

基于最优定权组合法的大气污染物 SO₂ 预测^①



谢磊¹, 铁治欣¹, 宋飞扬¹, 丁成富²

¹浙江理工大学 信息学院, 杭州 310018)

²(聚光科技(杭州)股份有限公司, 杭州 310052)

通讯作者: 铁治欣, E-mail: tiezx@zstu.edu.cn

摘要: 为了提高对大气污染物 SO₂ 的预测准确率, 基于多个空气质量预测模式 (WRF-CHEM、CMAQ、CAMx), 以过去一段时间内各单项空气质量预测模式的组合预测误差平方和最小为原则, 构建出针对大气污染物 SO₂ 的最优定权组合预测模型. 选取 2018 年云南省楚雄、昭通、蒙自三个站点 1 至 5 月份的实际观测数据和前述三个空气质量模式的预测数据作为实验样本, 然后分别采用多元线性回归法和动态权重更新法在相同的实验条件下与所提的最优定权组合预测法进行预测对比实验. 实验结果表明, 所提方法的预测值相较其他两种方法更加贴近实际观测值, 其两项误差评估指标值均最小. 总体而言, 最优定权组合预测模型很好地综合了各单项空气质量预测模式的优势, 提高了 SO₂ 的预测精度.

关键词: 空气质量预测模式; 动态权重更新; 多元线性回归; 最优定权组合

引用格式: 谢磊, 铁治欣, 宋飞扬, 丁成富. 基于最优定权组合法的大气污染物 SO₂ 预测. 计算机系统应用, 2019, 28(3): 80-87. <http://www.c-s-a.org.cn/1003-3254/6814.html>

Prediction of Atmospheric Pollutant SO₂ Based on Optimal Weighted Combination Method

XIE Lei¹, TIE Zhi-Xin¹, SONG Fei-Yang¹, DING Cheng-Fu²

¹(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

²(Focused Photonics (Hangzhou) Inc., Hangzhou 310052, China)

Abstract: To improve the prediction accuracy of atmospheric pollutant SO₂, a combined forecasting model with optimal weights for atmospheric pollutant SO₂ was constructed, which was based on multiple air quality prediction (WRF-CHEM, CMAQ, CAMx) modes according to the principle of minimum square sum of combined forecasting errors of each single air quality prediction model in the past years. The actual observation data and the aforementioned three air quality modes prediction data of Chuxiong, Zhaotong, and Mengzi stations in Yunnan Province from January to May in 2018 were selected as the experimental samples. Then the multivariate linear regression method and the dynamic weight updating method were used to compare the prediction results with the optimal weighted combination prediction method proposed in this study under the same experimental conditions. The experimental results show that the predicted values of the proposed method are closer to the observed values than those of the other two methods, and the two error evaluation indexes are the smallest. Generally speaking, the combined forecasting model with optimal weights synthesizes the advantages of each single air quality forecasting model and improves the forecasting accuracy of SO₂.

Key words: air quality prediction mode; dynamic weight update; multiple linear regression; optimal weighting combination

① 基金项目: 浙江省公益技术应用研究项目 (2014C31G2060072)

Foundation item: Science and Technology Plan for Public Welfare of Zhejiang Province (2014C31G2060072)

收稿时间: 2018-09-13; 修改时间: 2018-10-18; 采用时间: 2018-11-02; csa 在线出版时间: 2019-02-22

SO₂ 作为大气中主要污染物之一,既能导致空气质量变差,又能在空气中迅速氧化生成 SO₃ 进而通过异质核化过程形成液态气溶胶^[1],降低了大气的能见度.同时,高浓度的 SO₂ 也是形成酸雨的罪魁祸首^[2],严重影响水环境以及农作物的产量,给人们的身心以及日常活动带来严重的影响.所以,实现对大气中 SO₂ 的精准预测,从而达到对空气质量的精准预警预报变得尤为重要.目前,在空气质量预报领域展现出良好应用前景的预报模式包括: CMAQ、CAMx、WRF-CHEM、NAQPMS 等^[3].但是由于各模式本身的不确定因素较大,在一定程度上影响了空气质量的预报结果,导致对 SO₂ 的预测不精准,为了提高对大气中各污染物浓度的预测准确度,以集合的形式把多个空气质量预报模式的预报结果集合在一起的预报方法已经成为提高预测大气污染物浓度的有效决策和方法^[4].

