

基于组合模型的 PM_{2.5} 浓度预测及其不确定性分析

王学梅^{1,2} 王凤文^{1*} 陈 滔³ 张庆国⁴ 蒋跃林¹

(1. 安徽农业大学 资源与环境学院, 合肥 230036; 2. 安徽省宣城市气象局, 安徽 宣城 242000;
3. 安徽农业大学 工学院, 合肥 230036; 4. 安徽农业大学 理学院, 合肥 230036)

摘要: 运用 GIS 软件及克里金(Kriging)插值等方法分析合肥城市圈 PM_{2.5} 浓度的时空分布, 根据合肥市环境监测历史数据、地面气象站点数据及历史气象数据, 采用多元回归分析、相关分析等方法, 研究合肥市 PM_{2.5} 浓度的影响因素。结果表明: 1) PM_{2.5} 浓度整体变化情况为冬季>秋季>春季>夏季, 大部分城市 PM_{2.5} 浓度峰值出现在 1 月, 之后浓度开始逐渐下降, 7 月达到最低值, 此后浓度逐渐升高。2) PM_{2.5} 浓度与 CO 呈高度正相关, 相关系数高达 0.875; 与 PM₁₀、SO₂、NO₂ 的相关性也较高; 与 O₃ 呈负相关。PM_{2.5} 浓度与气压、风速、降雨量以及能见度呈负相关, 与温度、相对湿度呈强正相关。基于 2018—2019 年合肥市地面站点 PM_{2.5} 浓度监测数据, 构建预测 PM_{2.5} 浓度的组合模型: 对比三次指数平滑模型, 确定模拟退火+遗传+三次指数平滑为优组合模型, 拟合度达到 95%。通过 Kappa 及 MAPE 指数对组合模型不确定性进行分析评价, 两者分别为 0.654 和 0.072, 说明该模型具有高度稳定性。恰当的预测因子组合和模型不确定性研究有助于模型预测精度的提升和改善, 从而为大气环境质量监测和评价提供参考。

关键词: PM_{2.5} 浓度; 时空分布; 预测; 组合模型; 不确定性分析

DOI: 10.13205/j.hjgc.202008038

PM_{2.5} CONCENTRATION PREDICTION AND UNCERTAINTY ANALYSIS BASED ON A COMPOSITE MODEL

WANG Xue-mei^{1,2}, WANG Feng-wen^{1*}, CHEN Tao³, ZHANG Qing-guo⁴, JIANG Yue-lin¹

(1. School of Resources and Environment, Anhui Agricultural University, Hefei 230036, China; 2. Xuancheng Meteorological Bureau, Xuancheng 242000, China; 3. School of Technology, Anhui Agricultural University, Hefei 230036, China;
4. School of Science, Anhui Agricultural University, Hefei 230036, China)

Abstract: In this paper, GIS software and Kriging interpolation method were used to analyze the spatial and temporal distribution of PM_{2.5} concentration in Hefei city circle. According to historical environment monitoring data, ground meteorological stations and historical meteorological data of Hefei, multiple regression analysis, correlation analysis and other methods were adopted to study the influencing factors of PM_{2.5} concentration in Hefei. The results showed that: 1) the overall change of PM_{2.5} concentration in the above cities was in the order of winter>autumn>spring>summer, and PM_{2.5} concentration in most cities peaked in January, then gradually declined, reached the lowest value in July, and then gradually increased; 2) PM_{2.5} concentration showed a highly positive correlation with CO, with the correlation coefficient as high as 0.875. The correlation with PM₁₀, SO₂ and NO₂ was also high. There was a negative correlation with O₃. PM_{2.5} concentration was negatively correlated with air pressure, wind speed, rainfall and visibility, and strongly positively correlated with temperature and relative humidity. Based on PM_{2.5} concentration monitoring data of Hefei from 2018 to 2019, a composite model was built to predict PM_{2.5} concentration, and three times exponential smoothing model was compared to determine that simulated annealing+genetic+three times exponential smoothing was the optimal composite model, with the fitting degree reaching 95%.

收稿日期: 2019-11-14

基金项目: 国家重点基础研究发展计划(2018YFD0300905); 安徽省教育厅重点科研项目(KJ2019A0208)。

第一作者: 王学梅, 女, 硕士, 主要研究方向为应用气象及计算机模拟。xmwang@ahau.edu.cn

* 通信作者: 王凤文(1974-), 男, 博士, 讲师, 主要研究方向为应用气象、环境监测。wfw2008@ahau.edu.cn

©1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Kappa and MAPE indexes were used to analyze and evaluate the uncertainty of the composite model. Kappa and MAPE indexes were 0.654 and 0.072 respectively, indicating that the model was highly stable. The proper combination of prediction factors and the study of model uncertainty were helpful to improve the model prediction accuracy, so as to provide theoretical basis and method for the monitoring and evaluation of atmospheric environment quality.

