

# OPAQ 系统两种模式对 O<sub>3</sub> 预报准确率的探讨

王淑莹<sup>1</sup>, 许纯领<sup>2\*</sup>, 尹翠芳<sup>1</sup>, 李鹏帅<sup>1</sup>

(1. 北京立博威拓环境技术有限公司, 北京 100085; 2. 江苏省宿迁环境监测中心, 江苏 宿迁 223800)

**摘要:**以江苏省宿迁环境监测中心 OPAQ 系统为例, 基于神经网络算法的 OPAQ 空气质量预报系统的 2 种模式对 O<sub>3</sub> 预报准确率的进行了分析, 结果表明, 趋势最优模式(RMSE 模式)对预报当天及未来 3 d 的预报值与监测值的相关性系数均 > 0.78, 相对误差在 25% 以下, 在预测当天及未来 24、48 及 72 h 优 - 良天的预测准确率较高, 分别为 88.8%、87.2%、86.3% 及 84.7%, 在预测轻度污染 - 重度污染的准确率较低; 极值最优模式(SI 模式)对预报当天及未来 3 d 的预报值与监测值的相关性系数(R)均 > 0.76, 相对误差 < 32%, 预测未来 24 和 48 h 的轻度污染 - 中度污染的级别准确率 > 60%。OPAQ 系统的极值最优模式(SI 模式)更适合作为夏季  $\rho(\text{O}_3)$  较高时的预测工具。

**关键词:**神经网络算法; 空气质量预报业务系统; O<sub>3</sub>; 预报; 准确率

中图分类号: X839.2

文献标志码: B

文章编号: 1674-6732(2020)02-0013-04

## Discussion on the Accuracy of O<sub>3</sub> Prediction by Two Modules of OPAQ System

WANG Shu-ying<sup>1</sup>, XU Chun-ling<sup>2\*</sup>, YIN Cui-fang<sup>1</sup>, LI Peng-shuai<sup>1</sup>

(1. Beijing LIBOVITO Environmental Technology Company, Beijing 100085, China; 2. Jiangsu Suqian Environmental Monitoring Center, Suqian, Jiangsu 223800, China)

**Abstract:** In this paper, two models of OPAQ air quality prediction system based on artificial neural network algorithm are used to analyze the accuracy of O<sub>3</sub> prediction. Taking the OPAQ system of Jiangsu Suqian Environmental Center as an example, the results show that the correlation coefficients  $R$  of RMSE model for predicting the current day and the next three days are above 0.78 and the relative errors are below 25%. The prediction accuracy of RMSE model for the current day and the 24-hour, 48-hour and 72-hour excellent-good days is high, which are 88.8%, 87.2%, 86.3% and 84.7% respectively. The accuracy of RMSE model for predicting the light-heavy pollution is low. The correlation coefficients  $R$  of the extreme optimal model (SI model) for predicting the current day and the next three days are above 0.76, and the relative errors are below 32%. The accuracy of predicting the level of mild-moderate pollution in the next 24 hours and 48 hours is above 60%. In summary, the extreme optimal model (SI model) of OPAQ system is more suitable as a prediction tool when the concentration of O<sub>3</sub> is high in summer.

**Key words:** Neural network algorithm; OPAQ system; Ozone; Forecast; Accuracy

空气质量与人类健康之间的关系已经引起了广泛关注, 已有不少学者对此做了大量研究<sup>[1]</sup>。空气质量预报在污染事件发生前对公众的预警及制定减灾计划方面尤为重要, 国内外学者也利用数值机理模型及统计模型来对空气质量进行预报, 试图精准预测污染事件的发生, 为减灾计划提供支持<sup>[2-7]</sup>。因此, 评估预报模型的预报准确率对进一步提高模型预报准确性也具有较为重大的意义。现以江苏省宿迁环境监测中心空气质量预报业务