目前,运用在大气污染物集合预报中较为广泛的方法包括以下几种: a) 算术平均法.它具有算法简单且能够很好地反映出预测结果的整体趋势等优点.如:王自发等^[5]基于多个空气质量预测模式(CMAQ、CAMx、NAQPMS)构建出业务化的北京空气质量集合预报系统(EMS-Beijing),该系统预测结果主要依赖于各模式的算术平均值.但是仅仅利用算术平均法进行集合预报获得的结果值极易受到极个别异常值的影响. Mallet 等^[6]在不同的初始条件下,互相组合构建了几十种空气质量预报模式系统,发现只是简单的平均集合方法并不能显著地改进预报结果,并且平均集合方法的使用范围总存在着一定的局限性,需要引入一些其它的方法才能改进预报性能. b) 多元线性回归法.该方法通过研究多个自变量和因变量的线性关系,计算出一个拟合度较高的回归预测模型,能够很好地度量各个因素之间的相关程度.如:黄思等^[7]利用多模式集合和多元线性回归改进北京 PM₁₀ 预报,通过将历史观测信息纳入到北京空气质量集合预报系统(EMS-Beijing)预报体系,并对 EMS-Beijing 三个集合成员的预报结果进行集成,能够很好的分析出影响集成预报的影响因子,提高了预测的准确率. c) 动态权重分配法.这种方法通过分析在一段时间内各模式预测值的偏差率动态调整集合模式的权值,有效地提高了集合模式的预测准确率.如:姚文强等^[8]针对杭州市的空气质量情况,提出了一种动态权重更新模型并与已有的平均预测模型进行比较,结果表明比简单平均集成效果更好.但是通过单个预测模型的集成方法的适用性存在

一定的局限性,且预测准确率不稳定^[9].张春萍等^[10]在进行区域物流量的预测当中,基于多个单项预测模型的预测结果,采用最优定权组合模型以误差平方和最小为原则求得各单项预测模型的最优权值,有效地解决了单项预测模型准确率不稳定,适用性不高等问题,查阅相关资料得知有关最优定权组合预测模型在空气质量预测方面鲜有介绍.

因此,本文通过采用三种国内应用成熟、业务化程度高的空气质量预测模式(CMAQ、CAMx、WRF-CHEM),结合云南省的地势、气候、污染源等特点,提出了一种基于最优定权组合预测模型的集合预测方法将上述三种预测模式的结果进行集成,并通过误差分析验证所提方法的有效性.

1 模式系统与数据介绍

1.1 系统的介绍

云南省省级空气质量预报预警平台所采用的多模式集合预报系统成员分别是由美国国家海洋和大气管理局(NOAA)预报系统实验室开发的 WRF-CHEM^[11],美国国家环境保护局研发的 CMAQ^[12]系统以及美国环境技术公司 Environ 维护并研发的 CAMx^[13]组合而成.上述三种模型被广泛的应用于国内外的空气质量模拟领域,但各个模式输出的格式各有特点,在定制化的过程中,因为变量名、变量组别的不同,为集合预报带来了极大的挑战.为了实现模式的统一化管理与展示,参考国家环保部发布的《环境空气质量指数(AQI)技术规范》,将三种不同的模式集成到同一个系统平台上,气象场采用中尺度 WRF 模式,初始场采用 GFS 数据集.云南省多模式集合预报系统框架如图 1 所示.

首先基于不同污染源的排放特征,通过 SMOKE 源排放模型对排放源清单进行处理,为各单项空气质量模式提供所需要的大气排放源清单的时空分布数据.然后利用气象模式 WRF 对 GFS 数据集进行分析计算,并按照各单项空气质量模式的运行需求将上述计算结果转换成所需要的气象文件.最后将气象模式和源排放模型的计算结果作为各单项空气质量模式的输入计算出各模式对应的 SO₂ 预测值,并分别采用最优定权组合预测法、多元线性回归法和动态权重更新法对上述各单项空气质量模式的 SO₂ 预测值进行集成,从而实现多模式集合预报.这一系统已经应用于云南省环境检测中心,为云南省的空气质量预报提供更加准确的预报数据.

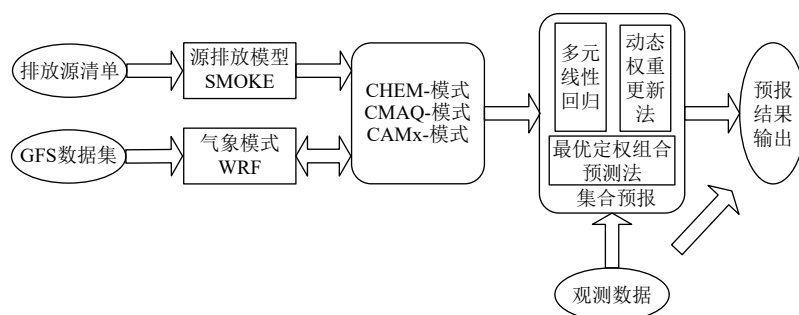


图1 云南省多模式集合预报系统框架

1.2 网格嵌套设置

本文重点针对云南省环境监测中心业务化运行的云南省省级环境空气质量预报预警业务平台进行分析与集成研究. 为了便于系统采用的三种不同模式的集成以及减小因为不同的区域设置而带来的误差, 本系统采用统一的区域设置、统一的排放清单、统一的气象场驱动, 以云南省为中心的三重网格设置. 第一重网格包括中国中西部地区, 范围如图2中的D1所示, 网格精度为27 km, 网格数(东西×南北)为141×134, 第二重网格包括中国西南地区, 范围如图2中的D2所示, 网格精度9 km, 网格数为219×219, 第三重网格包括云南全省, 范围如图2中的D3所示, 网格精度为3 km, 网格数为327×327.

