

多模式空气质量集成预报模型的研究

姚文强¹,王兆青¹,铁治欣¹,丁成富²,唐 敏¹

(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)

摘 要: 针对目前区域空气质量预报出现的预测不准确、误差较大、漏报重污染天气等问题,采用 WRF-Chem、WRF-CMAQ、MM5-CMAQ 和 MM5-CAMx 4 种数值模式集成预报的方式,对区域空气质量状况及污染物浓度进行预测预报,提出了一个动态权重更新模型,该模型通过评估一段时间内与特定条件下单个模式的预测准确度,动态调整各个模式在集成预报模型中的权重因子,然后利用各个模式预测值及其权重取加权平均值作为新模式的预测值。以收集到的浙江省空气质量监测网中的 47 个国家站点的 8 个月的监测数据为测试数据,分别用已有模型和动态权重更新模型进行预测预报,并对预测预报结果进行评估分析,结果表明:动态权重分配模型预报结果具有更高的准确度与更小的误差。

关键词: 空气质量;多模式;数值预报;动态权重更新

中图分类号: TP311.5

文献标志码: A

文章编号: 1673-3851 (2017) 03-0444-07

0 引 言

近年来,随着污染问题的日益突出,空气质量状况已经成为公众越来越关心的话题,空气质量预测模型也成为科学研究的热点。我国的大气污染问题比较复杂,特别是最近几年,已经由过去单一类型的污染逐步转变为多种污染物、多种作用机制同时存在、相互影响的复合型大气污染问题。复杂的污染问题必然会增加大气质量预测的难度,传统的以单一类型污染为研究对象的预测方法已无法全面揭示当前污染的特征和演变规律。所以改进优化空气质量预测方式,研究准确有效的预测模型尤为重要。

现有的空气质量预测预报方法主要分为统计预报与数值预报。统计预报是根据统计学原理,用数理统计的方法对空气质量的变化趋势进行预测。数值预报是根据大气实际情况,在一定的初值条件下,运用计算机作数值计算,预测未来一定时间大气污染物的浓度与变化趋势。数值预报起

源于国外,迄今已经发展到第三代^[1]。我国有关大气数值预报的研究起步较晚,但最新一代的空气质量数值预报模式已经在国内空气质量预报及污染物预报领域得到了广泛的应用^[2-7]。多模式集成预报日益成为数值预报的发展趋势,国内学者运用基于多模式的空气质量集成预报技术,实现对当前大气复合污染的预测预报。王自发等^[8]以 NAQPMS 模式、CMAQ 模式^[9]及 CAMx 模式为核心,采用统一的模式网格、统一的排放源、统一的气象场,构建了 EMS-Beijing 空气质量多模式集成预报系统,对于预测北京及周边污染物起到了重要作用。王茜等^[10]为上海市建立了包含 NAQPMS、CMAQ、CAMx、WRF-Chem 等多个国内外著名数值模型的集合业务预报模式系统,为上海市提供了丰富的气象和污染预报产品,经过评估后发现集合预报系统对于 PM₁₀、SO₂ 和 NO₂ 日均浓度的预测值与实测值相关系数达 0.5~0.6,并且在 2010 年上海世博会开幕前夕及 2010 年 5 月 19 日两次污染事件中,该集成预报系统为

收稿日期:2016-08-15 网络出版日期:2017-01-19

基金项目:浙江省公益技术应用研究项目(2014C31G2060072)

作者简介:姚文强(1991-),男,山西省运城市人,硕士研究生,主要从事环保软件开发方面的研究。

通信作者:王兆青,E-mail:zilggjb@163.com

上海市环保局启动世博预警联动方案提供了重要的决策依据。谢涛等^[11]将地理信息系统(GIS)技术与多模式集合预报系统结合的方式,在保障预测准确度的前提下,充分发挥GIS在数据管理、信息可视化、辅助决策支持等方面的优点,为广州市在环境质量预测方面提供了保障。黄思等^[12]运用多模式集合预报与多元线性回归的方法,将数值预报与统计预报的技术相结合使用,对北京地区的PM₁₀预报结果进行修正,并且大幅提高了对污染过程的预报能力。

