -47.31x - 173.39$	0.82	
				云南	$y = 240.48 - 150.27x + 26.67x^2$	1.00	
				浙江	$y = -9.27 + 84.37x - 23.54x^2$	0.19	
				广东	$y = 2678.43 - 1672.57x + 264.63x^2$	0.88	

12 个月份的拟合分组结果中(表 1),大多数小组具有良好的  $R^2$ ,其值在 0.60~0.80 范围占较大比重,表明拟合得到的数学模型可以应用到未来的 PM<sub>2.5</sub> 浓度预测中。极少数省份小组拟合得到的  $R^2$  介于 0.50~0.60 之间,其值略低于平均水平,但仍适用于未来的 PM<sub>2.5</sub> 浓度预测。12 月份浙江省拟合方程的  $R^2$  为 0.19,数值较小,不具有统计学意义,故判定为无效拟合,不参与后续讨论。

### 2.3 模型精度验证

采用 RMSE 和  $R^2$  来评价数学模型的精度:RMSE 越小说明模型精度越高, $R^2$  越大说明模型精度越高。对于分布规律和分布分散两类省份,采取不同的验证方法。对于分布规律省份,按照省份分别进行验证。将 19 个省份 2018—2020 年 PM<sub>2.5</sub> 月排放量数据分别输入到对应的数学模型中,从而得到预测浓度数据。将预测数据与真实数据先进行归一化处理,再进行拟合。从拟合

结果可以看出(图5),10个省份(占该类别的比例为57.89%)的 $R^2$ 高于0.60,8个省份(占比42.11%)介于0.50~0.60之间,仅贵州省(0.49)低于0.50。14个省份(占比73.68%)的RMSE

数值范围为6.00~16.00,3个省份(占比15.79%)为16.00~20.00,2个省份(占比10.53%)为20.00~25.00。可以看出,大多数分布规律省份的数学模型的预测效果良好。