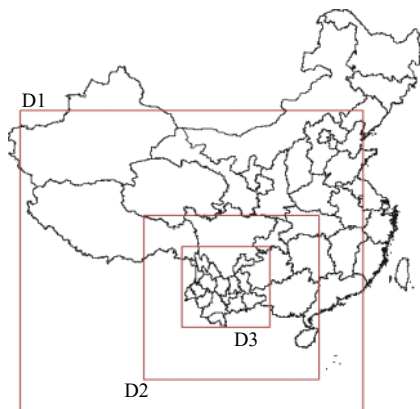


图2 云南省多模式集合预报系统网格设置简图

1.3 运行环境

云南省环境监测中心预报预警业务平台采用多模式集合预报系统的环境为 Red Hat Linux 7.3, 编译环境为 Intel Fortran, 并行环境为 MPI 和 OpenMP 混合并行, 运行所用核数为48核/96线程, 同时安装了 Intel C/C++编译器、GNU C/C++编译器.

1.4 实验数据

本文所采用的数据均来自于云南省省级环境空气质量预报预警业务平台数据库, 为保证数据的有效性, 该数据库数据均经过自动审核系统的过滤和筛选. 其中观测数据分别来源于云南省蒙自、楚雄、昭通三个区域站点在2018年1月至6月期间的SO₂日均实测值. 模式输出数据来自模式第三重网格的预报结果, 根据我国执行的空气污染指数标准, 利用当天的12时至第二天的11时的SO₂预报值计算日均值.

2 集成方法和评价指标

2.1 本文的最优定权组合预测模型

组合预测最早是由Bates和Granger提出来的^[14], 其基本思想是即使某一单项预测方法的预测精度很差, 我们也不能随意丢弃任何一种预测结果, 因为这种做法极有可能造成某些有用信息的丢失, 把它和某一种或几种预测精度较高的方法进行组合, 完全有可能提高系统的整体预测能力, 比较科学的做法是通过某种方法将各单项预测方法进行组合集成, 综合利用各单项预测方法的优势, 从而获得优于各单项预测方法的预测能力.

组合预测的关键^[15]是如何确定各组合成员的权重系数, 对于多模式空气质量数值模式集合来说, 传统的方法是主观赋权法, 通过相关领域的专家根据经验进行主观判断得到相应的权重系数, 这种方法的主观因素较多, 随意性较大, 不同的专家给出的结果不同. 还有一种是普通的平均定权法, 虽然方法简单, 但是如果某一组合预测成员的误差极大, 通过平均定权后很容易影响整个系统的预测精度. 而最优定权组合法基于误差平方和最小原则求得最优加权系数, 它综合了各种单项预测模型的优势, 充分利用各种单项模型提供

的有用信息达到提高预测准确度的效果, 不仅能克服上述缺点, 而且还能在一定程度上分散组合预测模型中选择不当的风险。

利用最优定权组合法集成的思路如下: 通过应用多个单项空气质量数值模式 (WRF-CHEM、CMAQ、CAMx) 对 SO₂ 的浓度进行预测, 并计算出相应的单项模式预测误差, 然后将各单项预测误差以不同的权值建立出组合预测误差方程, 并以误差平方和最小为原则建立目标函数对组合预测误差方程进行优化, 最后在一定的约束条件下求出目标函数的最优解。

具体建模步骤如下: 假设 SO₂ 的观测值序列为 $Y = \{y_t, t = 1, 2, \dots, n\}$, 其中 y_t 为第 t 时刻的实际观测数据。利用 m 种单项预测模型对 SO₂ 浓度进行预测, 设第 i ($i = 1, 2, \dots, m$) 个单项预测模型在第 t 时刻的预测值为 \hat{y}_{it} , 且第 i 个单项预测模型的权值为 k_i , 并且满足 $\sum_{i=1}^m k_i = 1$, 则第 t 时刻组合预测值可表示为:

$$\hat{y}_t = k_1 \hat{y}_{1t} + k_2 \hat{y}_{2t} + \dots + k_m \hat{y}_{mt} \quad (1)$$

令第 i 个单项模型在 t 时刻的单项预测误差为 e_{it} , 则 e_{it} 可表示为:

$$e_{it} = \hat{y}_{it} - y_t \quad (2)$$

令 e_t 为组合预测模型在第 t 时刻的组合预测误差, 则 e_t 可表示为:

$$e_t = \sum_{i=1}^m k_i e_{it} = (k_1, k_2, \dots, k_m)(e_{1t}, e_{2t}, \dots, e_{mt})^T = (e_{1t}, e_{2t}, \dots, e_{mt})(k_1, k_2, \dots, k_m)^T \quad (3)$$