Keywords: PM_{2.5} concentrations; spatial and temporal distribution; prediction; composite model; uncertainty analysis

0 引 言

随着我国社会经济的快速发展以及城市化进程的持续加快,环境污染的压力越来越大。环境空气质量作为生态安全体系的重要指标之一,越来越受到关注^[1]。PM_{2.5}是指空气动力学当量直径 $\leq 2.5 \mu\text{m}$ 的细颗粒物,会对人类的生活和身体健康造成严重危害^[2]。PM_{2.5}浓度的预测对大气环境质量监测和评价具有重要的现实意义^[3-4]。李松等^[5]利用 GIS 对贵州省遵义市 PM_{2.5}浓度变化特征进行了分析,发现 PM_{2.5}浓度的季节性变化特征明显,并且其日变化受到人类活动的影响。任才溶^[6]利用随机森林算法,基于 Spark 平台建立 PM_{2.5}浓度值预测模型和等级预测模型,对所建立的模型预测结果进行了评价。但 PM_{2.5}浓度的精准预测却很难实现,且建立模型后对该模型的稳定性加以评价的研究尚不多见^[7]。

本研究利用 2018—2019 年合肥城市圈空气质量监测站的日均 PM_{2.5}浓度数据,结合其他空气污染因子、气象、统计年鉴等多源数据,分析合肥城市圈 PM_{2.5}浓度的影响因素,探索合肥城市圈 PM_{2.5}浓度的时空分布特征,并建立模型对其进行预测。对比三次指数平滑模型,设计了基于模拟退火+遗传+三次指数平滑模型的组合预测模型,并对该组合模型的后期稳定性进行了不确定性分析,从而为大气环境质量监测和评价提供参考。

1 材料与方法

1.1 研究区域

合肥位于中国华东地区,长三角西端,江淮之间,是安徽省省会,地形以丘陵岗地为主,江淮分水岭自西向东横贯全境。气候属亚热带季风性湿润气候,四季分明,夏热冬冷,年均气温为 15.7℃,年均相对湿度为 77%,年均降水量约 1000 mm^[8-9]。近年来,合肥城市圈社会经济发展很快,给空气环境质量保护带来了较大压力。

1.2 数据来源

本文在综合现有研究现状的基础上,根据合肥市 2018-01-01—2019-04-30 的 PM_{2.5}、PM₁₀、SO₂、NO₂、

CO、O₃ 等空气污染因子^[10],以及该时间段内降水、温度、气压、大气能见度等气象因子和合肥市及周边地区监测点(如六安、安庆等)空气污染因子的空间分布数据,进行相关性分析和 PM_{2.5}浓度的预测。

1) PM_{2.5}浓度与其他污染因子的数据,由天气后报网(<http://www.tianqihoubao.com>)提供的全国国内城市历史天气中获取;2) 气象因子的数据,可从中国气象数据网(<http://data.cma.cn/>)获取;3) PM_{2.5}浓度空间分布特点的数据,从中国空气质量在线监测分析平台(<https://www.aqistudy.cn>)历史数据中获取。

1.3 研究方法

1.3.1 数据标准化处理

对原始数据进行处理,消除各数据不同数量级、量纲的影响,PM_{2.5}、气象因子数据进行标准化处理的公式为:

$$X_i = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中: X_i 为标准化后的数据; X 为原始数据; X_{\min} 为最小数据; X_{\max} 为最大数据。

1.3.2 Pearson 相关性

皮尔森相关系数(Pearson correlation coefficient)也称皮尔森积矩相关系数,是一种线性相关系数。Pearson 相关性是一种用线性相关系数反映 2 个变量之间线性程度的数学方法^[11-13]。本文主要应用 SPSS 软件来实现相关性的判定,使用 Oringin 8.5 绘制相关图表,对 2018 年已经标准化处理后的数据进行处理分析。求解皮尔森相关系数 R ,其计算公式如下:

$$R = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

式中: R 为皮尔森相关系数; σ_{xy} 为 2 个样本变量之间的协方差; σ_x 、 σ_y 分别为 2 个样本变量的标准差; x_i 、 y_i 分别为 2 个样本变量中的元素; \bar{x} 、 \bar{y} 分别为 x_i 、 y_i 的平均值。

1.3.3 模拟退火遗传算法(SA+GA)

