

# 基于数据预处理和支持向量回归的PM<sub>2.5</sub>、NO<sub>2</sub>和SO<sub>2</sub>组合预测研究

——以兰州市和西安市为例

方敏, 魏麟, 马晶, 李元林, 袁艳, 崔旭东, 施岱瑜, 朱素玲\*

兰州大学, 甘肃 兰州  
Email: zhusl@lzu.edu.cn

收稿日期: 2020年10月1日; 录用日期: 2020年10月16日; 发布日期: 2020年10月23日

## 摘要

由于空气污染能够引起多种呼吸道非传染性疾病, 甚至造成生命质量受损或者过早死亡, 所以空气污染已经成为全球死亡人数增加的第二大原因, 因此, 制定合理且有效的空气污染治理措施已经迫在眉睫。预测在空气污染预警过程中扮演着重要的角色, 准确且科学的预测能够帮助人们有效规避空气污染的危害, 因此, 提高预测的精度与科学性也成为众多学者关心的问题之一。本研究为了提高空气污染预测的准确性, 采用了数据预处理互补集合经验模式分解(complementary ensemble empirical mode decomposition, CEEMD)技术、支持向量回归(support vector regression, SVR)、广义回归神经网络(general regression neural network, GRNN)和粒子群优化(particle swarm optimization, PSO)算法建立组合模型。通过PM<sub>2.5</sub>、NO<sub>2</sub>和SO<sub>2</sub>时间序列数据检验建立的组合预测模型的有效性, 根据平均绝对百分比误差(mean absolute percentage error, MAPE)发现: 组合预测模型能够提高PM<sub>2.5</sub>、NO<sub>2</sub>和SO<sub>2</sub>指标的预测精度, 如西安市的SO<sub>2</sub>指标, 最优单项模型的MAPE值为6.13%, 而组合模型的MAPE值为5.86%。总之, 组合预测模型能够为空气污染治理提供更准确的预测信息, 为空气污染的防控提供理论支持。

## 关键词

空气污染, 组合预测, 分解与集成, 支持向量回归

# Research on Combined Forecasting of PM<sub>2.5</sub>, NO<sub>2</sub> and SO<sub>2</sub> Based on Data Preprocessing Technology and Support Vector Regression —Cases Study of Lanzhou and Xi'an

Min Fang, Lin Wei, Jing Ma, Yuanlin Li, Yan Yuan, Xudong Cui, Daiyu Shi, Suling Zhu\*

\*通讯作者。

文章引用: 方敏, 魏麟, 马晶, 李元林, 袁艳, 崔旭东, 施岱瑜, 朱素玲. 基于数据预处理和支持向量回归的 PM<sub>2.5</sub>、NO<sub>2</sub> 和 SO<sub>2</sub> 组合预测研究[J]. 统计学与应用, 2020, 9(5): 792-800. DOI: 10.12677/sa.2020.95082

Lanzhou University, Lanzhou Gansu  
Email: zhusl@lzu.edu.cn

Received: Oct. 1<sup>st</sup>, 2020; accepted: Oct. 16<sup>th</sup>, 2020; published: Oct. 23<sup>rd</sup>, 2020

## Abstract

Because air pollution can cause a variety of respiratory non-communicable diseases, and even lead to impaired quality of life or premature death, the air pollution has become the second leading cause of the increase in deaths worldwide. Therefore, it is extremely urgent to formulate reasonable and effective air pollution control measures. Prediction plays an important role in the early warning process of air pollution. Accurate and scientific predictions can help people avoid the hazards of air pollution effectively. Therefore, how to improve the accuracy and scientificity of prediction has become one of the concerns of many scholars. To improve the accuracy of air pollution prediction, this study adopts data preprocessing technology complementary ensemble empirical mode decomposition (CEEMD), support vector regression (SVR), general regression neural network (GRNN), and particle swarm optimization (PSO) algorithms to establish a combined model. By  $PM_{2.5}$ ,  $NO_2$  and  $SO_2$  index time-series data set checking the validity of combined forecasting model, according to the mean absolute percentage error (MAPE), it is found that the combined model can improve the prediction accuracy of  $PM_{2.5}$ ,  $NO_2$  and  $SO_2$  index. For example, the MAPE of the optimal individual model and the combined model of the  $SO_2$  index is 6.13% and 5.86% respectively in Xi'an. In short, the combined prediction model can provide more accurate prediction information for air pollution control and provide theoretical support for air pollution prevention and control.

