

DOI: 10.3969/j.issn.1673-6141.2019.03.004

## 基于气象因素的 PM10 浓度预测

蔡春茂, 何红弟

(上海海事大学物流研究中心, 上海 201306)

**摘要:** 为建立准确高效的空气质量预报系统, 建立以污染物、气象因素、污染物混合气象因素的三种预测因子模式, 并将该三种预测因子模式作为支持向量机回归 (Support vector machine regression, SVR) 的输入变量进行 PM10 浓度的每日预测, 寻找最优预测因子模式。并使用灰狼优化算法 (Grey wolf optimization, GWO) 对支持向量机回归模型进行优化, 形成 GWO-SVR 模型。实验结果表明, 污染物混合气象因素作为输入变量为最优预测因子模式, SVR 和 GWO-SVR 模型测试集确定系数分别达到  $R^2=0.79$  和  $R^2=0.81$ , 预测精度较高, 经比较发现 GWO-SVR 模型预测性能较好。之后, 依据风向条件对数据进行分类, 使用较优的 GWO-SVR 进行 PM10 浓度预测, 预测结果显示盛行西南风时, 预测集评测指标为  $R=0.91$ 、 $M_{SE}=47.15$ , 优于盛行东北风时的  $R=0.87$ 、 $M_{SE}=125.80$  和所有数据下的  $R=0.90$ 、 $M_{SE}=107.94$ 。

**关键词:** 气象因素; 污染物; GWO-SVR 模型; 分类预测

中图分类号: O431.2; P426

文献标识码: A

文章编号: 1673-6141(2019)03-0191-010

## Prediction of PM10 Concentration Based on Meteorological Factors

CAI Chunmao, HE Hongdi

(Logistics Research Center, Shanghai Maritime University, Shanghai 201306, China)

**Abstract:** In order to establish an accurate and efficient air quality forecasting system, three forecasting models based on pollutants, meteorological factors and pollutant mixed meteorological factors were established and used as input variables for support vector machine regression (SVR) for the daily forecast to look for the best predictor mode. The support vector machine regression model was optimized by using grey wolf optimization (GWO) to form the GWO-SVR model. The experimental results show that the meteorological factors of pollutants mixing acted as input variables is the optimal predictor model, and the determination coefficients of the test set of SVR and GWO-SVR model are  $R^2=0.79$  and  $R^2=0.81$ , respectively, which indicates both of the models have high prediction accuracy. By comparison, GWO-SVR model prediction performance is better. After that, the data is classified according to the wind direction conditions and the better GWO-SVR is used to predict the PM10 concentration. The prediction results show that when the prevailing southwest wind prevails, the evaluation index of prediction set is  $R=0.91$  and  $M_{SE}=47.15$ , which is better than the

**基金项目:** Supported by National Natural Science Foundation of China(国家自然科学基金, 11672176)

**作者简介:** 蔡春茂 (1991 -), 安徽省安庆人, 研究生, 主要从事空气质量预测方面的研究。 **E-mail:** caicm98@163.com

**导师简介:** 何红弟 (1980 -), 男, 陕西宝鸡人, 博士, 副教授, 主要从事低碳物流、绿色交通方面的研究。 **E-mail:** 4003963@qq.com

**收稿日期:** 2017-11-22; **修改日期:** 2018-01-03

status with prevailing northeasterly wind where  $R=0.87$  and  $M_{SE}=125.80$  and the whole data with  $R=0.90$  and  $M_{SE}=107.94$ .

Key words: meteorological factors; pollutants; GWO-SVR model; classified forecast

## 1 引言

近年来,随着大气中颗粒物浓度的增加,空气质量下降,雾霾天气成为热点话题,建立完善的空气污染治理机制成为了人民和政府面对的重点问题。大气中颗粒物主要包括细颗粒物 PM<sub>2.5</sub>(直径小于等于 2.5  $\mu\text{m}$  的颗粒物)和可吸入颗粒物 PM<sub>10</sub>(直径在 10  $\mu\text{m}$  以下的颗粒物)。相关研究表明,大气中的颗粒物会给人类健康带来许多不利的影响,比如增加心血管和呼吸系统疾病的死亡率等<sup>[1-2]</sup>。根据世界卫生组织(World health organization, WHO)的报告,全球每年超过 200 万人因为吸入过多可吸入颗粒物而死亡,原因是暴露于 PM<sub>10</sub> 中。空气中 PM<sub>10</sub> 每增加 10  $\mu\text{g}/\text{m}^3$ ,老年人呼吸死亡率增加 1.62%,年轻人呼吸死亡率增加 1.05%<sup>[3]</sup>,据 2012 年世界卫生组织 WHO(world health organization)估计,城市和农村地区的环境(室外)空气污染每年导致全世界 370 万人过早死亡。在香港,大气颗粒物 (Particulate matter, PM) 是影响健康的重要来源。这些相关问题已经得到香港民众和相关部门的高度关注<sup>[4]</sup>。