模拟退火算法(simulated annealing, SA)是基于Monte Carlo迭代求解法后种启发式随机搜索算法。模拟固体物质退火过程的热平衡问题与随机搜索寻优问题的相似性,来达到寻找全局最优或近似全局最优的目的^[14-16]。在搜索最优解的过程中,SA除了可以接受优化解外,还有一个随机接受准则(Metropolis 准则)有限度地接受恶化解,并且接受恶化解的概率逐渐趋向于0,既可找到全局最优解,又保证了算法的收敛性^[17]。遗传算法(genetic algorithm, GA)是按照个体对样本数据的适应程度对种群进行一系列调整,使问题解空间不断逼近最优解,遗传操作包括选择、交叉、变异^[17,18]。SA+GA 操作流程如图1所示。

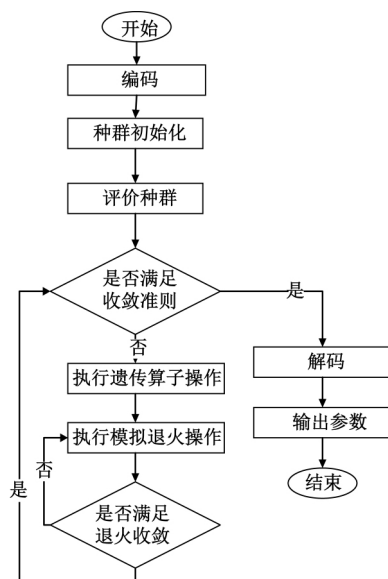


图1 模拟退火遗传算法操作流程

Figure 1 Simulated annealing genetic algorithm operation flow

1.3.4 三次指数模型

指数平滑法是在移动平均法基础上发展的一种算法,对过去已有的数据进行指数加权分析,来预测演变未来的发展趋势^[18,19]。三次指数平滑法是在指数加权二次的基础上进行的,适用于非线性时间序列值的分析预测^[17-19]。

一次指数平滑预测模型:

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha) \hat{x}_t \quad (3)$$

二次平滑模型:

$$\hat{x}_{t+T} = a_t + b_t T \quad (4)$$

三次指数平滑模型:

$$\hat{x}_{t+T} = a_t + b_t T + c_t T^2 \quad (5)$$

式中: \hat{x}_{t+1} 为 $t+1$ 时刻预测值; x_t 为 t 时刻预测值。

平滑参数 α 值一般情况下根据时间数据序列特性在 0~1 选取。 α 越小,参考之前的时间点越多;反之越少。设每次预测结果的绝对误差为 d ,相对误差为 β ,预测的总天数为 n , i 为预测日期,则有以下计算公式:

$$d_i = |\hat{x}_i - x_i| \quad (6)$$

$$\beta_i = \frac{d_i}{x_i} \times 100\% \quad (7)$$

2 结果与分析

2.1 PM_{2.5}浓度时空分布

为分析 PM_{2.5} 浓度的时空分布特点,根据气候统计法,采用四季分类法,这种方法比较适用四季分明的温带地区^[8-10]。春季为 3—5 月,夏季为 6—8 月,秋季为 9—11 月,冬季为 12—2 月。通过分析 2018—01—12 合肥城市圈的 PM_{2.5} 数据,研究安徽省合肥城市圈 PM_{2.5} 浓度月平均变化,结果如图 2 所示。

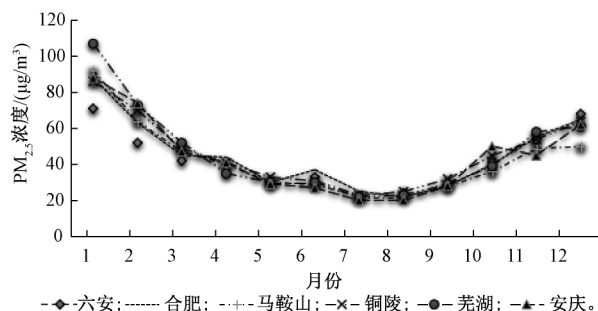


图2 合肥城市圈 2018 年 PM_{2.5} 浓度月均值曲线

Figure 2 PM_{2.5} monthly average concentration in Hefei city circle in 2018

由图 2 可知:各城市 PM_{2.5} 浓度月均浓度变化整体趋势一致。2018 年 PM_{2.5} 浓度季节整体变化情况为冬季>秋季>春季>夏季,月变化总体上表现出“V”形特点,大部分城市 PM_{2.5} 浓度峰值在 1 月,之后浓度开始逐渐下降,在 7 月达到最低值,此后浓度逐渐升高。同时,各地冬季 PM_{2.5} 浓度明显高于夏季。该区域夏季多为雨季,气候湿润,水分有利于吸附空气中的颗粒,且夏季风力较少,颗粒物不易飞扬;而在冬天,由于烧炭取暖及秸秆燃烧,且风力强,容易产生扬尘,所以冬季的 PM_{2.5} 浓度含量明显高于夏季。