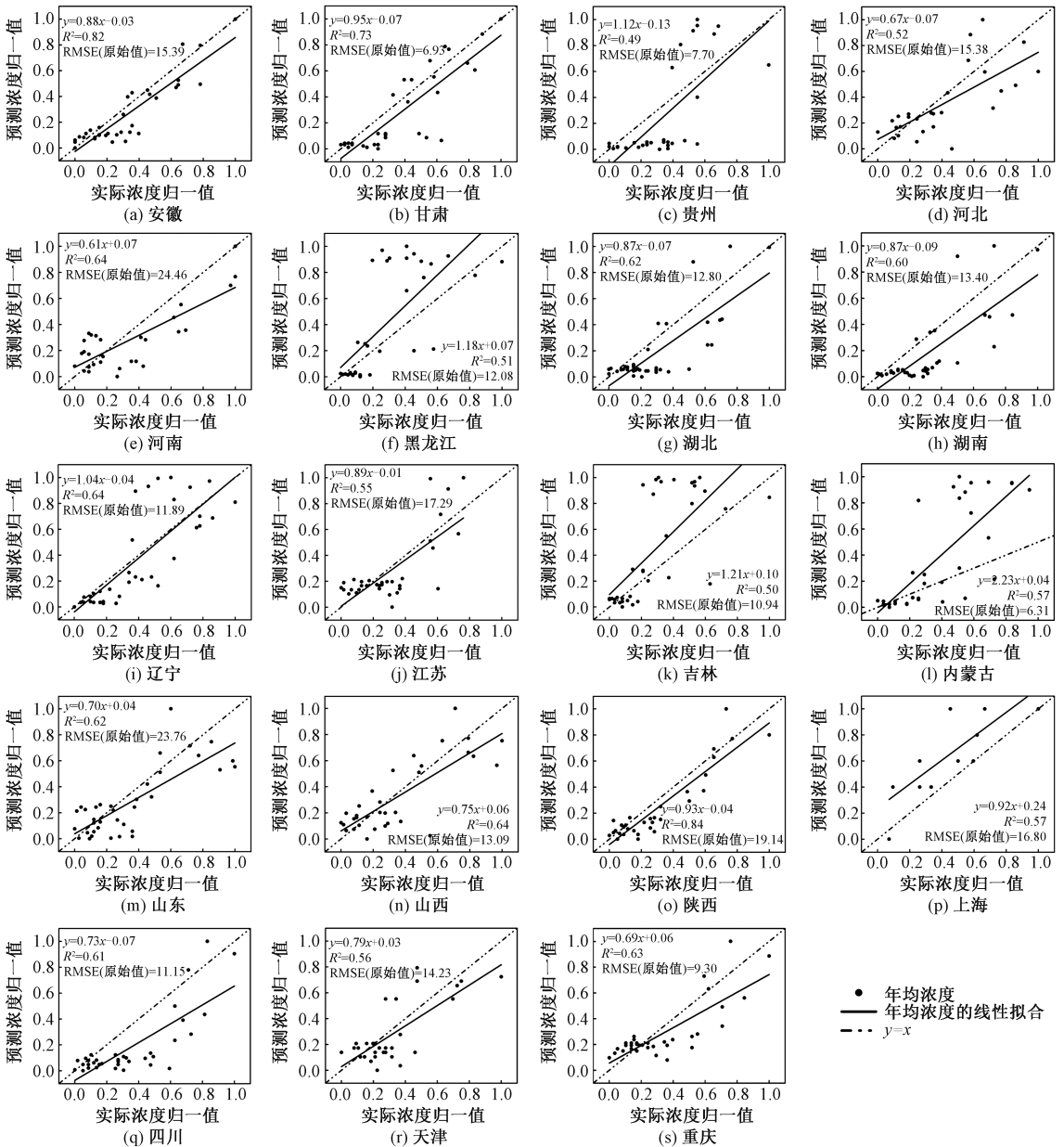


图5 分布规律省份模型验证结果

Fig. 5 Model validation results for provinces with distribution patterns

将分布分散省份2018—2020年 $PM_{2.5}$ 月排放量输入到数学模型中,得到预测浓度数据。采取分组的方法对所得数学模型进行验证,共分为10组,每组均包括所有分布分散的省份,各个省份对应的时间节点不同。

从分布分散省份模型验证结果可以看出(图6):5个小组(占该类别的比例为50%)的 $R^2$ 介

于0.50~0.80之间,4个小组的 $R^2$ (占比40%)介于0.50~0.60之间,只有第1组(占比10%)的 $R^2$ 低于0.50;RMSE数值介于6.00~16.00之间的小组有8个(占该类别的比例为80%),剩余两个小组的RMSE值分别为16.92和24.81。综上所述可知,大多数分布分散省份的 $PM_{2.5}$ 月均浓度预测效果良好。