系统(OPAQ 系统)为例, 基于神经网络算法的 OPAQ 系统的两种模式对臭氧(O<sub>3</sub>)预报准确率进行分析, 探讨 2 种模式的优劣。

### 1 OPAQ 系统

OPAQ 系统采用神经网络算法中的多层感知方法, 其中一个隐含层代表一个输出层。预测的结果是污染物的日均值或日最大值。假如模型在当天早上运行, 则当天 00:00 至模型开始运行最

收稿日期: 2019-09-09; 修订日期: 2019-09-29

作者简介: 王淑莹(1982—), 女, 工程师, 硕士, 从事环境空气质量预报预警工作。

\* 通讯作者: 许纯领 E-mail: 361167649@qq.com

近时刻的监测数据以及未来几天的气象数据作为输入数据。对每个监测站点配置几个不同的模型进行训练,不同模型的区别在于所采用的输入参数不同。所预测当天早上的监测数据及行星边界层高度(BLH)作为预测未来第  $N$  天的一个重要输入参数。行星边界层高度可认为是理查德系数( $Ri$ )超过 0.5 时的高度<sup>[2]</sup>。其中  $Ri$  是浮力(由垂直温廓线确定)与惯性力(由大气湍流确定)的比率。如果  $Ri$  非常小(甚至是负的),湍流非常强烈,并且足以打破稳定的温廓线将颗粒物带至垂直方向,则从底部到行星边界层高度之间,颗粒物可以随湍流在此之间扩散。研究表明行星边界层高度在基于神经网络算法的空气质量预报系统中有非常重要的作用<sup>[8]</sup>。另外,还使用了风速与风向、温度、湿度及云量等气象参数来提高预测的准确性<sup>[8-9]</sup>。对每个监测站点都配置几种模型进行训练,最后选择一个最优的模型作为日常预报模型。

OPAQ 系统具有两种模式,一种是预报趋势最优模式(RMSE 模式),另一种是预报极值最优模式(SI 模式)。预报趋势最优模式指的是在模型训练和验证阶段,选择预报值与监测值的均方根误差(RMSE)最小的模型作为日常业务预报时所采用的模型;预报极值最优模式指的是在模型训练和验证阶段,选择预报成功指数(SI)最大的模型作为日常业务预报时所采用的模型。

## 2 研究方法

### 2.1 数据时间

2018 年 1 月 1 日—2019 年 7 月 29 日。

### 2.2 统计分析的项目

O<sub>3</sub> 预报的散点图、正确预警指数(FFE)、成功指数(SI)、错误预警指数(FFA)、相关性系数( $R$ )以及级别预报准确率。统计方法技术说明见图 1。

$$SI = \frac{N3}{N3 + N4} + \frac{N1}{N1 + N2} - 1 \quad (1)$$

式中:SI——成功指数,为正确预报非超标天数和超标天数的比例,%;

$N1$ ——监测值达标预测值也达标的天数;

$N2$ ——监测值达标而预测值超标的天数。

$$FFE = \frac{N3}{N3 + N4} \quad (2)$$

式中:FFE——正确预警指数,为正确预报超标天数的比例,%;

$N3$ ——监测值超标且预测值也超标的天数;

$N4$ ——监测值超标而预测值达标的天数。

$$FFA = \frac{N2}{N1 + N2} \quad (3)$$

式中:FFA——错误预警指数,为错误预报超标,而实际未超标的天数比例,%。

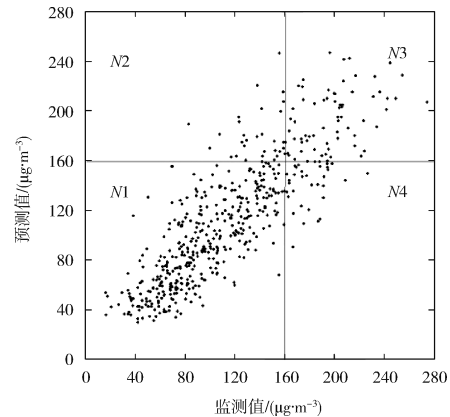


图 1 统计方法技术说明

## 3 结果与讨论

### 3.1 RMSE 模式预报准确率评价

O<sub>3</sub> 预报分析评价结果中, $m$  表示监测值, $d0$  表示预报当天, $d1$  表示预报未来 24 h, $d2$  表示预报未来 48 h, $d3$  表示预报未来 72 h。RMSE 模式预报结果评价见表 1。