通过 Arcgis 软件,运用空间插值法,得到合肥城市圈四季 PM_{2.5} 浓度的空间分布,如图 3 所示。

由图 3 可直观地看出:平均 PM_{2.5} 浓度的最高值主要分布在冬季,最高值达到 86.3 μg/m³。平均

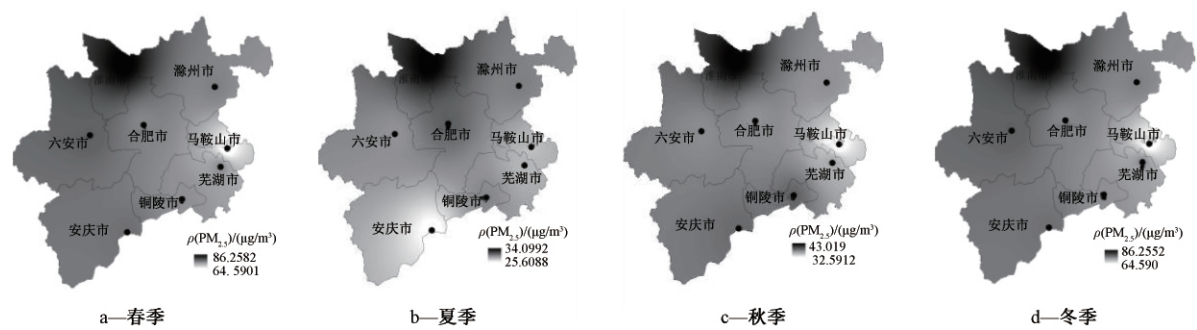


图 3 合肥城市圈四季 PM_{2.5} 浓度的空间分布

Figure 3 Spatial distribution of PM_{2.5} concentration in spring , summer , autumn and winter in Hefei city circle

PM_{2.5} 浓度最低值分布在夏季 ,最高为 25.6 μg/m³。同时 ,随着季节变化 ,PM_{2.5} 浓度值的范围和极值区域的分布也发生改变。在春季 ,PM_{2.5} 浓度的波动范围为 41.4~50.0 μg/m³ ,滁州市和淮南市的 PM_{2.5} 浓度值最高;合肥市和六安市的 PM_{2.5} 浓度值最低。夏季 ,PM_{2.5} 浓度的波动范围为 25.6~34.7 μg/m³ ,淮南市的 PM_{2.5} 浓度值最高;马鞍山市、芜湖市、安庆市的 PM_{2.5} 浓度值最低。秋季 ,PM_{2.5} 浓度的波动范围为 32.6~43.0 μg/m³ ,淮南市的 PM_{2.5} 浓度值最高;马鞍山市和芜湖市的 PM_{2.5} 浓度值最低。冬季 ,PM_{2.5} 浓度的波动范围为 64.6~86.3 μg/m³ ,淮南市的 PM_{2.5} 浓度值最高;马鞍山市和芜湖市的 PM_{2.5} 浓度值最低。

安徽省整体冬季受高压控制 ,由于冷空气下沉 ,降水较少 ,同时北方污染物由西北风传递 ,各种取暖设备以及春节期间大量私家车出行 ,烟花爆竹的燃烧 ,使得大量污染因子聚集 ,所以冬季 PM_{2.5} 浓度较高。秋季较强的太阳辐射加剧了气态污染物的氧化 ,此外 ,夏季积累的逆温现象使得 PM_{2.5} 难以短时间消散。春季盛行风带来北方污染物 ,与冬季的 PM_{2.5} 剩余物结合 ,使春季浓度较高 ,但春季相对充裕的降水量可净化空气。夏季下沉气流及东南季风带来大量水汽 ,有利于污染物的稀释和湿沉降。因此 ,安徽省 PM_{2.5} 浓度由高到低依次为冬季>秋季>春季>夏季 ,这与张海霞等的研究结果一致^[20]。

2.2 PM_{2.5} 浓度影响因素分析

2.2.1 PM_{2.5} 浓度与其他空气污染因子相关性

利用合肥市 2018-01—2019-04 的空气质量监测数据 ,建立多元线性回归数学模型解释其他污染因子(PM₁₀、SO₂、NO₂、CO、O₃) 与 PM_{2.5} 浓度的相关性。为了考察多元线性回归模型的应用合理性和数据结