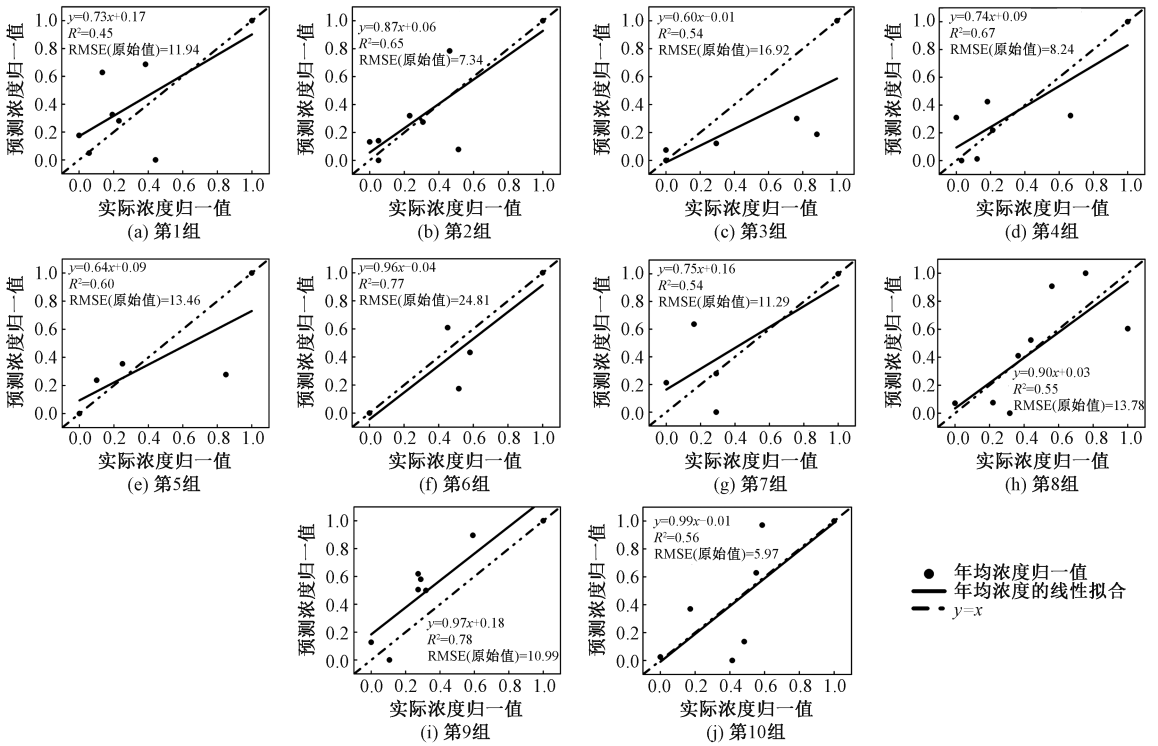


图6 分布分散省份模型验证结果

Fig. 6 Model validation results for provinces with decentralized distribution

为验证所得数学模型对PM<sub>2.5</sub>浓度的预测能力,对模型预测结果进行时序和空间分析。对于时序分析,选取2018—2020年PM<sub>2.5</sub>逐月预测浓度与真实浓度进行对比。由于涉及的省份较多,分别选取甘肃(分布规律)和宁夏(分布分散)两个省份进行分析(图7)。分析结果表明,两类数学模型较好地预测了PM<sub>2.5</sub>浓度的变化趋势,预

测值与真实值较为契合,波峰和波谷出现的时间也大体一致,可较好地反映PM<sub>2.5</sub>浓度的季节性变化特征。对于浓度峰值预测结果存在的偏低现象,从图7可以看出,其出现的时间主要集中在冬季,此时的相对湿度较高,相较于其他季节更有利于PM<sub>2.5</sub>的吸湿增长,进而出现雾霾现象,最终导致预测浓度低于观测浓度。