令 J 为组合预测模型的误差平方和, 则有:

$$J = \sum_{t=1}^n e_t^2 = (k_1, k_2, \dots, k_m) \begin{pmatrix} \sum_{t=1}^n e_{1t}^2 & \sum_{t=1}^n e_{1t}e_{2t} & \dots & \sum_{t=1}^n e_{1t}e_{mt} \\ \sum_{t=1}^n e_{2t}e_{1t} & \sum_{t=1}^n e_{2t}^2 & \dots & \sum_{t=1}^n e_{2t}e_{mt} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n e_{mt}e_{1t} & \sum_{t=1}^n e_{mt}e_{2t} & \dots & \sum_{t=1}^n e_{mt}^2 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{pmatrix} \quad (4)$$

$$\text{令 } \mathbf{K} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} \sum_{t=1}^n e_{1t}^2 & \sum_{t=1}^n e_{1t}e_{2t} & \dots & \sum_{t=1}^n e_{1t}e_{mt} \\ \sum_{t=1}^n e_{2t}e_{1t} & \sum_{t=1}^n e_{2t}^2 & \dots & \sum_{t=1}^n e_{2t}e_{mt} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n e_{mt}e_{1t} & \sum_{t=1}^n e_{mt}e_{2t} & \dots & \sum_{t=1}^n e_{mt}^2 \end{bmatrix}$$

则式 (4) 的矩阵表达式如下所示:

$$J = \mathbf{K}^T \mathbf{E} \mathbf{K} \quad (5)$$

令 $\mathbf{R} = (1, 1, \dots, 1)^T$, 则 $\sum_{i=1}^m k_i = 1$ 可以表示为:

$$\mathbf{R}^T \mathbf{K} = 1 \quad (6)$$

由此可知我们要做的就是式 (6) 的约束条件下, 求出最优加权向量 \mathbf{K} , 使得此时的误差平方和 J 最小, 即:

$$\begin{cases} J_{\min} = \mathbf{K}^T \mathbf{E} \mathbf{K} \\ \text{s.t. } \mathbf{R}^T \mathbf{K} = 1 \end{cases} \quad (7)$$

利用拉格朗日乘子 λ 对式 (7) 进行求解, 则 J 可以表示为^[16]:

$$J = \mathbf{K}^T \mathbf{E} \mathbf{K} + \lambda(\mathbf{R}^T \mathbf{K} - 1) \quad (8)$$

若想求得 J 的极值, 则 J 对 \mathbf{K} 的一阶偏导必为零^[16], 则:

$$\partial J / \partial \mathbf{K} = 2\mathbf{E} \mathbf{K} + \lambda \mathbf{R} = 0 \quad (9)$$

将式 (9) 两边左乘 \mathbf{E}^{-1} 可得:

$$2\mathbf{K} + \lambda \mathbf{E}^{-1} \mathbf{R} = 0 \quad (10)$$

将式 (10) 两边左乘 \mathbf{R}^T 可得:

$$2\mathbf{R}^T \mathbf{K} + \lambda \mathbf{R}^T \mathbf{E}^{-1} \mathbf{R} = 0 \quad (11)$$

将式 (6) 代入式 (11) 即可得到 λ 的值:

$$\lambda = -\frac{2}{(\mathbf{R}^T \mathbf{E}^{-1} \mathbf{R})} \quad (12)$$

将式 (12) 代入式 (10) 即可得到 \mathbf{K} 的值:

$$\mathbf{K} = \frac{\mathbf{E}^{-1} \mathbf{R}}{\mathbf{R}^T \mathbf{E}^{-1} \mathbf{R}} \quad (13)$$

又由式 (6) 可得:

$$\mathbf{R}^T \mathbf{K} = \frac{\mathbf{R}^T \mathbf{E}^{-1} \mathbf{R}}{\mathbf{R}^T \mathbf{E}^{-1} \mathbf{R}} = 1 \quad (14)$$

从而得出结论, 由式 (13) 求得的最优权重满足式 (6) 的约束条件, 即表明式 (13) 求得 \mathbf{K} 就是最优加权系数向量, 且此时的 J 值就是极小值。

$$\begin{cases} J_{\min} = \frac{1}{R^T E^{-1} R} \\ K = \frac{E^{-1} R}{R^T E^{-1} R} \end{cases} \quad (15)$$