为了尽快使空气质素达到并维持在可接受水平,香港环境保护署于 2014 年 1 月 1 日更新香港的空气质素指标,以进一步保障公众健康。香港制定的环境空气指标 PM<sub>10</sub> 年平均浓度和 24 h 平均浓度分别为 50  $\text{mg}/\text{m}^3$  和 100  $\text{mg}/\text{m}^3$ 。该指标的发布和实施在监测空气质量和改善生活环境方面发挥了重要作用。香港环保署发布的 2015 年香港空气质素报告显示:香港整体大气中颗粒物浓度呈下降趋势,反映了区域性的背景颗粒物水平在过去几年间已有所下降。香港特区政府近年实施了多项汽车废气管制措施,2015 年路边监测站监测的 PM<sub>10</sub> 水平均低于全年空气质素指标,同时 PM<sub>10</sub> 水平较 1991 年减少 51%。尽管香港特区政府为改善空气污染采取了多项措施,但依然不够。所以,较准确的估算,预测颗粒物水平,建立预报和预警系统,高效及时准确地为民众提供空气质素健康指数具有重要意义。为此,众多学者开发出了多种方法应用于预测颗粒物水平。如:张玮在基于颗粒物与气象因素相关性研究的基础上建立逐步回归和 BP 神经网络预测模型预测南京市颗粒物浓度<sup>[5]</sup>;而刘小兵通过对 PM<sub>2.5</sub> 与气象因素相关性的研究,利用支持向量机回归建立颗粒物预测模型<sup>[6]</sup>;李璐等建立气象相似准则,对颗粒物数据、气象数据进行三层筛选,寻找历史类似数据,然后利用 BP 神经网络进行预测,具有较优预测结果<sup>[7]</sup>。国外研究者开发出了一些新型混合模型应用于空气质量的预测,以消除单一模型的缺点,如:Ausati 等将 ANFIS、EEMD-GRNN、PCR 和 MLR 模型应用于 PM<sub>2.5</sub> 的预测研究,通过比较分析,发现混合模型 EEMD-GRNN 的准确性要高于线性模型<sup>[8]</sup>;Ul-Saufie 等将主成分分析 (Principal components analysis, PCA) 与多元线性回归 (Mutiple linear regression, MLR) 和前馈反向传播神经网络 (Feedforward back-propagation, FFBP) 结合在一起预测马来西亚森美兰 PM<sub>10</sub> 浓度水平,与单纯的 MLR 模型和 FFBP 相比,次日预测精度误差降低多达 18.1%(PCA-MLR) 和 17.68%(PCA-FFBP),对接下来两天和三天的预测精度也有较大程度的提升<sup>[9]</sup>。

以上研究主要考虑气象因素对颗粒物浓度的影响<sup>[10-11]</sup>,以气象因素作为输入变量,而本文将考虑颗粒物与污染物之间的关联性<sup>[12]</sup>,分别以污染物、气象因素,以及污染物加气象因素三种因子模式作为支持向量机回归 (Support vector regression, SVR) 的输入变量<sup>[13]</sup>,并使用一种新型智能优化算法灰狼优化 (Grey wolf optimization, GWO) 算法对支持向量机回归进行优化,形成 GWO-SVR 模型<sup>[14-15]</sup>,应用于香港旺角监测站监测的 PM<sub>10</sub> 数据进行预测,如此,可保证即使在某种预测变量数据缺失的情况依然能够对 PM<sub>10</sub> 浓度进行准确预测。

本文描述了颗粒物数据、污染数据和气象数据, 利用一些统计方法的基础知识进行统计分析, 并对本文所提出的模型进行重点介绍; 介绍了使用不同模型在进行预测获得的结果, 包括预测模型的选择、模型的训练, 以及验证其预测性能。最后, 进行总结。