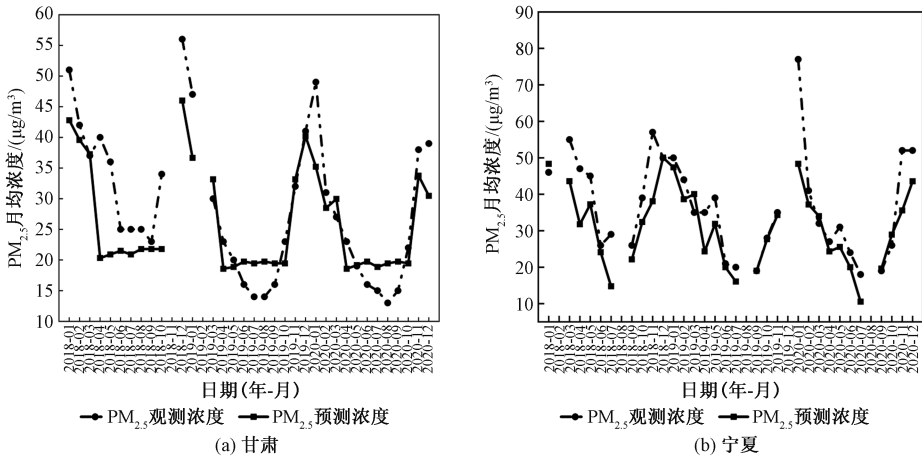


图7 基于数学模型的浓度预测结果

Fig. 7 Plot of concentration prediction results based on mathematical model

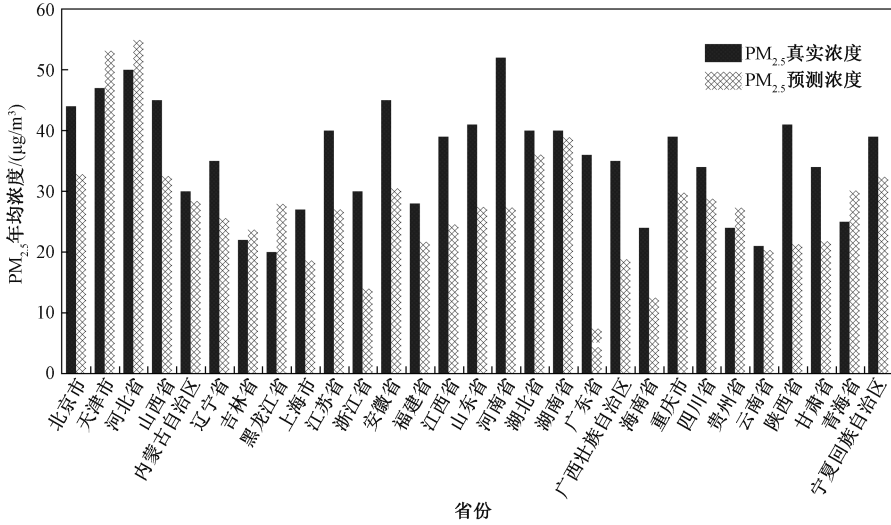
空间分布对比以2018年12月数据为例,将PM<sub>2.5</sub>真实浓度与预测浓度以柱状图的形式

展示,具体见图8。总体上看,大多数省份的PM<sub>2.5</sub>真实浓度高于预测浓度,这种情况在浙



江、广西和广东等省份尤为明显;预测效果较好的省份包括吉林和贵州,其真实浓度与预测浓度的差值(绝对值)分别为  $1.71 \mu\text{g}/\text{m}^3$  和

$3.31 \mu\text{g}/\text{m}^3$ 。因此,从整体上看,各省份的数学模型可以较准确地预测其未来的  $\text{PM}_{2.5}$  浓度变化。



注:由于数据缺失,图内不包含新疆、西藏、台湾、香港、澳门状况。

图8 2018年12月全国各省份真实浓度与预测浓度分布

Fig. 8 Distribution of real and predicted concentrations in each province in December 2018

### 3 结论

1)使用 Pearson 相关性分析方法判别各省份  $\text{PM}_{2.5}$  月排放量与月均浓度散点的分布情况,将  $P < 0.05$  且相关系数大于 0.70 的省份定义为分布规律省份,否则为分布分散省份。最后,共得到分布规律省份 19 个,分布分散省份 10 个。

2)对于散点分布规律的省份,通过对 2013—2017 年  $\text{PM}_{2.5}$  月排放量与月均浓度进行线性拟合建立数学模型。拟合结果显示,有 8 个省份的  $R^2$  介于 0.70~0.90 之间,另有 8 个省份的  $R^2$  介于 0.60~0.70 之间,表明模型拟合效果良好。模型精度验证结果显示:10 个省份的  $R^2$  大于 0.60,8 个省份的  $R^2$  介于 0.50~0.60 之间;14 个省份的 RMSE 介于 6.00~16.00 之间,3 个省份介于 16.00~20.00 之间。上述结果说明,该数学模型对于未来  $\text{PM}_{2.5}$  浓度的预测具有一定的有效性。

3)对于散点分布分散的省份,按照月份划分成 12 个大组进行线性拟合和多项式拟合。所得结果中,大多数组别的  $R^2$  在合理范围内,其中在 0.60~0.80 范围内的占比较大,表明拟合得到的数学模型可以应用到对未来  $\text{PM}_{2.5}$  浓度的预测中。采用分组方法进行模型精度验证,共分为 10