表 1 RMSE 模式预报结果评价

指标	d0	d1	d2	d3
$R$	0.84	0.81	0.80	0.78
FFE/%	65.7	61.4	58.6	50.5
SI/%	58.8	54.8	51.9	42.8
FFA/%	6.9	6.6	6.7	7.7
平均误差	21.5	21.5	24.0	24.3
相对误差/%	24.2	22.2	24.3	24.4

由表 1 可见,在预测当天及未来 24 h 的 FFE 都能达到 60% 以上,预测未来 48 和 72 h 的 FFE 略低;在预测当天及未来 24,48 及 72 h 的 FFA 较低,均 < 10%。

图 2(a)(b)(c)(d)为 OPAQ 系统 RMSE 模式当天及未来 24,48 及 72 h 的预测值与实际监测值对比的散点图,十字线为 O<sub>3</sub> 滑动 8 h 最大的超标临界值 160  $\mu\text{g}/\text{m}^3$ 。

由图 2 可见,RMSE 模式预测当天及未来 24,48 及 72 h 的结果与实际监测值的拟合度较好,有个别值离散度较大。

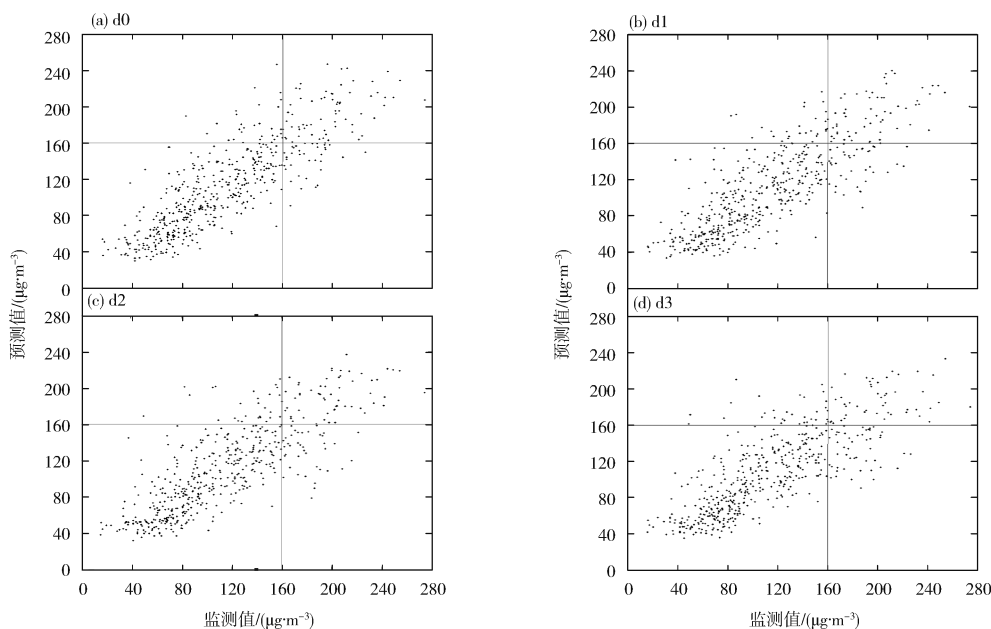


图 2 监测值与 RMSE 模式 d0、d1、d2、d3 预测值对比散点图

表 2 中展示了 OPAQ 系统统计预报模型 RMSE 模式预报当天及未来 24、48 及 72 h 的 O<sub>3</sub> 预报级别准确率的评价结果。表中天数列为监测值 ( $m$ ) 与预测值 (d0, d1, d2, d3) 在 O<sub>3</sub> 各级别出现的天数, 准确