最后将求出的最优权重 K 代入式 (1), 即可求出 t 时刻的最优预测值。

2.2 多元线性回归法

多模式集合预报系统的预报结果往往受到多个单项预报成员的影响, 由于各单项预报模式的预报性能不同, 不同模式对 SO_2 的预测准确程度不同, 所以为了分析各单项预报模式之间的相关性, 采用多元线性回归的方法^[7]对各模式的预报结果进行集成预报, 通过大量的样本值做回归分析, 建立拟合度较高的回归方程。

$$\hat{Y}_t = b_0 + b_1 \hat{x}_1 + b_2 \hat{x}_2 + \dots + b_k \hat{x}_k \quad (16)$$

其中, $\hat{Y}_t (t=1, 2, \dots, n)$ 表示 t 时刻的观测值 Y 的估计值, $b_k (k=1, 2, \dots, n)$ 表示各单项预报模式的回归系数, b_0 是常数项, $\hat{x}_k (k=1, 2, \dots, n)$ 表示各单项预报模式的预报值。

2.3 动态权重更新法

动态权重更新^[8]通过评估一段时间内单个预报模式的预测值和实际观测值的偏差情况, 从而可以动态调整各单项预报模式在集成预报中的权重, 其优点在于各单项预报模式的权重分配依赖于实际的监测数据, 具体操作步骤如下:

设第 i 个数值预报模式第 j 天的 SO_2 预测值为 \hat{Y}_{ij} , 第 j 天的实测日数据为 Y_j , 则该模式的偏差率为:

$$R_{ij} = \frac{|\hat{Y}_{ij} - Y_j|}{Y_j} \quad (17)$$

然后根据第 i 个模式每天的偏差率计算出总共 m 天内的平均偏差率 R_{im} 。定义一个过渡因子 V_i , 且该过渡因子与权重系数呈线性关系, 与预测偏差率呈反比, 则 V_i 可表示为:

$$V_i = 1 - \frac{R_{im}}{\sum_{i=1}^n R_{im}} \quad (18)$$

设单个模式的权重为 $w_i (i=1, 2, \dots, n)$, 且满足 $\sum_{i=1}^n w_i = 1$, 则 w_i 可表示为:

$$w_i = \frac{V_i}{\sum_{i=1}^n V_i} \quad (19)$$

最后将求得的权重和相对应的模式预报结果进行加权计算即可得到多模式集成预报结果。

2.4 评估方法

根据统计学的相关性原理, 本文通过使用误差平方和 (Sum of Squares for Error, SSE) 以及均方百分比误差 (Mean Square Percentage Error, $MSPE$) 两个评价指标对各预测模型的准确度进行评估分析。

(1) 误差平方和:

$$SSE = \sum_{t=1}^n (x_t - \hat{x}_t)^2 \quad (20)$$

(2) 均方百分比误差:

$$MSPE = \frac{1}{n} \sqrt{\sum_{t=1}^n [(x_t - \hat{x}_t)/x_t]^2} \quad (21)$$

其中, n 表示实验样本总个数, \hat{x}_t 为第 t 时刻单项预测方法的预测值, x_t 为第 t 时刻的实测值。

SSE 表征的是由于模式的输入、排放清单采集不合理以及模式自身的一些物理和化学参数导致预报结果与实际观测值的偏差。其值越大, 表明与测定值之间的差异越大。 $MSPE$ 则表征各模式的平均误差水平。其值越小, 表明模式的预测值与实际观测值的误差值越小, 即预测的偏差值更小。

3 实验结果

3.1 单项模式预测评估

为了了解单个空气质量模式的特点以及它在不同站点对 SO_2 的预测能力, 本文收集了 2018 年 6 月份云南省蒙自、楚雄、昭通三个站点各单项空气质量预报模式 (CMAQ、CAMx、WRF-CHEM) 的 SO_2 预测值和日均实测值。通过对比不同区域的各单项预报模式预测结果, 利用 SSE 与 $MSPE$ 两项指标评估多模式集合系统中三个空气质量模式对不同站点 SO_2 的预测能力, 结果如表 1 所示。

观察表 1 中数据发现, 在蒙自站点的误差分析中, CMAQ 模式的 SSE 以及 $MSPE$ 在三种模式误差对比中最小, 说明该模式的预测能力最好, CAMx 模式的预测能力次之, WRF-CHEM 模式的预测能力最差。而在楚雄站点的误差分析中, 对 SO_2 预测较为精准的模式是 WRF-CHEM 和 CAMx, CMAQ 模式的预测能力最差。在昭通站点的误差分析中, WRF-CHEM 和 CMAQ 模式的预测能力较为精准, CAMx 的预测能力最差。

由此可知,对于不同的区域,不同的模式预测能力相差较大,每种模式在不同区域的优点各有不同.通过进一步对比不同区域的三种模式的预测结果可得,没有任何一种模式在不同区域的预测结果完全优于其它两个模式,从而在一定程度上证明了单项空气质量模式对大气污染物浓度的预测仍存在一定的局限性.