## 2 材料和方法

### 2.1 数据来源

香港是位于华南地区的沿海城市, 是典型的高密度、高层次城市, 空气质量问题重大。作为世界人口密度最高的地区之一, 截至 2017 年底, 香港的人口密度为 7039 人 /km<sup>2</sup>, 总人口为 740.98 万人<sup>[16]</sup>。收集的气象数据由香港天文台提供<sup>[17]</sup>, 保留的数据包括平均气压、平均气温、平均露点温度、平均相对湿度、总雨量、盛行风向、平均风速。颗粒物数据和污染数据来源于香港环境保护署, 由香港 14 个监测站中的旺角监测站测得, 其地理位置靠近香港天文台<sup>[18]</sup>。主要数据包括: PM10、CO、NO<sub>2</sub>、NO<sub>x</sub>、O<sub>3</sub> 和 SO<sub>2</sub>。数据为 2011~2016 年剔除缺失值后的 2050 日。其中, 风速、风向代表两个参数, 但两者关系密不可分, 因此将其融合处理, 进行极坐标化, 即风向为极坐标的  $\theta$  角, 风速为极坐标的矢量  $\rho$ , 投影到直角坐标系上, 分别得到  $X$ 、 $Y$  轴两个方向上的风速值, 以其作为输入参数; 总雨量, 由于其离散型较大, 因此将其按等级划分, 即使用整数 0、1、2、3... 代表无雨、小雨、中雨、大雨... 的总雨量等级。计算了香港旺角的 PM10、气象因素和污染物六年的基本资料结果, 如表 1 所示。表 1 中信息表明, 当受到西南风影响时, PM10 浓度将低于盛行东北风向的天气状况, 这是因为所吹的西南季候风也可为香港补充较洁净的海洋气流。

表 1 数据统计汇总表

Table 1 Data summary table

		Mean	Max	Min	SD
PM10	PM10 concentration( $\mu\text{g}\cdot\text{m}^{-3}$ )	46.87	181	7	22.74
	PM10(southwest) concentration( $\mu\text{g}\cdot\text{m}^{-3}$ ) <sup>a</sup>	31.25	118	12	14.67
	PM10(northeast) concentration( $\mu\text{g}\cdot\text{m}^{-3}$ ) <sup>b</sup>	53.08	181	7	22.39
Meteorological factors	Mean pressure (hPa)	1012.76	1034.6	992.2	6.65
	Mean temperature (°C)	23.5	32.4	4.9	5.41
	Mean dew point temperature (°C)	19.35	26.8	-4.3	6.2
	Mean relative humidity (%)	78.52	99	29	10.35
	Rainfall level (mm)	6.18	230.8	0	18.57
	$X$ (Km·h <sup>-1</sup> )	4.64	55.91	-51.7	14.94
	$Y$ (Km·h <sup>-1</sup> )	10.58	57.5	-39.7	15.94
Pollutants	CO concentration (10 $\mu\text{g}\cdot\text{m}^{-3}$ )	98.40	197	23	24.68
	NO <sub>2</sub> concentration ( $\mu\text{g}\cdot\text{m}^{-3}$ )	105.06	260	35	31.77
	NO <sub>x</sub> concentration ( $\mu\text{g}\cdot\text{m}^{-3}$ )	265.08	860	49	91.70
	O <sub>3</sub> concentration ( $\mu\text{g}\cdot\text{m}^{-3}$ )	14.65	85	0	12.40
	SO <sub>2</sub> concentration ( $\mu\text{g}\cdot\text{m}^{-3}$ )	9.38	79	0	8.23

注: a 为 583 日, b 为 1467 日, 其它数据为 2050 日

## 2.2 算法描述

### 2.2.1 支持向量机回归

支持向量机 (Support vector machine, SVM) 由 Vapnik 在 1995 年提出<sup>[19]</sup>, 是一种基于统计学习理论的学习方法, 更精确地说, 是结构风险最小化的近似实现。SVM 的主要特点是内核功能的应用没有局部最小, 并且通过间隔或维度独立于控制容量的量。SVM 应用于函数回归时, 称为支持向量机回归 (Support vector regression, SVR) 主要针对非线性问题。一般是先使用非线性映射, 把数据映射到一个高维特征空间  $H$  通过  $\varphi$  变换:  $R^d \rightarrow H$ , 将  $x$  映射为  $\varphi(x)$ , 再在高维特征空间  $H$  中进行回归。SVR 模型具有较好的泛化能力, 它巧妙地解决了维数问题, 算法复杂度与样本维数无关。SVR 模型的最终形式为

$$y = f(x) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) k(x_i, x) + b, \quad (1)$$