率为预测当天及未来 3 d 对 O<sub>3</sub> 各级别的预测准确率。由表 2 可见, OPAQ 系统的 RMSE 模式在预测当天及未来 24、48 及 72 h 优 - 良天的预测准确率较高, 在预测轻度污染 - 重度污染的准确率较低。

表 2 RMSE 模式预报准确率评价

O <sub>3</sub> 污染级别	天数					准确率/%			
	$m$	d0	d1	d2	d3	d0	d1	d2	d3
优 - 良	439	390	383	379	372	88.8	87.2	86.3	84.7
轻度污染	87	51	48	43	37	58.6	55.2	49.4	42.5
中度污染	16	5	4	4	4	31.3	25.0	25.0	25.0
重度污染	1	0	0	0	0	0.0	0.0	0.0	0.0

### 3.2 SI 模式预报准确率评价

表 3 为 OPAQ 系统统计预报模型 SI 模式预报当天及未来 24、48 及 72 h 的预报结果评价指标。由表 3 可见, 在预测当天及未来 24、48 及 72 h 的正确预警指数 FFE 都 > 60%。

表 3 SI 模式预报结果评价

指标	d0	d1	d2	d3
$R$	0.85	0.81	0.78	0.76
FFE/%	84.3	72.3	69.4	64.3
SI/%	71.6	61.3	57.3	50.6
FFA/%	12.7	11.0	12.1	13.6
平均误差	10.5	29.7	29.7	31.7
相对误差/%	27.8	29.5	29.8	31.9

图 3(a)(b)(c)(d) 为 OPAQ 系统 SI 模式当天及未来 24、48 及 72 h 的预测值与实际监测值对比的散点图, 十字线为 O<sub>3</sub> 滑动 8 h 最大的超标临界值 160  $\mu\text{g}/\text{m}^3$ 。

由图 3 可见, RMSE 模式预测当天及未来 24、48 及 72 h 的结果与实际监测值的拟合度较好, 有个别值离散度较大。

表 4 为 OPAQ 系统统计预报模型 SI 模式预报当天及未来 3 d 的 O<sub>3</sub> 预报级别准确率的评价结果。由表可见, OPAQ 的 SI 模式在预测当天及未来 24、48 及 72 h 优 - 良天的预测准确率较高, 在长达一年半的统计时段里, O<sub>3</sub> 重度污染发生过 1 次, 预测值为中度污染。