表 1 各单项模式在不同区域的误差分析

区域	模式	SSE	MSPE(%)
蒙自	WRF-CHEM	514	0.085
	CMAQ	254	0.056
	CAMx	394	0.075
楚雄	WRF-CHEM	548	0.089
	CMAQ	936	0.120
	CAMx	556	0.088
昭通	WRF-CHEM	291	0.050
	CMAQ	322	0.050
	CAMx	431	0.059

3.2 最优定权组合法集成预测评估

为了充分利用各单项空气质量预测模式的优势,验证本文所提出的最优组合预测模型(OWCF)的有效性,选取多元线性回归法(MLR)和动态权重更新法(DWA)与本文提出的最优组合预测法分别对各单项空气质量预报模式(CMAQ、CAMx、WRF-CHEM)的预报结果进行集成对比实验.即在相同的实验条件下,分别利用蒙自、楚雄、昭通三个站点2018年1至5月份的SO₂实测值和上述模式预报数据构建出预测模型,对各站点2018年6月份的SO₂浓度值进行预测,并与各站点六月份的实测数据(REAL)进行对比验证,最后通过绘制对比图的方式将各方法的预测结果与实测结果进行对比分析,实验结果如图3-图11所示.

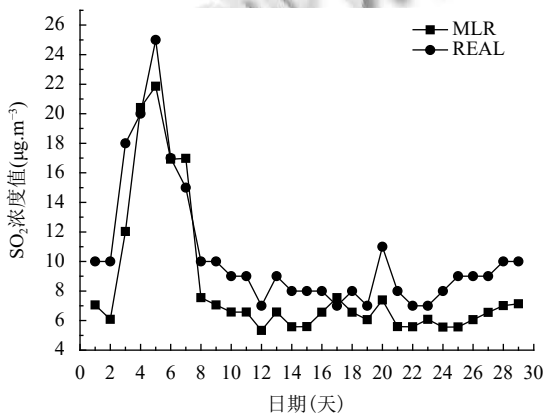


图 3 MLR 预测结果图(蒙自)

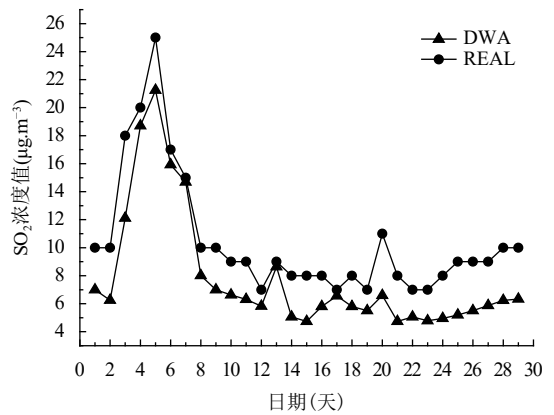


图 4 DWA 预测结果图(蒙自)

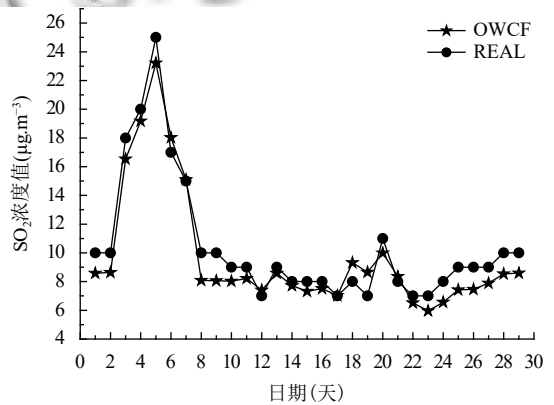


图 5 OWCF 预测结果图(蒙自)

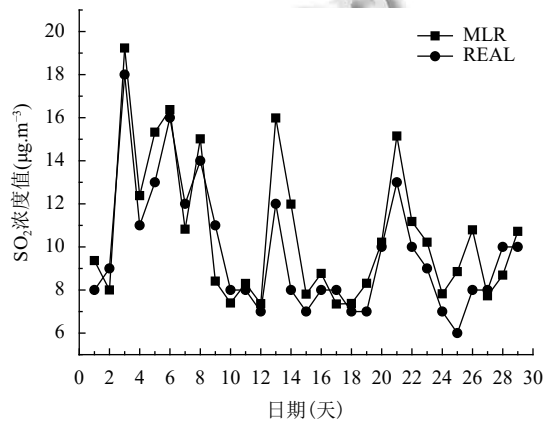


图 6 MLR 预测结果图(楚雄)

由图3-图11中不同站点的预测结果对比曲线可以看出,在大多数时间内,三种方法的预测结果均与实测结果(REAL)的拟合曲线趋势一致.但是相较于本文所提的OWCF,MLR和DWA所预测的结果准确率不高,偏差较大,而OWCF的结果值更加接近于实测值,预测准确率更高.