果的可靠性 ,利用 SPSS 进行残差分析并绘制残差直方图 ,结果见图 4。可知: 监测数据满足多元线性回归的假设条件。

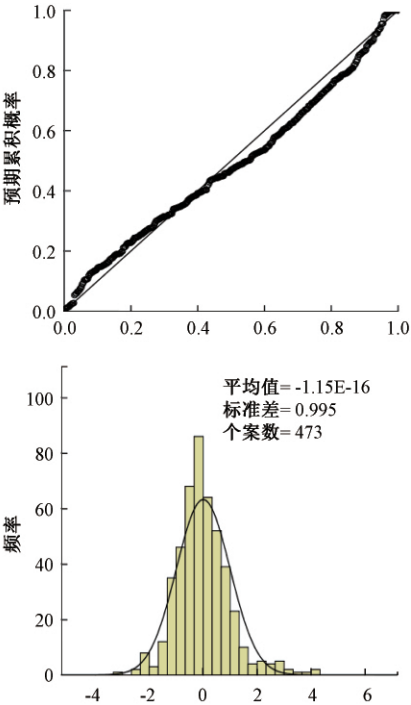


图 4 PM_{2.5} 浓度残差分析与残差直方图

Figure 4 Residual analysis and residual histogram of PM_{2.5} concentration

结合回归方程与数据分析 ,在进行残差分析和多元线性回归模型检验后 ,利用 SPSS 对上述监测数据进行多元线性回归模型评价以及方差分析 ,见表 1 和表 2。可知: 模型对观测值的拟合程度高达 78.9% ,*F* 统计量为 353.695 ,*P* 值为 2×10⁻⁵ ,表明结果具有统计学意义。

表 1 模型评价

Table 1 Model evaluation result

<i>R</i>	<i>R</i> ²	调整后 <i>R</i> ²	标准估算的误差
0.889	0.791	0.789	14.374

表 2 ANOVA(方差分析)
Table 2 ANOVA (analysis of variance) result

项目	平方和	自由度	均方	F	显著性
回归	365364.862	5	73072.972	353.695	0.000
残差	86481.561	467	206.599		
总计	461846.423	472			

根据多元线性模型分析结果(表 3)可得到回归方程如下:

$$y = 0.194x_1 - 0.043x_2 - 0.020x_3 + 0.785x_4 - 0.120x_5 \quad (8)$$

式中:被解释变量 y 为 PM_{2.5};解释变量 x_i ($i=1, 2, \dots, 5$) 分别为 PM₁₀、SO₂、NO₂、CO、O₃。

表 3 多元线性模型结果
Table 3 The results of multivariate linear model

解释变量	未标准化系数	标准误差	标准化系数	t	显著性	皮尔森相关系数
x_1	0.165	0.027	0.194	6.229	0.000	0.471
x_2	-0.398	0.326	-0.043	-1.221	0.003	0.406
x_3	-0.035	0.072	-0.020	-0.489	0.001	0.579
x_4	96.477	3.341	0.785	28.876	0.000	0.875
x_5	-0.125	0.026	-0.120	-4.729	0.000	-0.351

根据上述模型结果可知:PM₁₀、SO₂、NO₂ 等解释变量在 95% 的显著性水平上,均通过显著性检验,其中 PM₁₀ 和 CO 偏回归系数均>0,SO₂、NO₂ 和 O₃ 的均<0。PM_{2.5} 浓度与 CO 呈高度正相关,相关系数高达 0.875;与 PM₁₀、SO₂、NO₂ 的相关性也较高;与 O₃ 的相关性系数<0,呈负相关。尽管许多工业都选用了含硫率较低的煤种,并在燃烧废气排放到大气之前采取了脱硫措施,依然有很多硫化物被排放到大气中。这也是安徽省冬季和春季大气中 SO₂ 的主要来源。PM_{2.5} 的化学成分主要包括有机碳(OC)、元素碳(EC)、硝酸盐、硫酸盐、铵盐和钠盐(Na⁺)等。所以 PM_{2.5} 与 SO₂、NO₂ 质量浓度具有一定的相关性,与崔淑华等^[21]的研究结果基本一致。

2.2.2 PM_{2.5} 浓度与气象因子的相关性

由于气象因子的观测具有实时性,利用 2019-05-06—20 共计 15 d(每天 24 h)的气象台实时监测数据进行 PM_{2.5} 浓度与气象因子相关性分析。通过建立 PM_{2.5} 浓度与各气象因子之间的相关性曲线,分析得到各影响因子的拟合曲线函数见表 4,并求解其皮尔森相关系数 R 。

由表 4 可得:PM_{2.5} 浓度与气压、风速、降雨量以及能见度呈负相关,与温度、相对湿度呈强正相关。因为气压较低时,空气中细颗粒物更多地漂浮在空气

表 4 PM_{2.5} 浓度与各气象因子拟合公式以及相关性分析
Table 4 Fitting formula and correlation analysis of PM_{2.5} concentration and various meteorological factors