组。5 个小组的  $R^2$  介于 0.50~0.80 之间,4 个小组的  $R^2$  介于 0.50~0.60 之间;8 个小组的 RMSE 介于 6.00~16.00 之间。由此可以看出,大多数散点分布分散的省份的  $\text{PM}_{2.5}$  月均浓度预测效果良好。

4)探讨长时间序列  $\text{PM}_{2.5}$  排放量与浓度的关系具有重要意义,特别是以月为时间单元,有助于加强对月尺度环境气象条件的预报和评估,而地方政府也可据此来提前制定计划,及时采取相应的污染排放管控措施,从而为当地大气环境质量管控提供科学指导。

### 参考文献 (References):

- [1] ZHANG B, LIU Y, YONG R, et al. A Spatial Correlation Prediction Model of Urban  $\text{PM}_{2.5}$  Concentration Based on Deconvolution and LSTM [J]. Neurocomputing, 2023, 544: 126280.
- [2] YANG H, WANG W, LI G. Prediction Method of  $\text{PM}_{2.5}$  Concentration Based on Decomposition and Integration [J]. Measurement, 2023, 216: 112954.
- [3] 陈强,梅琨,朱慧敏,等.郑州市  $\text{PM}_{2.5}$  浓度时空分布特征及预测模型研究 [J]. 中国环境监测, 2015, 31(3): 105-112.  
CHEN Qiang, MEI Kun, ZHU Huimin, et al. Study on Spatiotemporal Variability of  $\text{PM}_{2.5}$  Concentrations