本文通过研究国内外现有较好的数值预报模式,并且结合浙江省的气候、地形、污染源排放、大气污染等方面的特点,建立起包括WRF-Chem、WRF-CMAQ、MM5-CMAQ^[13]和MM5-CAMx 4套模式组合在内的多模式空气质量集成预报系统,并设计了一个基于预测准确度的动态权重更新模型,通过评估单个模式在一段时间与特定条件下的预测准确度,来动态调整各个模式在集成

预报过程中的权重因子,以提高集成预报的准确度。

1 多模式预报系统的总体框架设计

1.1 系统框架

浙江省大气质量多模式预报系统运行在统一的计算平台和运行环境中,以WRF/CHEM、Models-3/CMAQ、CAMx为核心模式构成WRF-Chem、WRF-CMAQ、MM5-CMAQ、MM5-CAMx预报组合,并基于4套模式组合的预报结果建立集成预报技术,整合不同模式的预报性能优势来实现对污染过程影响的分析和污染来源的量化。

在区域空气质量模式本地化的基础上,采用模块化设计、编程技术和网络技术实现全球预报场的自动下载、初始数据的自动预处理、空气质量模式的自动运行、预报的结果的后处理及产品的自动生成等功能。多模式预报系统的构成框架如图1所示。

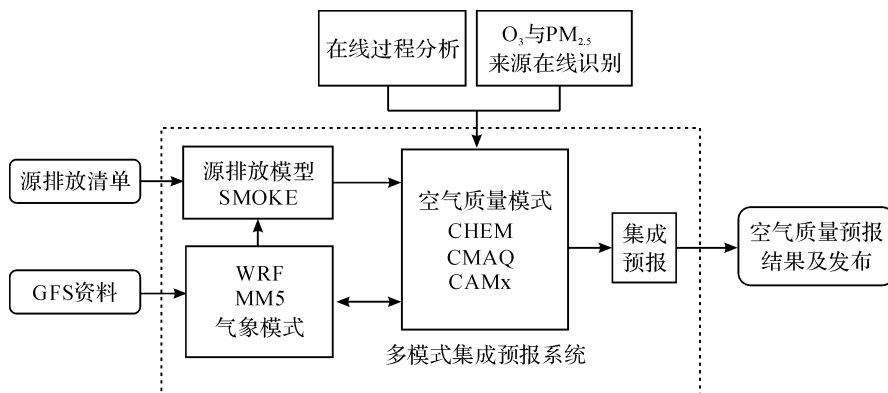


图1 浙江省多模式预报系统框架

1.2 模式模拟网格设置

水平模拟网格采用Lambert Conformal投影,投影参数为北纬25°、北纬47°以及东经108°55';坐标系原点取中国大地坐标原点东经108°55'、北纬34°32'。预报系统采用36、12、4 km三重嵌套网格:36 km网格覆盖中国大部分地区,为更高分辨率的嵌套网格提供气象和空气质量的边界条件;12 km网格覆盖浙江省和上海市全境、江苏省和安徽省绝大部分地区以及江西省东北部和福建省北部,实现浙江省及周边地区的空气质量预报,能够以较高的空间分辨率反映浙江省及周边地区的污染态势,并可用于相互影响分析;4 km网格覆盖浙江省全境,提供浙江省各城市更高分辨率的空气质量预报结果。设置的网格区域如图2所示。垂直模拟网格分为16层,最底层高度距离地面约40 m。

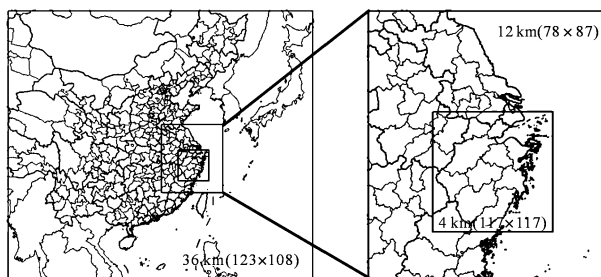


图2 浙江省多模式预报系统嵌套网格设置

1.3 输入数据

气象模式最外层的气象初始和边界条件采用全球预报系统GFS每天的预报结果,系统设定每天北京时间12:30开始自动下载全球预报模式在世界时当天0时的预报结果。最外层空气质量模式的边界条件采用全球模式输出的污染物浓度分布廓线,初始场采用预报系统在前一天的预报结果。