## Keywords

Air Pollution, Combined Forecasting Model, Decomposition and Ensemble, Support Vector Regression

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

空气污染是由能够改变空气自然特性的化学性、物理性的气体或者固体组成的混合物对室内外环境造成的污染,其主要污染物包括:颗粒物(细颗粒物,  $PM_{2.5}$  和粗颗粒物,  $PM_{10}$ )、臭氧( $O_3$ )、二氧化硫( $SO_2$ )、一氧化碳(CO)和二氧化氮( $NO_2$ ) [1]。根据世界卫生组织报道,城市、郊区和农村的空气污染每年大约导致全球 500 万人死亡,仅次于吸烟[2]。Wang Y.等[3]研究了 31 个城市的空气污染状况,结果发现绝大多数城市中占比最大的空气污染成分都是  $PM_{2.5}$ 。此外,  $PM_{2.5}$  暴露是慢性阻塞性肺疾病(COPD)、下呼吸道感染、脑血管等疾病的重要影响因素[4]。  $NO_2$  进入人体后能形成硝酸和亚硝酸,并对肺组织产生剧烈的刺激从而引起肺炎、COPD 等疾病[5]。Khaniabadi Y.等[6]研究了霍拉马巴德地区的  $SO_2$  发现  $SO_2$  的浓度与急性心梗和 COPD 住院人数的增加呈正相关。

空气污染对人类健康的危害越来越严重,其中主要是由于空气污染物浓度的增加,如  $PM_{2.5}$ 、 $NO_2$  和  $SO_2$ , 所以降低空气污染物水平已经成为改善人类健康的重要措施。根据中国环境空气质量标准的规

定, PM<sub>2.5</sub>、NO<sub>2</sub> 和 SO<sub>2</sub> 一级标准的年平均限值分别为 15 μg/m<sup>3</sup>、40 μg/m<sup>3</sup> 和 20 μg/m<sup>3</sup>, 二级标准分别为 35 μg/m<sup>3</sup>、40 μg/m<sup>3</sup> 和 60 μg/m<sup>3</sup>; PM<sub>2.5</sub>、NO<sub>2</sub> 和 SO<sub>2</sub> 平均 24 小时的一级标准为 35 μg/m<sup>3</sup>、80 μg/m<sup>3</sup> 和 50 μg/m<sup>3</sup>, 二级标准为 75 μg/m<sup>3</sup>、80 μg/m<sup>3</sup> 和 150 μg/m<sup>3</sup> [7]。为了有效地降低空气污染物水平, 减少过早死亡的发生, 因此相关部门需要尽快制定更合理的环境管理措施, 而空气污染预测作为一种科学的空气质量预警技术, 其预测信息能够指导相关部门制定改善空气质量的措施。王建州等的研究指出: 组合预测模型结合了不同单项预测模型的优点, 能够更准确地预测空气污染趋势[8]。因此本文基于互补集合经验模式分解 (complementary ensemble empirical mode decomposition, CEEMD)、支持向量回归 (support vector regression, SVR) 和广义回归神经网络 (general regression neural network, GRNN) 建立组合模型, 该模型的权重由粒子群优化算法 (particle swarm optimization, PSO) 搜索获得。PM<sub>2.5</sub>、NO<sub>2</sub> 和 SO<sub>2</sub> 数据的预测结果表明, 所建立的组合模型优于单项模型。

## 2. 方法

### 2.1. 互补集合经验模式分解

基于 EMD 在分解非线性、不平稳信号的优势和 EEMD 解决“模式混合”问题的特点提出了 CEEMD 方法[9]。以兰州市的 NO<sub>2</sub> 为例, 用 CEEMD 分解图如图 1 所示, 具体过程如下:

第一步, 在原始信号  $x$  的基础上添加一对相反的白噪声信号  $n_t$ , 得到具有正负信号的噪声序列  $x_t^+$  和  $x_t^-$ ,

$$\begin{bmatrix} x_t^+ \\ x_t^- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ n_t \end{bmatrix}$$

第二步, 将噪声信号用 CEEMD 分解为  $IMF_{d,t}^+$  和  $IMF_{d,t}^-$ , 重复  $\kappa$  次, 其中  $IMF_{d,t}$  是第  $t$  次的第  $d$  个本征模态函数 (intrinsic mode function, IMF)。