- and Prediction Model over Zhengzhou City [ J ]. Environmental Monitoring in China, 2015, 31 ( 3 ) : 105-112.
- [ 4 ] 黄春桃,范东平,卢集富,等.基于深度学习模型的广州市大气 PM<sub>2.5</sub> 和 PM<sub>10</sub> 浓度预测[J].环境工程,2021,39(12):135-140.  
HUANG Chuntao, FAN Dongping, LU Jifu, et al. Prediction of PM<sub>2.5</sub> and PM<sub>10</sub> Concentration in Guangzhou Based on Deep Learning Model [ J ]. Environmental Engineering, 2021, 39(12) :135-140.
- [ 5 ] 范丽行,杨晓辉,宋春杰,等.基于时空混合效应模型的京津冀 PM<sub>2.5</sub> 浓度变化模拟[J].环境科学,2022,43(5):2 262-2 273.  
FAN Lihang, YANG Xiaohui, SONG Chunjie, et al. Modeling of PM<sub>2.5</sub> Concentrations in the Beijing-Tianjin-Hebei Region Using a Space-Time Linear Mixed Effects Model [ J ]. Environmental Science, 2022, 43(5) :2 262-2 273.
- [ 6 ] ZHANG Q, WU S, WANG X, et al. A PM<sub>2.5</sub> Concentration Prediction Model Based on Multi-task Deep Learning for Intensive Air Quality Monitoring Stations [ J ]. Journal of Cleaner Production, 2020, 275:122722.
- [ 7 ] 顾阔,焦瑞莉,薄宇,等.基于复合 LSTM 模型的 PM<sub>2.5</sub> 浓度预测[J].中国环境监测,2023,39(1):170-180.  
GU Kuo, JIAO Ruili, BO Yu, et al. PM<sub>2.5</sub> Concentration Prediction Based on the Composite LSTM Model [ J ]. Environmental Monitoring in China, 2023, 39(1) :170-180.
- [ 8 ] GONG J, DING L, LU Y, et al. Scientometric and Multidimensional Contents Analysis of PM<sub>2.5</sub> Concentration Prediction [ J ]. Heliyon, 2023, 9 ( 3 ) : e14526.
- [ 9 ] HAN D, ZHANG T, ZHANG X, et al. Study on Spatiotemporal Characteristics and Influencing Factors of Pedestrian-Level PM<sub>2.5</sub> Concentrations in Outdoor Open Spaces of Harbin in Winter, Using a Generalized Additive Model ( GAM ) [ J ]. Urban Climate, 2022, 46: 101313.
- [ 10 ] NI X Y, HUANG H, DU W P. Relevance Analysis and Short-Term Prediction of PM<sub>2.5</sub> Concentrations in Beijing Based on Multi-source Data [ J ]. Atmospheric Environment, 2017, 150: 146-161.
- [ 11 ] SHARMA E, DEO R C, PRASAD R, et al. A Hybrid Air Quality Early-Warning Framework: An Hourly Forecasting Model with Online Sequential Extreme Learning Machines and Empirical Mode Decomposition Algorithms [ J ]. Science of the Total Environment, 2020, 709: 135934.
- [ 12 ] 马俊文,严京海,孙瑞雯,等.基于 LSTM-GCN 的 PM<sub>2.5</sub> 浓度预测模型[J].中国环境监测,2022,38(5):153-160.  
MA Junwen, YAN Jinghai, SUN Ruiwen, et al. Prediction Model of PM<sub>2.5</sub> Concentration Based on LSTM-GCN [ J ]. Environmental Monitoring in China, 2022, 38(5) :153-160.
- [ 13 ] ZHANG L, WANG J, LI Z, et al. Uncertainty Quantification of PM<sub>2.5</sub> Concentrations Using a Hybrid Model Based on Characteristic Decomposition and Fuzzy Granulation [ J ]. Journal of Environmental Management, 2022, 324: 116282.
- [ 14 ] ZHU J, DENG F, ZHAO J, et al. Attention-Based Parallel Networks ( APNet ) for PM<sub>2.5</sub> Spatiotemporal Prediction [ J ]. Science of the Total Environment, 2021, 769: 145082.
- [ 15 ] YANG H, LIU Z, LI G. A New Hybrid Optimization Prediction Model for PM<sub>2.5</sub> Concentration Considering Other Air Pollutants and Meteorological Conditions [ J ]. Chemosphere, 2022, 307: 135798.
- [ 16 ] SUN W, ZHANG H, PALAZOGLU A, et al. Prediction of 24-Hour-Average PM<sub>2.5</sub> Concentrations Using a Hidden Markov Model with Different Emission Distributions in Northern California [ J ]. Science of the Total Environment, 2013, 443: 93-103.
- [ 17 ] LIU H, DUAN Z, CHEN C. A Hybrid Multi-resolution Multi-objective Ensemble Model and Its Application for Forecasting of Daily PM<sub>2.5</sub> Concentrations [ J ]. Information Sciences, 2020, 516: 266-292.
- [ 18 ] BERA B, BHATTACHARJEE S, SENGUPTA N, et al. PM<sub>2.5</sub> Concentration Prediction During COVID-19 Lockdown over Kolkata Metropolitan City, India Using MLR and ANN Models [ J ]. Environmental Challenges, 2021, 4: 100155.
- [ 19 ] WANG P, ZHANG H, QIN Z, et al. A Novel Hybrid-Garch Model Based on ARIMA and SVM for PM<sub>2.5</sub> Concentrations Forecasting [ J ]. Atmospheric Pollution Research, 2017, 8(5) : 850-860.
- [ 20 ] DAI H, HUANG G, ZENG H, et al. PM<sub>2.5</sub> Volatility Prediction by XGBoost-MLP Based on GARCH Models [ J ]. Journal of Cleaner Production, 2022, 356: 131898.
- [ 21 ] BAI W, LI F Y. PM<sub>2.5</sub> Concentration Prediction Using Deep Learning in Internet of Things Air Monitoring System [ J ]. Environmental Engineering Research, 2023, 28(1) : 210456.