$k(x_i, x)$  被称为内核函数, 采用的径向基 RBF 核函数定义为

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\alpha^2}\right) = \exp(-g\|x_i - x_j\|^2), g > 0 \quad (2)$$

使用 GWO 来优化 SVR 参数  $C$  和核函数参数  $g$ 。

### 2.2.2 GWO 算法

GWO 算法是由 Mirjalili 等通过模拟自然界中灰狼群体的社会等级机制和捕食行为而提出的一种新型群体搜索方法<sup>[20]</sup>。其中  $\alpha$  称为头狼, 领导灰狼群体;  $\beta$  是下属的灰狼, 协助头狼  $\alpha$  作出决策;  $\delta$  称为普通狼, 听从  $\alpha$  及  $\beta$  的指令, 也可以指挥其他底层狼  $\omega$ 。

在捕食过程中, 灰狼群对猎物采取包围行动为狩猎行为第一阶段, 可以表示为

$$D = |CX_p(i_t) - X(i_t)|, \quad (3)$$

$$A = 2\alpha r_1 - \alpha, \quad (4)$$

$$C = 2r_2, \quad (5)$$

$$X(i_t + 1) = X_p(i_t) - AD, \quad (6)$$

式中  $i_t$  为当前迭代,  $X_p$  和  $X$  分别为猎物和灰狼的位置向量,  $A$  和  $C$  是系数向量,  $r_1$ 、 $r_2$  是  $[0,1]$  中的随机数, 并且  $\alpha$  在迭代过程中会从 2 逐渐减小到 0。

灰狼群在包围猎物后, 通常会受到头狼命令猎捕猎物。但是, 无法知道猎物在虚拟搜索空间中的位置信息。因此, 为了模仿狼的狩猎行为, 假设前三等级狼 ( $\alpha, \beta, \delta$ ) 有更多关于猎物位置的信息。因此, 可以节省目前的三项最佳解决方案, 并根据最佳解决方案的位置更新选择灰狼的位置。更新公式为

$$D_j = |C_j X_j - X|, j = \alpha, \beta, \delta, \quad (7)$$

$$X_i = X_j - A_j D_j, i = 1, 2, 3; j = \alpha, \beta, \delta, \quad (8)$$

$$X(i_t + 1) = \frac{X_1 + X_2 + X_3}{3}. \quad (9)$$

通过搜索和攻击猎物可以获得该优化算法的探索和开发能力。

### 2.2.3 GWO-SVR 算法

由于 SVR 参数优化过程容易出现早熟现象, 为了提高优化参数的准确性和效率, 重点介绍 GWO 算法优化 SVR 参数, 优化参数为惩罚系数  $C$  和核函数参数  $g$ , 将其命名为 GWO-SVR 算法。

GWO-SVR 算法步骤如下:

- 1) 初始化灰狼种群和  $\alpha, \beta, \delta$  狼的位置;
- 2) 计算适应度函数值, SVR 依照每个灰狼的位置作为参数  $C$  和  $g$ , 进行模型训练, 以预测集均方误差 (Mean-square error, MSE) 作为优化的目标函数值, 最好的为  $\alpha$  狼, 其次为  $\beta$  狼, 第三为  $\delta$  狼;
- 3) 当所有个体都迭代完成, 更新个体位置, 寻优过程结束时, 返回最优个体  $\alpha$  狼的位置, 即为最优参数  $C$  和  $g$ ;
- 4) 采用优化后的  $C$  和  $g$  建立 SVR 预测模型。

### 2.2.4 评测指标

实现了 SVR 和 GWO-SVR 算法, 将其应用于香港旺角监测站 PM10 浓度的训练及预测。为了验证预测实验结果的优劣, 使用两个常用的性能评测指标进行定量比较: 计算出训练集和测试集中实际值和预测值之间的相关系数  $R$  和均方误差  $M_{SE}$

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (10)$$

$$M_{SE} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (11)$$

式中  $x_i$  和  $y_i$  分别为 PM10 浓度的预测输出值和实际监测值。

## 3 结果与分析

### 3.1 数据相关性

表 2 所示为 PM10 与气象因子和污染物的相关系数。可以发现: 首先, PM10 与气象因素和污染物有一定的相关性。一般来说, PM10 与污染物高度相关, 且呈正相关; 与气象因素相关性则相对较低, 且有正有负。其次, 在不同天气条件下, PM10 与各因素之间的相关性也不同。因此, 分三个部分进行讨论: 首先, 将污染因子、气象因素、混合气象因素与污染物作为预测因子, 并对预测结果进行了比较。其次, 比较 SVR 和 GWO-SVR 的预测结果, 评价两种方法的预测性能和 GWO 的优化性能。最后, 根据上述研究结果, 采用最优预测因子模式和预测模型并依据不同风向进行分类的数据, 进行分类预测, 并比较在不同气象条件下的预测性能。