第三步, 计算由正负信号分解的 IMFs 的均值, 其中  $\bar{T}_\kappa$  是指 CEEMD 分解的 IMFs 的均值,

$$\bar{T}_\kappa = \left( \frac{1}{2\kappa} \right) \sum_{t=1}^{\kappa} (IMF_{d,t}^+ + IMF_{d,t}^-)$$

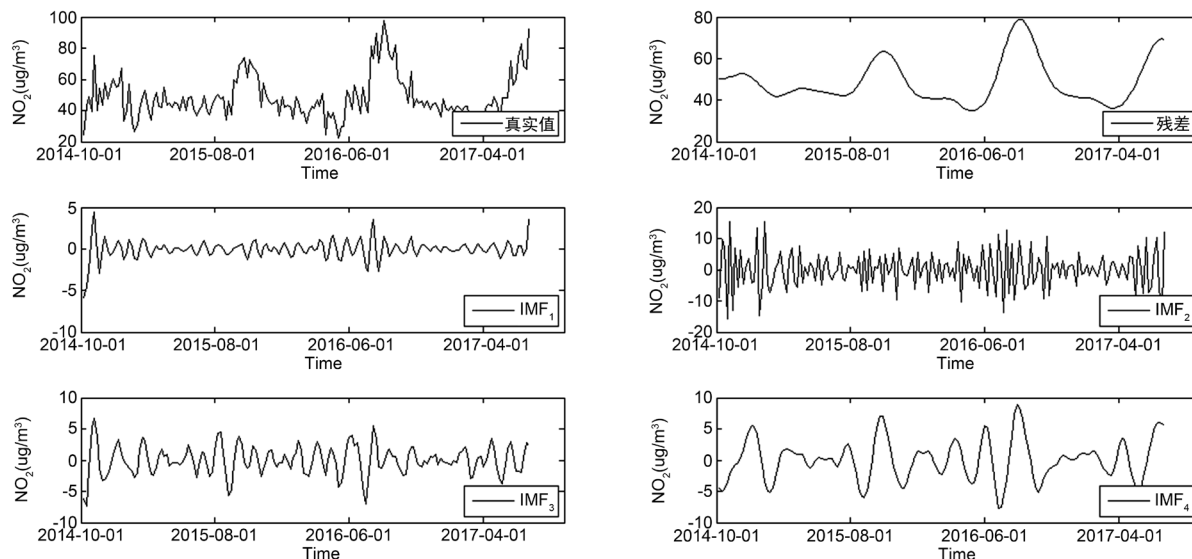


Figure 1. The true series and decomposition series of NO<sub>2</sub> indicator in Lanzhou

图 1. 兰州市 NO<sub>2</sub> 的真实序列和分解序列

## 2.2. 支持向量回归

回归的基本思想是指给定训练样本, 使得样本的拟合值与真实值尽可能接近。由于样本的不规则性和非线性, 传统的回归方法使得拟合值与真实值之间存在较大误差, 因此, 本文运用优化算法优化 SVR 拟合由 CEEMD 分解的 IMF 和残差序列。支持向量回归的基本思想是将样本数据通过一个非线性映射  $\phi$  映射到高维特征空间, 并在这个高维特征空间中进行线性回归[10], 即

$$f(X) = \omega\phi(X) + b$$

$$\phi: R^l \rightarrow R$$

其中  $b$  是阈值,  $\omega$  是权向量,  $l$  是维数,  $X$  是影响因素矩阵。支持向量回归机与传统回归的区别是: 当模型输出的  $f(X)$  与真实  $y$  之间的损失大于偏差  $\epsilon$  时计算损失。

$\phi(\cdot)$  定义之后, 采用  $\epsilon$  不敏感损失函数并引入松弛变量  $\xi_i$  和  $\xi_i^*$ , 于是 SVR 问题就转换为:

$$\min_{\omega} 1/2 \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\text{s.t. } f(X_i) - y_i \leq \epsilon + \xi_i^*$$

$$y_i - f(X_i) \leq \epsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0$$

其中  $C$  是误差惩罚参数, 它的作用是平衡模型复杂性和经验风险, 其对 SVR 的性能也有很大影响。解决上述凸优化问题可以引入拉格朗日函数将其转化为对偶问题解决, 即

$$f(X) = \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i) \kappa(X, X_i) + b$$