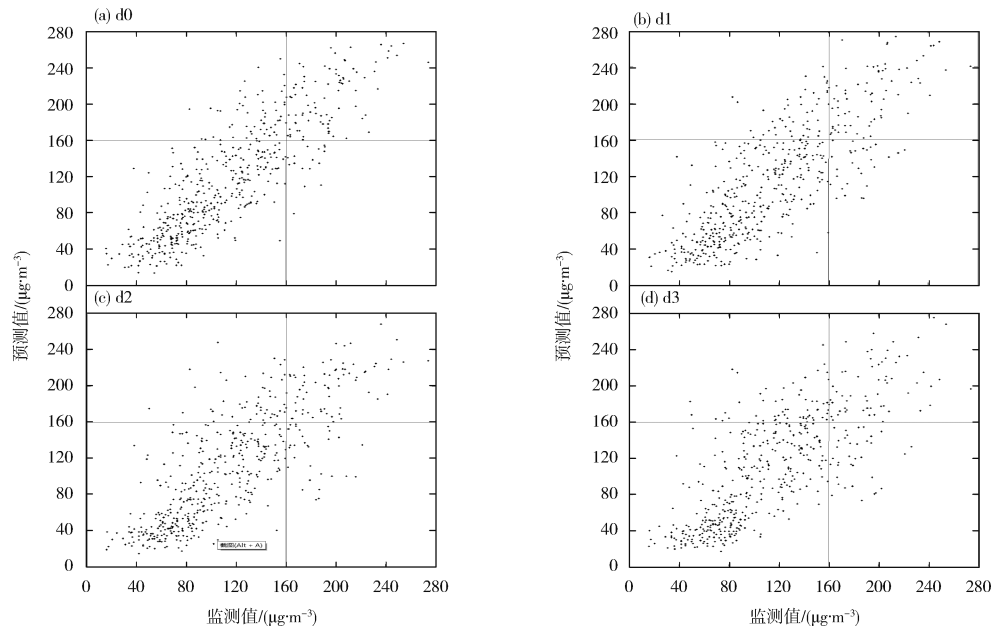


图 3 监测值与 SI 模式 d0、d1、d2、d3 预测值对比散点图

表 4 SI 模式预报准确率评价

O <sub>3</sub> 污染级别	天数					准确率/%			
	<i>m</i>	d0	d1	d2	d3	d0	d1	d2	d3
优 - 良	439	365	364	357	348	83.1	82.9	81.3	79.3
轻度污染	87	69	58	53	50	79.3	66.7	60.9	57.5
中度污染	16	12	11	9	6	75.0	68.8	56.3	37.5
重度污染	1	0	0	0	0	0.0	0.0	0.0	0.0

#### 4 结论

(1) 2 种模式预报评价指标显示, SI 模式的正确预警指数 FFE 和成功指数 SI 要高于 RMSE 模式; 而 RMSE 模式的错误预警指数 FFA 要低于 SI 模式, 该指标越低越好; 并且从长期来看, 2 种模式的相关性系数相差不大。

(2) 2 种模式预报准确率表明, 在 O<sub>3</sub> 优 - 良级别天时, RMSE 模式的预测准确率较高; 而在轻度污染以上时, SI 模式的预测准确率明显高于 RMSE 模式。建议在夏季 O<sub>3</sub> 污染较严重的时段里, 采用 OPAQ 系统的 SI 模式来预报  $\rho(\text{O}_3)$ 。

#### [参考文献]

- [1] World Health Organization. Review of evidence on health aspects of air pollution - REVIHAAP Project 309[R]. 2013.
- [2] 万显烈, 杨凤林, 王慧卿, 等. 利用人工神经网络对空气中 O<sub>3</sub> 浓度进行预测[J]. 中国环境科学, 2003, 23(1): 110 - 112.
- [3] STOCKWELL W R, MIDDLETON P, CHANG J S, et al. The Second generation regional acid deposition model chemical mechanism for Regional Air Quality Modeling[J]. J of Geophysical Research, 1990, 95(D10): 16343 - 16367.
- [4] 王淑莹, 许荣, 尹翠芳, 等. 基于 OPAQ 的城市空气质量预报系统研究[J]. 中国环境监测, 2016, 32(3): 13 - 20.
- [5] 沈路路, 王聿绚, 段雷. 神经网络模型在 O<sub>3</sub> 浓度预测中的应用[J]. 环境科学, 2011, 32(8): 2231 - 2235.
- [6] 叶斯琪, 陈多宏, 谢敏, 等. 珠三角区域空气质量预报方法及预报效果评估[J]. 环境监控与预警, 2016, 8(3): 10 - 13.
- [7] 王厚俊, 程滢, 吴莹, 等. 几种空气质量预报方法对冬季预报效果的评估与对比[J]. 环境监控与预警, 2018, 10(5): 36 - 40.
- [8] 程念亮, 李云婷, 邱启鸿, 等. 2013 年北京市 PM<sub>2.5</sub> 重污染日时空分布特征研究[J]. 中国环境监测, 2015, 31(3): 38 - 42.
- [9] 陈强, 梅琨, 朱慧敏, 等. 郑州市 PM<sub>2.5</sub> 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.