对于源排放输入数据,浙江省范围内采用本项目研制的高分辨率污染源排放清单的处理结果,浙江省以外的地区采用包括中国在内的东亚地区近年的排放清单资料作为输入。

1.4 运行环境

预报系统的运行环境的操作系统为 Red Hat Enterprise Linux 5, 安装有 GNU C/C++ 编译器、GNU Fortran77 编译器、Intel C/C++ 编译器、Intel Fortran 77/90 编译器以及 MPICH2 并行计算环境。

2 预报性能评估

2.1 评估方法

根据统计学中关于评估实测与预测相关性原理,本文通过特定的预报评估模块对预报系统预报结果的准确性进行评价,评估模块采用标准化平均偏差(NMB)、标准化平均误差(MME)、相关系数(COR)来评估污染物预测结果与实测值的吻合程度。计算公式如下:

a) 标准化平均偏差:

$$NMB = \frac{\sum_{i=1}^N (M_i - O_i)}{\sum_{i=1}^N O_i} \times 100\% \quad (1)$$

b) 标准化平均误差:

$$MMB = \frac{\sum_{i=1}^N |M_i - O_i|}{\sum_{i=1}^N O_i} \times 100\% \quad (2)$$

c) 相关系数:

$$COR = \frac{\sum_{i=1}^N (M_i - M_a)(O_i - O_a)}{\sqrt{\sum_{i=1}^N (M_i - M_a)^2 \sum_{i=1}^N (O_i - O_a)^2}} \quad (3)$$

其中: N 代表评估的总月份数, M_i 代表第*i*个月模式的预测值, M_a 代表总月份的模式预测平均值, O_i 为第*i*个月的实际测量值, O_a 代表总月份的实际测量值的平均值。标准化平均偏差 NMB 可以反应模式预测的正负相关性大小,正数代表总体预测较高于实际测量值,负数则相反,标准化平均偏差 NME 可以反应模式预测的偏差大小,而相关系数 COR 可以直接反应模式预测的准确度,相关系数大,则表示模式预测值的吻合程度比较高。

2.2 污染物浓度的预报评估

收集了浙江省空气质量监测网中 47 个国家站点 2014 年 1 月 1 日至 8 月 31 日期间每天各类污染物监测数据及各个预测模型的预测值,通过式(1) — (3) 计算各预测模型对各类污染物的预报效果统计评估参数,如表 1 所示,其中表 1 中“平均集成”模型各类污染物数值是由 4 种模型相对应的污染物数值求平均值后,通过式(1) — 式(3) 计算平均集成模型的评估参数。

表 1 各个模型对不同污染物的预测评估参数指标

模型系统	评估指标	SO ₂	NO ₂	O ₃	PM ₁₀	PM _{2.5}	CO
MM5-CAMx	NMB/%	35	-13	-35	-64	-54	-53
	NME/%	80	48	44	66	58	61
	COR	0.15	0.13	0.50	0.28	0.39	0.13
MM5-CMAQ	NMB/%	-12	-11	-24	-60	-55	-57
	NME/%	63	48	36	62	58	63
	COR	0.25	0.16	0.61	0.31	0.41	0.12
WRF-Chem	NMB/%	42	32	-42	-3	-8	-2
	NME/%	82	55	52	50	46	40
	COR	0.32	0.18	0.45	0.32	0.42	0.37
WRF-CMAQ	NMB/%	34	10	19	-10	-17	0
	NME/%	67	41	33	38	35	37
	COR	0.51	0.48	0.72	0.62	0.68	0.45
平均集成	NMB/%	19	5	-22	-32	-31	-24
	NME/%	40	27	25	38	36	33
	COR	0.46	0.43	0.83	0.63	0.72	0.60