气象因子	拟合公式	相关系数 R
气压(x_1)	$y = 2.9291e^{-2.01x_1}$	-0.762
风速(x_2)	$y = 54918e^{-0.338x_2}$	-0.649
温度(x_3)	$y = 9.2618e^{0.0189x_3}$	0.959
相对湿度(x_4)	$y = 44.015x_4^{-1.143}$	0.681
降水量(x_5)	$y = -9.7107x_5^{3.1321}$	-0.813
能见度(x_6)	$y = -18676x_6^{0.704}$	-0.917

中,PM_{2.5} 浓度相对较高;风速越大,空气流通速度越快,同时也易吹散空气中漂浮的细颗粒物,输送至较远距离。另外,在风的作用下,空气污染物不断与洁净空气相融合,进而导致空气污染较轻,PM_{2.5} 浓度含量有一定程度降低。降雨对空气中细颗粒物有一定程度的吸附作用,对空气有净化效果,降雨量越大时,空气中 PM_{2.5} 浓度含量越低。在开放环境中,温度越高,气压越低,所以在不考虑其他气象因素以及环境因素的情况,温度与 PM_{2.5} 浓度含量呈强正相关,其相关系数为 0.959。

2.3 运用三次指数平滑模型预测 PM_{2.5} 浓度

根据 2018—2019 年合肥市地面站点 PM_{2.5} 浓度监测数据,运用三次指数平滑模型预测 PM_{2.5} 浓度。以 2018-01-01—2019-03-31 的数据作为训练集,2019-04-01—30 的数据作为测试集,进行拟合。平滑参数值一般情况下根据时间数据序列特性在 0~1 选取,本文选取 0.1,0.2,⋯,0.9 共 9 个平滑参数值带入平滑预测模型中进行预测,并计算出每次预测结果的平均相对误差 η ,进行对比找到一个较为精确的参数值^[19] 将其应用于 PM_{2.5} 浓度值的预测中。利用 MATLAB 求解,得到各平滑参数值 α 对应的预测结果产生的相对误差结果见表 5。

表 5 各平滑参数 α 预测结果的相对误差
Table 5 Relative errors of prediction results of each smoothing parameter

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
平均相对误差 $\eta/\%$	10.5	8.6	6.7	7.6	10.3	13.0	17.5	22.6	28.1

根据表 5 得知,当平滑参数 α 取值为 0.3 时,相对误差最小,即测试值与实际值的拟合程度较高,其拟合度为 89.6%,因此选用 0.3 作为三次指数平滑预测的平滑参数,运行 MATLAB,对 2019 年 4 月的 PM_{2.5} 浓度进行预测,结果见图 5。

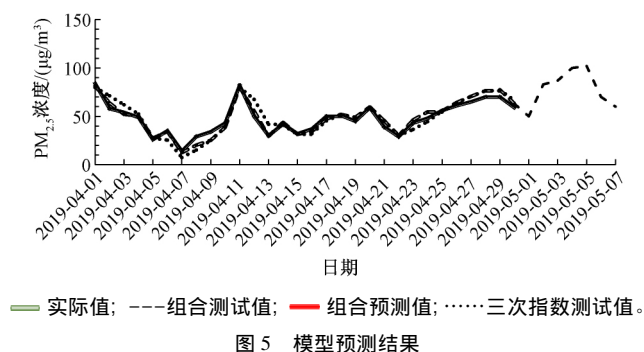


Figure 5 Prediction results of the models and actual values

2.4 运用组合模型预测 $PM_{2.5}$ 浓度

为提高模拟精度,在三次平滑预测模型的基础上,利用遗传算法的全局搜索能力和模拟退火算法的局部搜索能力结合对平滑参数 α 进行寻优,建立模拟退火+遗传+三次指数平滑组合模型对 $PM_{2.5}$ 浓度进行预测。

利用模拟退火遗传算法搜索平滑参数的三次指数平滑预测模型流程为: 1) 收集 $PM_{2.5}$ 浓度的原始时间序列值; 2) 对原始数据进行预处理,减少部分外在因素的影响; 3) 进行三次指数平滑模型的平滑初值的选取; 4) 融合改进遗传算法与模拟退火算法,进行预测搜索得到最佳平滑参数 α 的值; 5) 带入原始数据进行预测,并计算出相对误差。利用 MATLAB 求解得到合肥市 2019 年 4 月拟合效果,并预测 2019-05-01-07 的 $PM_{2.5}$ 浓度,结果如图 5 所示。