### 3.2 基于整体数据的预测

随机筛选采集的全部数据中的 4/5(1640 个) 作为训练集, 剩余 1/5(410 个) 作为测试集, 并保留筛选后的训练集和测试集数据, 以保证每次实验过程数据的一致性。之后, 考虑分别以污染因素、气象因素以及将污染物和气象因素混合作为预测因子, 并在不同预测因子模式下, 使用 SVR 模型和 GWO 模型进行 PM10 每日浓度的预测。最后, 分别比较三种预测因子模式和两种预测模型的预测性能。

表 3 所示为基于不同因子模式分别使用 SVR 和 GWO-SVR 模型进行 PM10 预测的结果检验值  $R$  和  $M_{SE}$ 。首先, 可以看到, 在基于不同预测变量的模式下, 以污染和气象因素共同作为预测因子时的预测精

度最高, 而单独以污染因素作为预测因子时的预测性能略差, 这是由于其气象因素中包含污染物信息, 而污染物中也包含气象因素信息。同时, 在基于同种预测变量不同预测模型进行预测时, SVR 模型测试集表现往往低于训练集, 而 GWO-SVR 预测模型却没有体现出这一特性, 原因在于在灰狼个体适应度的寻优过程中, 以测试集的 SVR 预测结果检验值  $M_{SE}$  作为优化的目标函数值, 往往会导致测试集有较好表现。而且从 SVR 和 GWO-SVR 两种模型预测性能的比较上可以看出, GWO-SVR 对比 SVR 在训练过程中没有体现出优势, 但是在测试过程中均体现出一定优势, 预测精度得到提高, 且  $M_{SE}$  均有不同程度的降低, 表明 GWO-SVR 的预测性能明显优于 SVR 预测性能, GWO 优化效果良好。

表 2 PM10 与气象因素和污染物的相关系数  $R$

Table 2 Correlation coefficient of PM10 with meteorological factors and pollutants

		$R$		
		PM10 (whole data)	PM10 (southwest)	PM10 (northeast)
Meteorological factors	Mean pressure	0.44	-0.06	0.30
	Mean temperature	-0.40	-0.24	-0.21
	Mean dew point temperature	-0.50	-0.41	-0.35
	Mean relative humidity	-0.43	-0.18	-0.43
	Rainfall level	-0.35	-0.10	-0.39
	X	0.33	0.13	-0.05
	Y	0.30	0.05	0.00
Pollutants	CO	0.54	0.48	0.45
	NO <sub>2</sub>	0.69	0.76	0.65
	NO <sub>x</sub>	0.21	0.37	0.30
	O <sub>3</sub>	0.41	0.42	0.29
	SO <sub>2</sub>	0.48	0.55	0.57

综上所述, 可以得出两个重要结论: 1) 将气象因素和污染物共同作为预测因子时, 其预测精度高于气象因素和污染物单独作为预测因子时的预测精度; 2) GWO 预测性能良好, GWO-SVR 模型预测性能强于 SVR 模型。基于 1), 可以考虑在某些变量不能获得的情况下, 将能获得的几个因素作为预测因子进行 PM10 预测。图 1 给出的是在使用气象因素和污染物共同作为预测因子时, SVR 和 GWO-SVR 的实际监测值和预测输出值的散点图, 其中  $X$  轴为实际监测值,  $Y$  轴为预测输出值。从图中可以更加直观地发现在最终的预测结果中, GWO-SVR 模型预测集点的位置更紧密且靠近拟合线, 确定系数  $R^2=0.81$  大于 SVR 的  $R^2=0.71$ , 预测性能良好。

### 3.3 基于不同风向进行分类预测

在不同风向的气象条件下, 香港旺角监测站 PM10 分布情况如图 2 所示。从图中可以看出香港在 2011~2016 年间盛行风向介于  $10^\circ \sim 110^\circ$  和  $280^\circ \sim 360^\circ$  之间, PM10 浓度范围较大, 且更易出现污染天气, 而在盛行风向介于  $120^\circ \sim 270^\circ$  之间时, PM10 浓度值较低, 空气质量较好。同时, 也可发现, 香港盛行东南风 ( $280^\circ \sim 360^\circ$ ) 的天气相对较少。因此, 考虑将上述数据按风向分为两类, 并依据其主要风向和数据占比命名, 即西南风向 ( $120^\circ \sim 270^\circ$ ), 东北风向 ( $10^\circ \sim 110^\circ$  &  $280^\circ \sim 360^\circ$ )。然后基于 GWO-SVR 为最优