- [22] XU H, CHEN H. Impact of Urban Morphology on the Spatial and Temporal Distribution of  $PM_{2.5}$  Concentration: A Numerical Simulation with WRF/CMAQ Model in Wuhan, China [J]. *Journal of Environmental Management*, 2021, 290: 112427.
- [23] HAO X, HU X, LIU T, et al. Estimating Urban  $PM_{2.5}$  Concentration: An Analysis on the Nonlinear Effects of Explanatory Variables Based on Gradient Boosted Regression Tree [J]. *Urban Climate*, 2022, 44: 101172.
- [24] WU W, ZHANG M, DING Y. Exploring the Effect of Economic and Environment Factors on  $PM_{2.5}$  Concentration: A Case Study of the Beijing-Tianjin-Hebei Region [J]. *Journal of Environmental Management*, 2020, 268: 110703.
- [25] YUN G, ZHAO S. The Imprint of Urbanization on  $PM_{2.5}$  Concentrations in China: The Urban-Rural Gradient Study [J]. *Sustainable Cities and Society*, 2022, 86: 104103.
- [26] YU M Y, CAI X H, SONG Y, et al. A Fast Forecasting Method for  $PM_{2.5}$  Concentrations Based on Footprint Modeling and Emission Optimization [J]. *Atmospheric Environment*, 2019, 219: 117013.
- [27] MA X Y, SHA T, WANG J Y, et al. Investigating Impact of Emission Inventories on  $PM_{2.5}$  Simulations over North China Plain by WRF-Chem [J]. *Atmospheric Environment*, 2018, 195: 125-140.
- [28] CHUAI X W, LU Y, XIE F J, et al. A New Approach to Evaluate Regional Inequity Determined by  $PM_{2.5}$  Emissions and Concentrations [J]. *Journal of Environmental Management*, 2021, 277: 111335.
- [29] ZHANG X L, SHI M J, LI Y J, et al. Correlating  $PM_{2.5}$  Concentrations with Air Pollutant Emissions: A Longitudinal Study of the Beijing-Tianjin-Hebei Region [J]. *Journal of Cleaner Production*, 2018, 179: 103-113.
- [30] 党莹, 张小玲, 饶晓琴, 等. 北京与成都大气污染特征及空气质量改善效果评估 [J]. *环境科学*, 2021, 42(8): 3 622-3 632.  
DANG Ying, ZHANG Xiaoling, RAO Xiaoqin, et al. Evaluation of Air Pollution Characteristics and Air Quality Improvement Effect in Beijing and Chengdu [J]. *Environmental Science*, 2021, 42(8): 3 622-3 632.
- [31] LI M, LIU H, GENG G N, et al. Anthropogenic Emission Inventories in China: A Review [J]. *National Science Review*, 2017, 4(6): 834-866.
- [32] ZHENG B, TONG D, LI M, et al. Trends in China's Anthropogenic Emissions Since 2010 as the Consequence of Clean Air Actions [J]. *Atmospheric Chemistry and Physics*, 2018, 18(19): 14 095-14 111.
- [33] MEIC 团队. 中国污染物排放 [EB/OL]. (2022-09-26) [2024-02-01]. [http://meicmodel.org.cn/?page\\_id=560](http://meicmodel.org.cn/?page_id=560).
- [34] ZHANG C, WANG X, CHEN S, et al. Coupling Detrended Fluctuation Analysis of the Relationship Between  $PM_{2.5}$  Concentration and Weather Elements [J]. *Physica A: Statistical Mechanics and Its Applications*, 2019, 531: 121757.
- [35] WANG Y, WANG H, ZHANG S. Prediction of Daily  $PM_{2.5}$  Concentration in China Using Data-Driven Ordinary Differential Equations [J]. *Applied Mathematics and Computation*, 2020, 375: 125088.
- [36] LIM C H, RYU J, CHOI Y, et al. Understanding Global  $PM_{2.5}$  Concentrations and Their Drivers in Recent Decades (1998-2016) [J]. *Environment International*, 2020, 144: 106011.
- [37] YANG H, WANG C, LI G. A New Combination Model Using Decomposition Ensemble Framework and Error Correction Technique for Forecasting Hourly  $PM_{2.5}$  Concentration [J]. *Journal of Environmental Management*, 2022, 318: 115498.
- [38] YU Y, LI H, SUN S, et al.  $PM_{2.5}$  Concentration Forecasting Through a Novel Multi-scale Ensemble Learning Approach Considering Intercity Synergy [J]. *Sustainable Cities and Society*, 2022, 85: 104 049.
- [39] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-Learn: Machine Learning in Python [J]. *Journal of Machine Learning Research*, 2011, 12: 2 825-2 830.