通过模型拟合,组合模型的测试值和实际值间拟合程度高达 95%,说明基于遗传模拟退火的三次指数平滑预测组合模型可以精确预测 $PM_{2.5}$ 浓度。对比原三次指数平滑模型拟合度(89.6%),组合模型拟合度更高。在此模型算法的基础上设计开发了《 $PM_{2.5}$ 浓度智能数据分析及预测系统》^[22]。运算结果表明,该模型具有高效、准确、快速等特点,但更适合短期预测,同时可以应用到实际生活中及其他相关领域。

2.5 不确定性分析

所有数学模型在使用过程中均存在不确定性,主要体现在模型参数不确定性以及输出结果不确定性^[23-25]。对于本文中的组合模型,其参数不确定性,主要来源于平滑初值和平滑参数值 α 的选取存在不确定因素。输出结果的不确定性主要来源于预测值与实际值之间存在的误差。为了量化每个不确定性来源,分别使用 Kappa 值和变化平均绝对误差百分率(mean absolute percentage error, MAPE) 作为响应变

量,在 R 软件中进行双因素方差分析。Kappa 的计算公式为

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \quad (9)$$

式中: p_0 为一致性单元的比例; p_e 为期望的偶然一致的单元的比例。

MAPE 是进行需求预测准确率评价中常用也是最有效的指标(预测准确率 = 1 - MAPE)。

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - F_i|}{Y_i} \times 100\% \quad (10)$$

式中: Y_i 为实际值; F_i 为预测值; n 为时间段。

通常 Kappa 介于 0~1,可分为 5 组来表示不同级别的一致性: 0.0~0.20 极低的一致性, 0.21~0.40 一般的一致性, 0.41~0.60 中等的一致性, 0.61~0.80 高度的一致性, 0.81~1 几乎完全一致^[26]。

基于遗传模拟退火算法的三次指数平滑模型中,平滑初值和平滑参数值 α 2 个参数的选取决定着最终输出预测结果的精度,对参数的不确定性进行分析是必要的。基于 2019 年 4 月合肥市 $PM_{2.5}$ 浓度的训练数据,用组合模型对 $PM_{2.5}$ 浓度超过阈值(阈值设定为 $75 \mu g/m^3$) 的概率进行预测,进行不确定性分析。根据组合模型预测值,假设 A: $PM_{2.5}$ 浓度值 $< 75 \mu g/m^3$ 的时间个数; 假设 B: $\geq 75 \mu g/m^3$ 的时间个数。设计 Kappa 系数分类见表 6。

表 6 Kappa 系数分类

Table 6 Kappa coefficient classification		
实际值与预测值	A	B
A	26	0
B	2	2

由表 6 可知: 预测值与实际值 $PM_{2.5}$ 浓度均 $< 75 \mu g/m^3$ 的数据有 26 个,两者均 $> 75 \mu g/m^3$ 的数据有 2 个,预测值 $> 75 \mu g/m^3$ 而实际值小于的有 2 个,预测值 $< 75 \mu g/m^3$ 而实际值大于的有 0 个,利用 Kappa 和 MAPE 进行评价,经计算得到,两者分别为 0.634 和 0.072。

MAPE 的值为 0.072,表示变动越小,输出结果越稳定。Kappa 值输出结果为 0.61~0.80,则表示与实际值具有高度的一致性,即该模型稳定性较高。

3 结论与讨论

1) 研究区域 $PM_{2.5}$ 浓度整体季节变化情况为冬季>秋季>春季>夏季,月变化总体上表现出“V”形特点,大部分城市 $PM_{2.5}$ 浓度峰值在 1 月,之后浓度开始

逐渐下降,在7月达到最低值,此后浓度逐渐升高。

2) PM_{2.5}浓度与CO呈强正相关,相关系数高达0.875;与PM₁₀、SO₂、NO₂的相关性也较高;与O₃的相关性系数小于0,呈负相关。PM_{2.5}浓度与气压、风速、降雨量以及能见度呈负相关,与温度、相对湿度呈正相关。

3) 组合模型的测试值和实际值间拟合程度高达95%,这说明基于遗传模拟退火的三次指数平滑预测模型可以精确预测PM_{2.5}浓度。利用该模型预测了1周后的PM_{2.5}浓度,该模型具有高效、准确、快速等特点,可以应用到实际生活以及其他相关领域。

4) 通过研究文中组合算法的不确定性分析,MAPE的值为0.072,Kappa输出结果为0.61~0.80,该模型预测与实际值具体高度一致性,得知该模型稳定性较高。