其中  $\alpha_i$  和  $\hat{\alpha}_i$  是拉格朗日乘子,  $\kappa(X, X_i)$  为核函数, 本研究中采用高斯核函数:

$$\kappa(X, X_i) = \exp(-g \|X - X_i\|^2)$$

其中  $g$  为核参数。由于核参数( $g$ )与误差惩罚参数( $C$ )对于预测结果存在一定影响, 因此我们引入优化算法求得最优的参数[11]。

Mirjalili S. 和 Hashim S. [12] 基于鸟群的觅食行为提出 PSO, 该算法的基本思想是在搜索空间中通过更新 fitness 求得最优全局解。Sarafrazi S. 和 Nezamabadi-pour H. [13] 基于万有引力定律, 即搜索空间中质量越大的粒子具有更大的吸引力, 提出重力搜索算法(gravitational search algorithm, GSA)。本研究主要运用 PSO 和 GSA 算法的快速局部收敛优势和开发能力求得 SVR 模型的最优核参数( $g$ )与误差惩罚参数( $C$ ), 并利用 PSO 算法搜索组合预测中单项模型的最优权重。

## 2.3. 广义回归神经网络

Specht D. [14] 在 1991 年提出一种具有非线性映射能力且学习收敛速度快的 GRNN 模型, 其简单的网络结构同时具有较好的容错性和鲁棒性。GRNN 由输入层、模式层、求和层和输出层构成, 如图 2 所示, 其中输入层用来输入训练的样本数据, 根据样本向量的维数确定神经元的个数; 模式层的神经元相互独立且传递函数为径向基函数; 求和层是对模式层各单元输出和各单元输出的加权值进行求和。

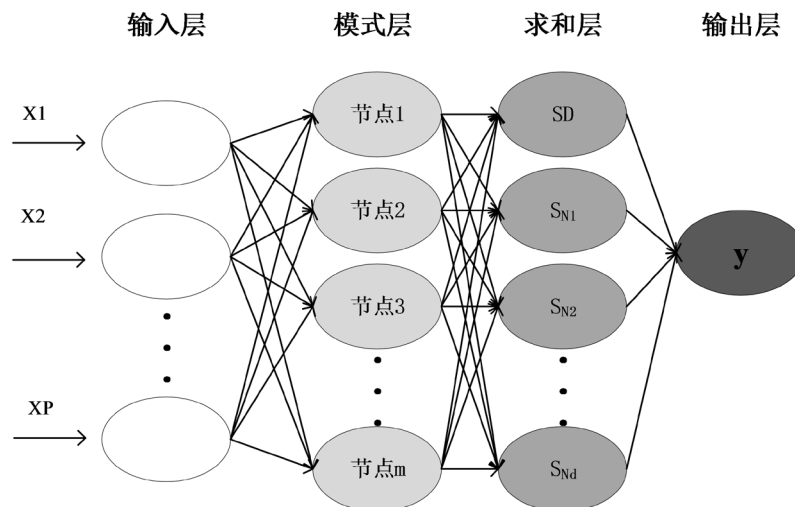


Figure 2. The structure of GRNN  
图 2. GRNN 神经网络结构

### 2.4. 组合预测模型

将2014年10月1日~2017年12月31日的PM<sub>2.5</sub>、NO<sub>2</sub>和SO<sub>2</sub>浓度的167个周数据分为训练集(1-122, 122个周数据)、验证集(123-137, 15个周数据)和测试集(138-167, 30个周数据)。训练集数据用来优化IMF和lag的个数以获得高精度的单项模型, 基于验证集数据用PSO算法搜索组合模型的权重, 测试集数据用来测试组合模型的预测能力。

第一步, 分解。CEEMD预处理方法将PM<sub>2.5</sub>、NO<sub>2</sub>和SO<sub>2</sub>浓度的122个周数据分别分解为不同的低频IMF<sub>i</sub><sup>低</sup>、高序列IMF<sub>i</sub><sup>高</sup>和残差序列。根据试错法选择合适的IMF个数和滞后阶数(lag),

$$y = IMF_i^{\text{高}} + IMF_i^{\text{低}} + \text{残差}_i$$

第二步, 拟合。高频和低频序列用PSO-SVR算法拟合获得 $\widehat{IMF}_i^{\text{高}}$ 和 $\widehat{IMF}_i^{\text{低}}$ , 残差序列分别用GRNN、PSOGSA-SVR、PSO-SVR、GSA-SVR算法进行拟合。将高频、低频和残差的拟合结果相加获得单项模型的预测值,

$$\hat{y}_M = \widehat{IMF}_i^{\text{高}} + \widehat{IMF}_i^{\text{低}} + \widehat{\text{残差}}, M = 1, 2, \dots, 48$$