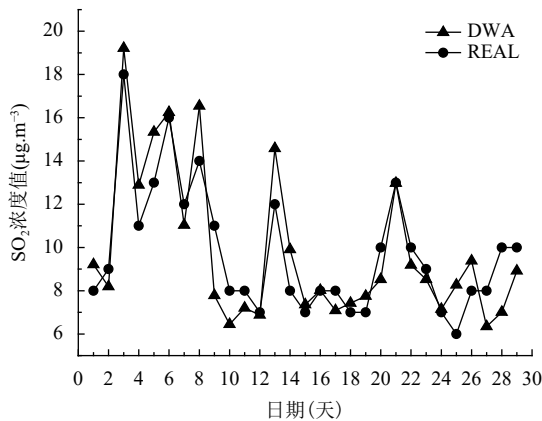


图7 DWA 预测结果图(楚雄)

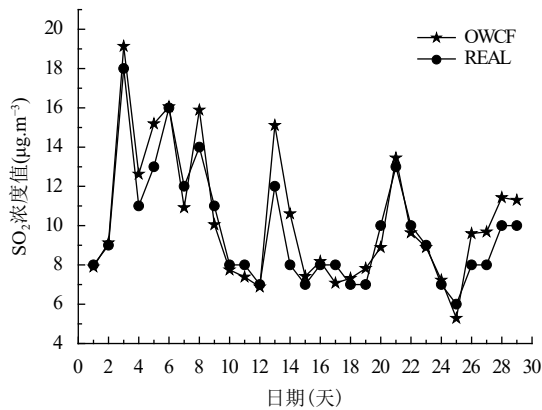


图8 OWCF 预测结果图(楚雄)

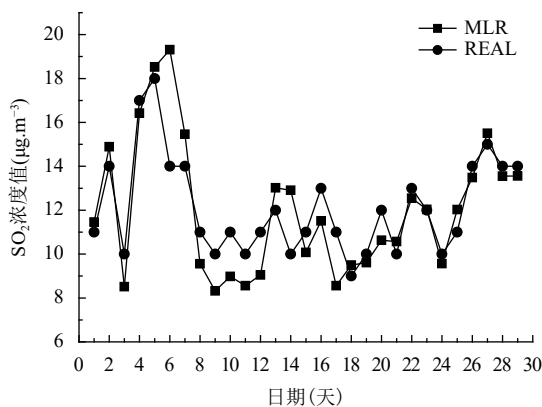


图9 MLR 预测结果图(昭通)

为了定量地分析最优定权组合预测模型的特点, 分别运用 *SSE* 与 *MSPE* 对三种预测方法的预测精度实行评估检验, 其结果如表 2 所示.

表 2 表示通过采用式 (20) 和式 (21) 对三种预测方法在不同区域的预测结果进行评估分析, 由表中数

据可知, 在蒙自站点中, MLR 和 DWA 的 *SSE* 分别为 179.82、218.05, 本文所提的 OWCF 的 *SSE* 与以上两种方法的 *SSE* 相比分别降低了 143.65、181.88. 从 *MSPE* 的角度来看, MLR、DWA 的均方百分比误差分别为 0.047、0.054, OWCF 的 *MSPE* 为 0.022, 与以上两种方法相比分别降低了 0.025、0.032 个百分点. 同样, 在楚雄站点和昭通站点的数据中, 本文所提的 OWCF 预测法的预测精度都存在不同程度的提升, *SSE*、*MSPE* 均最小.

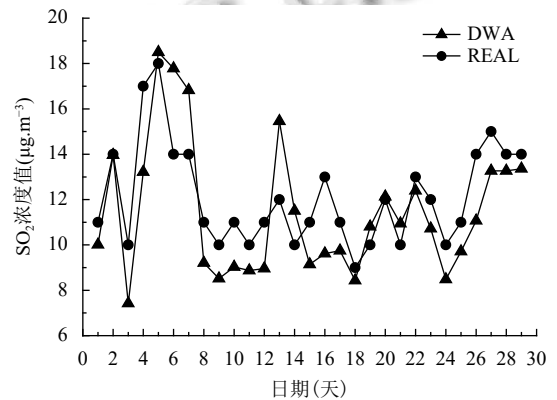


图10 DWA 预测结果图(昭通)

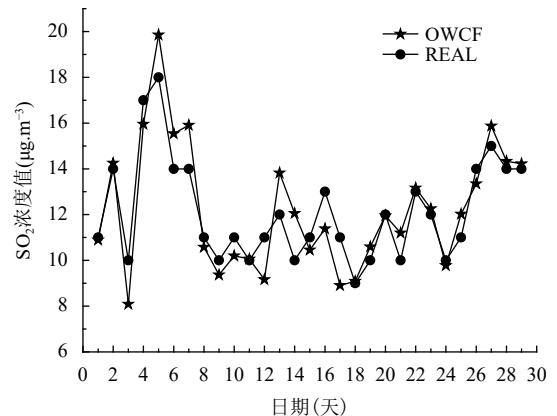


图11 OWCF 预测结果图(昭通)