本文中所采用的方法是通用的遗传算法,三次指数平滑值的设定在一定程度上会影响组合模型参数的优化,如何对遗传算法做进一步的改良,从而降低预测模型的误差率将是下一步的研究重点。同时,本文空气质量的1周预测是一种小范围预测,接下来的工作将考虑把该算法用于中长期预测,并推广到高原、平原、丘陵地区等更为复杂的环境,进一步验证算法的性能。

参考文献

- [1] 郭家瑜,张英杰,郑海涛,等.北京2015年大气细颗粒物的空间分布特征及变化规律[J].环境科学学报,2017,37(7):2409-2419.
- [2] 姜磊,周海峰,赖志柱,等.中国城市PM_{2.5}时空动态变化特征分析:2015—2017年[J].环境科学学报,2018,38(10):3816-3825.
- [3] 李大年,余家燕,刘海立.重庆市PM_{2.5}浓度时空变化特征分析研究[J].环境与发展,2018,30(1):159-160.
- [4] 王嫣然.北京市PM_{2.5}浓度时空变化特征及影响因素分析[D].北京:北京林业大学,2016.
- [5] 李松,罗绪强,徐红勤.基于GIS的贵州遵义市PM_{2.5}污染时空特征及变化分析[J].环境工程,2015,33(11):63-66.
- [6] 任才溶.基于并行随机森林的城市PM_{2.5}浓度预测[D].太原:太原理工大学,2018.
- [7] 苏筱倩,安俊琳,张玉欣.基于支持向量机回归和小波变换的O₃预报方法[J].中国环境科学,2019,39(9):3719-3726.
- [8] 顾康康,祝玲玲.合肥市主城区PM_{2.5}时空分布特征研究[J].生态环境学报,2018,27(6):1107-1112.
- [9] 杨云,付彦丽.关于空气中PM_{2.5}质量浓度预测研究[J].计算机仿真,2016,33(3):413-418.
- [10] 杨洋,陈春,高梦,等.合肥市空气中PM_{2.5}的计量分析[J].河北北方学院学报(自然科学版),2016,32(11):50-55.
- [11] 施惘惘,张庆国,张浩,等.合肥市大气能见度变化特征及其影响因素[J].应用生态学报,2017,28(2):485-492.
- [12] 潘本锋,汪巍,李莉娜.我国典型城市PM_{2.5}空间分布均匀性分析[J].环境工程,2015,33(5):109-113.
- [13] 李会霞,史兴民.西安市PM_{2.5}时空分布特征及气象成因[J].生态环境学报,2016,25(2):266-271.
- [14] 王云中.基于神经网络的PM_{2.5}浓度预测研究与实现[D].西安:西安电子科技大学,2018.
- [15] 殷国栋.GASA-BP神经网络在安徽省工业用电量预测中的应用研究[D].合肥:安徽建筑大学,2018.
- [16] 王敏,邵滨,郭宇,等.基于BP人工神经网络的城市PM_{2.5}浓度空间预测[J].环境污染与防治,2013,35(9):63-66,70.
- [17] 王惠琳,胡树根,王耘.基于模拟退火遗传算法优化的BP网络在质量预测中的应用[J].轻工机械,2011,29(4):26-31,35.
- [18] 马健,盛魁,朱庆友,等.基于遗传模拟退火三次指数平滑模型的中药材价格预测研究[J].佳木斯大学学报(自然科学版),2017,35(5):857-860,874.
- [19] 李慧敏.基于三次指数平滑模型的雾霾天气预测[J].中国环境管理干部学院学报,2017,27(3):52-55,74.
- [20] 张海霞,程先富,陈冉慧.安徽省PM_{2.5}时空分布特征及关键影响因素识别研究[J].环境科学学报,2018,38(3):1080-1089.
- [21] 崔淑华,倪艳姝.城市主要气体污染物与PM_{2.5}相关性建模分析[J].森林工程,2016,32(2):65-68.
- [22] 王学梅,陈滔,张庆国.PM_{2.5}浓度智能数据分析及预测系统:2019SR0976437.V1.0[P].
- [23] SGAMBI L, GARAVAGLIA E, BASSO N, et al. Monte Carlo simulation for seismic analysis of a long span suspension bridge[J]. Engineering Structures, 2014, 78: 100-111.
- [24] 骆勇鹏,黄方林,刘景良,等.基于有限测试数据的结构响应不确定性量化分析[J].铁道科学与工程学报,2018,15(12):3217-3224.
- [25] 刘庚,毕如田,张朝,等.某焦化场地苯并(a)芘污染空间分布范围预测的不确定性分析[J].环境科学学报,2013,33(2):587-593.
- [26] 张剑波,江泽友,房俊,等.Kappa统计量和Bland-Altman分析法在HbA1c试剂盒一致性评价中的应用[J].数学的实践与认识,2019,49(20):167-175.