第三步, 搜索权重。通过第一步和第二步为每个污染物选择三个精度最高的单项模型 $\hat{x}_1$ 、 $\hat{x}_2$ 和 $\hat{x}_3$ , 基于验证集数据(15个周数据)使用PSO优化算法在搜寻三个单项模型的最优权重 $\hat{W}^1$ 、 $\hat{W}^2$ 和 $\hat{W}^3$ 。因此我们可以得到组合预测为:

$$\hat{y} = \hat{W}^1 \cdot \hat{x}_1 + \hat{W}^2 \cdot \hat{x}_2 + \hat{W}^3 \cdot \hat{x}_3$$

第四步, 通过30个周数据测试建立的组合模型的效果。组合模型的建模过程如图3所示, 以PM<sub>2.5</sub>为例。

## 3. 结果

### 3.1. 预测评估指标

对建立的组合模型和单项模型用平均绝对百分比误差(mean absolute percentage error, MAPE)、均方误差(mean square error, MSE)、平均绝对误差(mean absolute error, MAE)和一致性指标(index of agreement,

IA)进行评价。MAPE、MAE 和 MSE 值越小,说明预测效果越好,IA 是一种评价不同模型之间预测能力的无量纲指标,IA 越接近 1,模型的预测能力越好。分别定义如下:

$$\text{MAPE} = \frac{1}{i} \sum_{i=138}^{167} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

$$\text{MSE} = \frac{1}{i} \sum_{i=138}^{167} (y_i - \hat{y}_i)^2,$$

$$\text{MAE} = \frac{1}{i} \sum_{i=138}^{167} |y_i - \hat{y}_i|,$$

$$\text{IA} = 1 - \frac{\sum_{i=138}^{167} (y_i - \hat{y}_i)^2}{\sum_{i=138}^{167} (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2}$$

其中  $y_i$  是真实的浓度值,而  $\hat{y}_i$  分别表示预测值,  $\bar{y}$  是真实浓度的平均值。

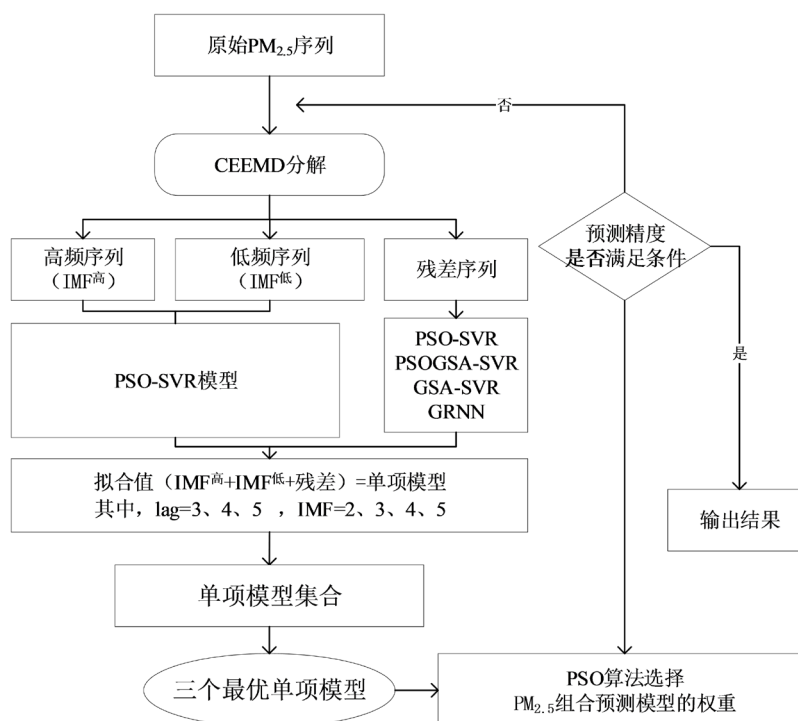


Figure 3. The flow chart of the combined model  
图 3. 组合模型的流程图

### 3.2. 模型选择

本研究中 IMF 的个数( $d$ )分别设置为 2、3、4 和 5,滞后的阶数(lag)分别设置为 3、4 和 5,对每个污染序列来说总共获得  $4 \times 3 \times 4 = 48$  个单项模型。根据模型评价指标 MAPE 从 48 个单项模型中选择三个精度最高的模型,此外,由于气候条件、经济状况等方面的原因,最优单项模型的 lag 和 IMF 个数在不同城市的不同污染指标中表现不同,具体信息见表 1,如兰州市  $\text{PM}_{2.5}$  的滞后的阶数为 3,IMF 的个数为 5,而西安市滞后的阶数为 3,IMF 的个数为 3。