由表 1 可知,从单个模式预测各个污染物角度出发,参考各污染物的预报评估参数可以看出,不同模式系统在预报过程中对不同污染物的预报效果存在一定差异。对 SO₂ 的预测表现最佳的模型系统为 MM5-CMAQ 和 WRF-CMAQ,MM5-CMAQ 模型系统对 SO₂ 预测的偏差及误差为最低,虽然 WRF-CMAQ 预测的偏差及误差大于 MM5-CMAQ,但是一致性系数优于 MM5-CMAQ;WFR-CMAQ 预测 NO₂ 的偏差及误差均为最低,同时一致性也最优,因此其对 NO₂ 的预测效果最佳,MM5-CMAQ 与 MM5-CAMx 对 NO₂ 的预测偏差及误差略高于 WRF-CMAQ,对 NO₂ 预测的效果较优;综合各项评估参数,对 O₃ 预测最佳的模型系统为 MM5-CMAQ 与 WRF-CMAQ,这两套模型对 O₃ 的预测偏差和误差均较低并且一致性系数较高,其中 MM5-CMAQ 的预测结果偏低而 WRF-Chem 的预测结果偏高;对 CO 的预测表现较好的模型系统为 WRF-Chem 和 WRF-CMAQ,MM5-CMAQ 与 MM5-CAMx 对 CO 的预测偏低显著,可能由于排放清单的低估导致。各套模型系统对颗粒物的预测均存在偏低现象,其中 WRF-Chem 的偏低水平最不显著,但是其预测误差高于 WRF-CMAQ,综合来看 WRF-CMAQ 对颗粒物的预测效果最佳。

对表 1 中通过平均集成后获得的评估参数与其他单个模式进行比较,发现就浙江全省的平均预测效果而言,“平均集成”模型的预测误差优于任何一

个单模式的预测结果,对 SO₂、NO₂ 和 O₃ 的预测效果提升较为明显,优于或至少相当于任一单一模式的预测结果。

2.3 空气质量指数(AQI)与空气质量等级的预测评估

空气质量指数(air quality index,AQI)是定量描述空气质量状况的无量纲指数。AQI 值可以直观反映出空气质量的等级状况,所以我们运用特定的评估参数对多模式集成预报系统的 AQI 预报进行评估。

表 2 给出了各套模型系统及平均集成模型的 AQI 预报效果的统计评估结果。观察评估参数后发现各“平均集成”模型的结果相对单个模式预报误差有所减少,与观测值的相关性有了明显提高。

表 2 日均 AQI 预测效果评估结果

评估参数	平均集成	MM5-CAMx	MM5-CMAQ	WRF-Chem	WRF-CMAQ
NMB/%	-18	-42	-48	-9	-10
NME/%	25	45	50	29	27
COR	0.73	0.38	0.24	0.46	0.64

对每日预报的污染等级与实际观测等级进行比较和逐月统计,表 3 是 2014 年 1 月至 8 月各个模式与“平均集成”模型预报的污染等级预报准确率的逐月统计,比较后发现就浙江省整体预报的情况而言,“平均集成”模型的污染等级预报结果准确率方面有了显著提高。

表 3 2014 年 1—8 月污染等级预测准确率逐月统计

模式	指标	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	平均值
MM5-CAMx	准确	50	42	52	53	43	32	40	39	43.8
	偏高	22	25	18	32	17	33	30	40	26.1
	偏低	28	33	30	15	50	35	30	21	30.1
MM5-CMAQ	准确	47	51	63	60	43	59	70	51	55.5
	偏高	23	25	7	20	26	24	20	39	23.0
	偏低	30	24	30	20	31	17	10	10	21.5
WRF-Chem	准确	55	54	67	60	54	55	60	51	57.0
	偏高	15	22	10	20	26	24	20	19	19.5
	偏低	30	24	23	20	20	21	20	30	23.5
WRF-CMAQ	准确	57	61	73	70	53	69	80	61	65.5
	偏高	3	25	0	20	16	24	20	39	18.4
	偏低	40	14	7	10	31	7	0	0	13.6
平均集成	准确	58	68	83	73	64	69	80	81	72.0
	偏高	0	18	0	0	6	7	10	19	7.5
	偏低	42	14	17	27	30	24	10	0	20.5