预测模型进行基于三种因子模式的 PM10 浓度预测。

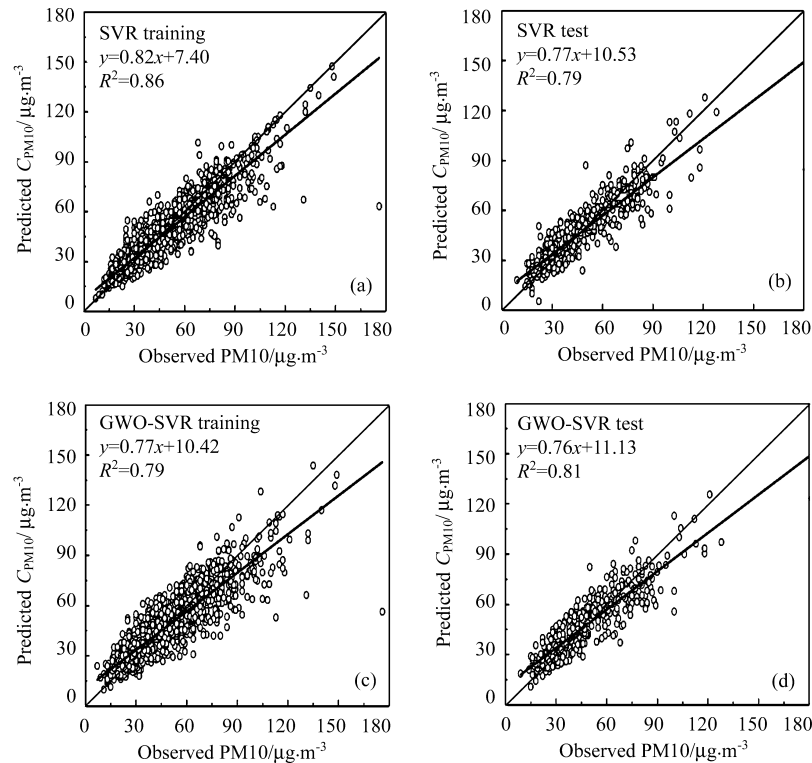


图 1 PM10 浓度实际值与预测值相关性拟合。(a)SVR 训练集; (b)SVR 测试集; (c)GWO-SVR 训练集; (d)GWO-SVR 测试集。

Fig.1 Correlation between the actual value of PM10 concentration and the predicted value.

(a) SVR training; (b) SVR test; (c) GWO-SVR training; (d) GWO-SVR test

表 3 PM10 预测结果检验值  $R$  和  $M_{SE}$

Table 3 PM10 predicted results test values  $R$  and  $M_{SE}$

Predictors	Statistical indices	SVR		GWO-SVR	
		Training	Test	Training	Test
Pollution	$R$	0.86	0.87	0.88	0.89
	$M_{SE}$	130.12	135.85	111.27	113.96
Meteorological	$R$	0.76	0.68	0.73	0.69
	$M_{SE}$	219.37	301.61	234.69	290.84
Pollution & Meteorological	$R$	0.92	0.89	0.89	0.90
	$M_{SE}$	74.34	118.25	107.91	107.94

表 4 所示为使用 GWO-SVR 模型基于三种预测因子模式进行 PM10 预测的预测结果检验值  $R$  和  $M_{SE}$ 。可以发现依据风向进行分类预测时, 训练集预测检验值相关系数存在一定的偶然性, 不能明确判断两种风向条件下的训练结果优劣情况; 而在预测集中, 盛行风向为西南风时, 检验值  $R=0.91$  均明显高于东北风向时的  $R=0.87$ , 并且通过检验值  $M_{SE}$  也可以发现, 盛行西南风时  $M_{SE}$  为 47.15, 低于盛行东北风时的 125.80, 即在依据风向分类进行预测时, 盛行西南风向时的预测精度较高。