**Table 1.** The number of Lag, IMF and the residuals of the three individual models**表 1.** 三个单项模型的 lag、IMF 个数和残差拟合的模型

城市	模型	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>
兰州	1	(3, 5, P)	(5, 4, PG)	(3, 5, NN)
	2	(3, 5, PG)	(5, 4, G)	(3, 4, P)
	3	(3, 5, G)	(5, 3, P)	(3, 4, PG)
西安	1	(3, 3, P)	(3, 4, P)	(4, 5, P)
	2	(3, 3, PG)	(3, 3, P)	(4, 5, PG)
	3	(3, 3, G)	(3, 3, PG)	(4, 5, G)

注释：第一个参数是滞后的阶数，第二个参数表示分解的 IMF 的个数，第三个参数表示残差序列的预测模型，其中 NN、P、G 和 PG 分别表示 GRNN、PSO-SVR、GSA-SVR 和 PSOGSA-SVR 模型。

根据原始数据的特点与内在的价值，通过 MAPE、MAE、MSE 和 IA 值选择三个最优的单项预测模型；然后，基于三个最优单项模型用 PSO 算法寻得三个模型的权重值，见表 2。从表 2 中可以看出单项模型的预测精度越高，PSO 算法寻得权重值越大，如兰州市 PM<sub>2.5</sub> 最优模型的 MAPE 值为 4.44%，其权重为 0.6372。

**Table 2.** The weights of three optimal individual model**表 2.** 三个最优单项模型的权重

城市	模型	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>
兰州	1	0.6372	0.0031	0.9935
	2	0.3603	0.0085	0.0025
	3	0.0025	0.9884	0.0040
西安	1	0.0958	0.4612	0.8948
	2	0.0046	0.5372	0.0287
	3	0.8996	0.0016	0.0765

### 3.3. 预测结果分析

组合模型与单项模型的 MAPE 值如表 3 所示，从表中可以发现：各城市不同指标的组合预测模型优于其最优的单项预测模型，说明组合模型不仅泛化性能较高，而且稳定性较高。此外，组合模型可以显著提高空气污染指标的预测精度，如西安市 SO<sub>2</sub> 指标中最优单项模型是 6.13%，组合模型为 5.86%。表 4 是组合模型的 MAE、MSE 和 IA 值，从中可以看出组合模型的 IA 值都在 0.95 之上，说明组合模型总体预测效果较好。

**Table 3.** The MAPE of combined model and individual model**表 3.** 组合模型和单项模型的 MAPE 值

城市	模型	PM <sub>2.5</sub> (%)	SO <sub>2</sub> (%)	NO <sub>2</sub> (%)
兰州	单项模型-1	4.44	8.28	4.90
	单项模型-2	4.52	8.65	5.96
	单项模型-3	4.51	7.09	5.28
	组合模型	<b>4.38</b>	<b>7.09</b>	<b>4.89</b>

## Continued

西安	单项模型-1	9.58	6.13	4.69
	单项模型-2	9.97	8.95	4.85
	单项模型-3	8.81	9.33	4.85
	组合模型	<b>8.71</b>	<b>5.86</b>	<b>4.66</b>

Table 4. The MSE, MAE and IA of combined model

表 4. 组合模型的正态、MAE 和 IA 值

城市	评价指标	PM <sub>2.5</sub>	SO <sub>2</sub>	NO <sub>2</sub>
兰州	MAE	2.2207	1.7348	2.8926
	MSE	7.3915	5.3027	14.4667
	IA	0.9897	0.9944	0.9868
西安	MAE	6.5880	1.0462	2.7325
	MSE	86.9372	1.8734	11.9436
	IA	0.9769	0.9917	0.9819