表 2 三种预测方法在不同区域的误差分析

区域	方法	<i>SSE</i>	<i>MSPE</i> (%)
蒙自	MLR	179.82	0.047
	DWA	218.05	0.054
	OWCF	36.17	0.022
楚雄	MLR	83.29	0.036
	DWA	68.49	0.030
	OWCF	44.39	0.024
昭通	MLR	87.88	0.029
	DWA	132.55	0.034
	OWCF	37.94	0.018

且通过对比表1中单项模式的误差可知,在蒙自站点中,预测能力最好的单项模式是CMAQ,其中SSE为254,MSPE为0.056.而通过采用本文所提的OWCF法将各模式结果进行集成所获得的SSE为36.17,MSPE为0.022.说明本文所提方法的预测精度即使和单项预测模式中预测精度最优的模式相比仍然存在一定的优势.同样,在楚雄和昭通站点中,OWCF和各站点单项最优的预测模式相比,OWCF的SSE和MSPE均最低.

由以上分析可知,通过与单项预测精度最优的模式以及采用不同方式的集成方法预测效果相比较,本文所提的OWCF有效的提升了SO₂的预测准确度,从而证明了本文所提方法的可行性.

4 结语

本文基于多个空气质量预测模式(WRF-CHEM、CMAQ、CAMx),以误差平方和最小为原则,提出了最优定权组合预测模型.采用2018年1至5月份云南省三个监测站点(蒙自、楚雄、昭通)的SO₂浓度实测值和多个单项空气质量模式的预测值作为实验数据,与现有的多元线性回归法、动态权重分配法等预测方法对前述三个站点2018年6月份的SO₂浓度进行预测,并对各模型的预测误差进行评估分析.实验结果表明最优定权组合预测模型的预测准确率最高,效果最佳,为云南省空气质量预警预报平台提供一种高效的预测方法.

参考文献

- 李锐,李俊伟,刘之杰,等.卫星遥感研究中国气溶胶光学厚度、NO₂和SO₂的相关性.科学通报,2016,61(22): 2524-2535.
- 王梓廷,王艳,宋洪,等.SO₂及酸雨胁迫对不同时期冬小麦叶绿素含量的影响.西北农业学报,2012,21(4): 131-136. [doi: 10.3969/j.issn.1004-1389.2012.04.026]
- 李勇,白云,李川.大气污染物SO₂预测模型研究综述.四川环境,2016,35(1): 144-148. [doi: 10.3969/j.issn.1001-3644.2016.01.027]
- 杜钧.集合预报的现状和前景.应用气象学报,2002,13(1): 16-28. [doi: 10.3969/j.issn.1001-7313.2002.01.002]
- 王自发,吴其重, Gbaguidi A, 等.北京空气质量多模式集成预报系统的建立及初步应用.南京信息工程大学学报:自然科学版,2009,1(1): 19-26.
- Mallet V, Sportisse B. Ensemble-based air quality forecasts: A multimodel approach applied to ozone. Journal of Geophysical Research: Atmospheres, 2006, 111(D18): D18302. [doi: 10.1029/2005JD006675]
- 黄思,唐晓,徐文帅,等.利用多模式集合和多元线性回归改进北京PM10预报.环境科学学报,2015,35(1): 56-64.
- 姚文强,王兆青,铁治欣,等.多模式空气质量集成预报模型的研究.浙江理工大学学报(自然科学版),2017,37(3): 444-450.
- 王永刚,郑红运.基于最优变权组合模型的航空运输事故征候预测.中国安全科学学报,2013,23(4): 26-31.
- 张春萍,殷莹.基于最优定权组合模型的物流量预测.物流工程与管理,2013,35(1): 85-88,91. [doi: 10.3969/j.issn.1674-4993.2013.01.032]
- Grell GA, Peckham SE, Schmitz R, et al. Fully coupled "online" chemistry within the WRF model. Atmospheric Environment, 2005, 39(37): 6957-6975. [doi: 10.1016/j.atmosenv.2005.04.027]
- Dennis RL, Byun DW, Novak JH, et al. The next generation of integrated air quality modeling: EPA's models-3. Atmospheric Environment, 1996, 30(12): 1925-1938. [doi: 10.1016/1352-2310(95)00174-3]
- Nopmongcol U, Koo B, Tai E, et al. Modeling Europe with CAMx for the air quality model evaluation international initiative (AQMEII). Atmospheric Environment, 2012, 53: 177-185. [doi: 10.1016/j.atmosenv.2011.11.023]
- Bates JM, Granger CWJ. The combination of forecasts. Journal of the Operational Research Society, 1969, 20(4): 451-468. [doi: 10.1057/jors.1969.103]
- 唐小我.最优组合预测方法及其应用.数理统计与管理,1992,11(1): 31-35.
- 唐小我.组合预测计算方法研究.预测,1991,(4): 35-39.