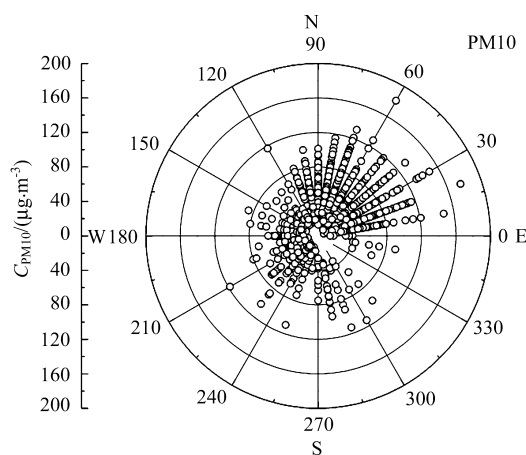


图 2 PM10 浓度随风向变化的分布散点图

Fig.2 Distribution scatter plot of PM10 concentration with wind direction

表 4 基于不同风向分类的预测结果检验值  $R$  和  $M_{SE}$ Table 4 Predicted values of  $R$  and  $M_{SE}$  based on different wind direction classification

Predictors	Statistical indices	Southwest(120° ~270°)		Northeast(10° ~110°&280° ~360°)	
		Training	Test	Training	Test
Pollution	$R$	0.82	0.90	0.85	0.86
	$M_{SE}$	67.15	54.29	132.72	135.34
Meteorological	$R$	0.71	0.64	0.68	0.63
	$M_{SE}$	103.4	153.32	272.9	320.2
Pollution & Meteorological	$R$	0.88	0.91	0.89	0.87
	$M_{SE}$	0.88	47.15	102.16	125.80

## 4 结 论

1) 建立以污染物、气象因素、污染物混合气象因素的三种预测因子模式，然后分别将三种因子模式作为 SVR 模型的输入变量建立预测模型用于香港旺角监测站 PM10 浓度预测，对比预测结果发现：将污染物和气象因素混合共同作为输入变量时预测精度最高，为最优预测因子模式。

2) 使用 GWO 算法用于支持向量机回归参数的寻优，得到的 GWO-SVR 模型与原 SVR 模型预测结果对比具有明显优势，其测试集在各预测因子模式下的评价指标  $R$  至少提高 0.01，预测精度有一定提高。

3) 基于上述两个结论，依据风向将数据进行分类，其中西南风向为一类，东北风向为一类，然后以三种预测因子模式作为 GWO-SVR 模型的输入变量，其预测结果显示盛行西南风时，预测集评测指标  $R$  为 0.91、 $M_{SE}$  为 47.15，优于盛行东北风时的  $R$  为 0.87、 $M_{SE}$  为 125.80 和所有数据下的  $R$  为 0.90、 $M_{SE}$  为 107.94，为不同气象条件下提高 PM10 的预测精度提供方向。