## 4. 结论

本研究提出组合预测模型对大气污染进行建模分析, 并使用兰州市和西安市的 PM<sub>2.5</sub>、NO<sub>2</sub> 和 SO<sub>2</sub> 浓度数据进行了验证和测试。为了提高组合模型的预测精度, 使用了 CEEMD 方法对数据进行分解处理, 即将原始数据分解为不同的低频和高频数据, 这种处理方法显著提高了模型预测的准确性。此外, 将机器学习方法(SVR)的参数用三种不同的算法进行了优化, 结果发现不同城市不同指标的数据结构适用不同的优化算法。最后, 基于所选择的最优单项模型, 使用了 PSO 算法搜索出适合的权重值。数据分析结果表明组合模型的预测精度高于最优的单项模型。因此, 组合预测模型能够为空气污染治理提供更准确的预测信息, 为空气污染的防控提供理论支持。

## 基金项目

本研究在全国统计科学研究项目(2018LZ30)和 2018 年度教育部春晖计划项目资助下完成。

## 参考文献

- [1] 世界卫生组织. 空气污染[S]. [https://www.who.int/topics/air\\_pollution/zh/](https://www.who.int/topics/air_pollution/zh/), 2016.
- [2] Bao, J., Yang, X., Zhao, Z., et al. (2015) The Spatial-Temporal Characteristics of Air Pollution in China from 2001-2014. *International Journal of Environmental Research and Public Health*, **12**, 15875-15887. <https://doi.org/10.3390/ijerph121215029>
- [3] Wang, Y., Ying, Q., Hu, J., et al. (2014) Spatial and Temporal Variations of Six Criteria Air Pollutants in 31 Provincial Capital Cities in China during 2013-2014. *Environment International*, **73**, 413-422. <https://doi.org/10.1016/j.envint.2014.08.016>
- [4] Zhou, M., Wang, H., Zeng, X., et al. (2019) Mortality, Morbidity, and Risk Factors in China and Its Provinces, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet*, **394**, 1145-1158. [https://doi.org/10.1016/S0140-6736\(19\)30427-1](https://doi.org/10.1016/S0140-6736(19)30427-1)
- [5] 马洪群, 崔莲花. 大气污染物(SO<sub>2</sub>、NO<sub>2</sub>)对中国居民健康效应影响的 meta 分析[J]. 职业与健康, 2016, 32(8): 1038-1044.
- [6] Khaniabadi, Y., Usef, O., Daryanoosh, S.M., Hopke, P.K., et al. (2017) Acute Myocardial Infarction and COPD Attributed to Ambient SO<sub>2</sub> in Iran. *Environmental Research*, **156**, 683-687. <https://doi.org/10.1016/j.envres.2017.04.028>
- [7] 中华人民共和国环境保护部. 环境空气质量标准[S].



- 
- <http://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqhjzlbz/201203/W020120410330232398521.pdf>, 2012.
- [8] Wang, J., Du, P., Hao, Y., *et al.* (2020) An Innovative Hybrid Model Based on Outlier Detection and Correction Algorithm and Heuristic Intelligent Optimization Algorithm for Daily Air Quality Index Forecasting. *Journal of Environmental Management*, **255**, 109855. <https://doi.org/10.1016/j.jenvman.2019.109855>
- [9] Chen, J., Zhou, D., Lyu, C., *et al.* (2018) An Integrated Method Based on Ceemd-Sampen and the Correlation Analysis Algorithm for the Fault Diagnosis of a Gearbox under Different Working Conditions. *Mechanical Systems and Signal Processing*, **113**, 102-111. <https://doi.org/10.1016/j.ymssp.2017.08.010>
- [10] Vapnik, V. (1999) An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, **10**, 988-999. <https://doi.org/10.1109/72.788640>
- [11] 闫国华, 朱永生. 支持向量机回归的参数选择方法[J]. 计算机工程, 2009, 35(14): 218-220.
- [12] Mirjalili, S. and Hashim, S. (2010) A New Hybrid PSOGSA Algorithm for Function Optimization. *The 2010 International Conference on Computer and Information Application*, Tianjin, 3-5 December 2010, 374-377. <https://doi.org/10.1109/ICCI.2010.6141614>
- [13] Sarafrazi, S. and Nezamabadi-Pour, H. (2013) Facing the Classification of Binary Problems with a GSA-SVM Hybrid System. *Mathematical and Computer Modelling*, **57**, 270-278. <https://doi.org/10.1016/j.mcm.2011.06.048>
- [14] Specht, D. (1991) A General Regression Neural Network. *IEEE Transactions on Neural Networks*, **2**, 568-576. <https://doi.org/10.1109/72.97934>