## 参考文献:

- [1] Sun Zhaobin, Tao Yan, An Xingqin, *et al.* Exposure-response relationship between respiratory system diseases and PM10 in different seasons [J]. *Journal of Environment and Health*, 2016, **33**(8): 663-669(in Chinese).  
孙兆彬, 陶燕, 安兴琴, 等. 不同季节 PM10 与呼吸系统疾病暴露 - 反应关系研究 [J]. *环境与健康杂志*, 2016, **33**(8):663-669.
- [2] Zheng M. Hong Kong: particulate air pollution and health impacts [J]. *Encyclopedia of Environmental Health*, 2011(3): 56-61.
- [3] Zhou M G, He G J, Liu Y N, *et al.* The associations between ambient air pollution and adult respiratory mortality in 32 major Chinese cities, 2006-2010 [J]. *Environmental Research*, 2015, **137**: 278-286.
- [4] Genc D D, Yesilyurt C, Tuncel G. Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere [J]. *Environmental Monitoring & Assessment*, 2010, **166**(1-4): 11.
- [5] Zhang Wei. *The Pollution Situation of Atmospheric Particulate Matter in Nanjing Area and Its Relation with Meteorological Factors* [D]. Nanjing: Master's Thesis of Nanjing University of Information Science & Technology, 2016(in Chinese).  
张玮. 南京地区大气颗粒物污染近况及其与相关气象因子的关系 [D]. 南京: 南京信息工程大学硕士论文, 2016.
- [6] Liu Xiaobing. *Analysis and Research of Forecast Model on PM2.5 Using Support Vector Regression* [D]. Chengdu: Master's Thesis of Southwestern University of Finance and Economics, 2016(in Chinese).  
刘小兵. 基于支持向量回归机的 PM2.5 浓度预测模型分析与研究 [D]. 成都: 西南财经大学硕士论文, 2016.
- [7] Li Lu, Liu Yonghong, Cai Ming, *et al.* A forecast model for urban air quality based on meteorological similarity criteria [J]. *Environmental Science & Technology*, 2013, **36**(5): 156-161(in Chinese).  
李璐, 刘永红, 蔡铭, 等. 基于气象相似准则的城市空气质量预报模型 [J]. *环境科学与技术*, 2013, **36**(5): 156-161.
- [8] Ausati S, Amanollahi J. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM2.5 [J]. *Atmospheric Environment*, 2016, **142**: 465-474.
- [9] Ul-Saufie A Z, Yahaya A S, Ramli N A, *et al.* Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA) [J]. *Atmospheric Environment*, 2013, **77**(3): 621-630.
- [10] Zhao Wenhui, Gong Huili, Zhao Wenji, *et al.* Spatial and temporal variation of Beijing inhalable particulate matter and its meteorological factors based on geostatistical analysis [J]. *Acta Scientiae Circumstantiae*, 2010, **30**(11): 2154-2163(in Chinese).  
赵文慧, 宫辉力, 赵文吉, 等. 基于地统计学的北京市可吸入颗粒物时空变异性及气象因素分析 [J]. *环境科学学报*, 2010, **30**(11): 2154-2163.
- [11] Wang Fang, Wu Huanbo, Chen Qiang. The influence of meteorological conditions on PM10 mass concentration in Lanzhou City [J]. *Ecology and Environmental Sciences*, 2016, **25**(1): 99-102(in Chinese).  
王芳, 吴焕波, 陈强. 气象因素对兰州市 PM10 质量浓度的影响 [J]. *生态环境学报*, 2016, **25**(1): 99-102.
- [12] Ni X Y, Huang H, Du W P. Relevance analysis and short-term prediction of PM2.5 concentrations in Beijing based on multi-source data [J]. *Atmospheric Environment*, 2017, **150**: 146-161.
- [13] Hou Weizhen, Li Zhengqiang, Zhang Yuhuan, *et al.* Using support vector regression to predict PM10 and PM2.5 [J]. *Iop Conference Series: Earth & Environmental Science*, 2014, **17**(1): 12268-12273.

- [14] Cong S L, Sun J, Mao H P, *et al.* Non-destructive detection for mold colonies in rice based on hyperspectra and GWO-SVR [J]. *Journal of the Science of Food and Agriculture*, 2018, **98**(4): 1453-1459.
- [15] Xu Dayu, Ding Shuai. Research on improved GWO-optimized SVM-based short-term load prediction for cloud computing [J]. *Computer Engineering and Applications*, 2017, **53**(7): 68-73(in Chinese).  
徐达宇, 丁 帅. 改进 GWO 优化 SVM 的云计算资源负载短期预测研究 [J]. *计算机工程与应用*, 2017, **53**(7):68-73.
- [16] Vapnik V N. *The Nature of Statistical Learning Theory* [M]. Springer, 1995. **8**(6): 988-999.
- [17] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer [J]. *Advances in Engineering Software*, 2014, **69**: 46-61.
- [18] The Government of the Hong Kong Special Administrative Region Census and Statistics Department[OL]. [https://www.censtatd.gov.hk/home/index\\_tc.jsp](https://www.censtatd.gov.hk/home/index_tc.jsp).  
香港特别行政区政府统计处 [OL].[https://www.censtatd.gov.hk/home/index\\_tc.jsp](https://www.censtatd.gov.hk/home/index_tc.jsp).
- [19] The Government of the Hong Kong Special Administrative Region Hong Kong Observatory. Daily data excerpt from March 2016 [OL]. [http://www.hko.gov.hk/cis/dailyExtract\\_uc.htm?y=2016&m=3](http://www.hko.gov.hk/cis/dailyExtract_uc.htm?y=2016&m=3).  
香港特别行政区政府香港天文台. 二零一六年三月每日数据摘录 [OL]. [http://www.hko.gov.hk/cis/dailyExtract\\_uc.htm?y=2016&m=3](http://www.hko.gov.hk/cis/dailyExtract_uc.htm?y=2016&m=3).
- [20] Environmental Protection Department. Query and download air quality monitoring data[OL]. <http://cd.epic.epd.gov.hk/EPICDI/air/download/>.  
环境保护署. 查询和下载空气质素监测数据 [OL]. <http://cd.epic.epd.gov.hk/EPICDI/air/download/>