3 多模式动态权重分配模型

近几十年来,空气质量模型有了快速发展,虽然在模式结构、数值算法和物理化学过程等方面都有很大改进,大气污染模式对主要的污染过程处理还存在很多不确定性。这主要在于不同模式因开发时所关注的主要科学问题不同,重点考虑的大气物理、化学过程有所侧重,而简化其他一些过程,这使得即使是评估中最优的模式也存在模拟、预报上的某些缺陷,单一模式并不能完全模拟所有大气物理、化学过程,且针对不同地域及各天气过程各模式预报效果差异显著。这种不确定性的存在,使得目前空气质量模式在空气质量预报和空气质量管理政策制定上存在一定的风险。而多模式集合预报正成为改进预报和模拟的一种重要手段,通过对不同模式预报结果,采用合理的集成方法能有效改进模式的整体模拟效果。

基于以上评估分析,发现多模式空气质量预报取算术平均值的引入,相对单个模式而言提高了浙江省空气质量预测的准确度,但是数值预报模式有不同的输入条件与计算规则,模式在预报过程中的不同会得到不同的预报结果,通常情况下单个模式不可能完全准确地预报出空气质量状况,往往会有偏差,但是单个模式在不同条件下会呈现出不同的预测准确度,如上文分析得到 WRF-CMAQ 在浙江省空气质量预测中具有较高的准确度,只是简单取各个模式的算术平均值,往往不能发挥单个模式在特定条件下具有较高准确度的优势,于是本文提出并且设计了一个动态权重分配模型,动态权重分配的优势在于可以通过分析各个模式在一段时间与特定条件下的预测准确度,来分配集成预报的权重因子。基于以上思路,系统通过分析集成预报与单个模式的准确度来动态调整权重因子。

3.1 模型构建

动态权重分配模型的数学模型如下:

定义单个模式的权重 W_i , 各个模式的权重因子相加等于 1, 那么 W_i 应该满足:

$$\sum_{i=1}^n W_i = 1 \quad (4)$$

其中 n 表示参与集成预报的模式数量。

偏差率是预测值与实际测量值做减法运算后求绝对值与实际测量值的除法运算结果。偏差率可以反应出单个模式预测的误差大小。设第 i 个预报模

式第 j 天预测值 M_{ij} , 实际观测值 Q_j , 则该 i 模式第 j 天的偏差率 R_{ij} 可用下式计算:

$$R_{ij} = \frac{|M_{ij} - Q_j|}{Q_j} \quad (5)$$

根据模式每天的预测偏差率, 做算术平均值求单个模式一段时间内的预测偏差率。定义第 i 个模式一段时间内, 总共 m 天内的偏差率 R_{im} 为:

$$R_{im} = \frac{\sum_{j=1}^m R_{ij}}{m} \quad (6)$$

过渡因子是该模型定义的一个变量, 定义过渡因子与权重因子呈线性关系并且与预测偏差率成反比。设第 i 个模式的权重过度因子 V_i , 可用下式计算:

$$V_i = 1 - \frac{R_{im}}{\sum_{i=1}^n R_{im}} \quad (7)$$

因为过渡因子与权重因子呈线性关系, 并且满足各个模式权重因子和为 1。第 i 个模式预报的权重因子 W_i 可用下式计算:

$$W_i = \frac{V_i}{\sum_{i=1}^n V_i} \quad (8)$$

定义 M_i 为第 i 个模式的预测结果, F_m 为最后的集成结果, 求解公式为:

$$F_m = \sum_{i=1}^n M_i W_i \quad (9)$$

其中 n 表示参与集成预报的模式数, 集成结果通过每个模型预测结果 M_i 乘以对应的权重因子 W_i 之后相加得到。

3.2 模型评估

将动态权重分配模型运用到多模式空气质量集成预报系统中, 并对浙江省空气质量监测网中 47 个站点 2014 年 1 月 1 日至 8 月 31 日的 AQI 数据进行评估分析, 用动态权重分配模型与平均集成计算出来的 AQI 预测结果的评估参数指标对比结果列于表 4。经过动态权重分配模型处理后的 AQI 的逐日变化情况与观测结果的时间序列对比结果见图 3。

表 4 动态权重分配模型处理后的预测结果的评估参数

集成方法	NMB/%	NME/%	COR
动态权重分配模型	-11	22	0.72
平均集成	-18	25	0.67

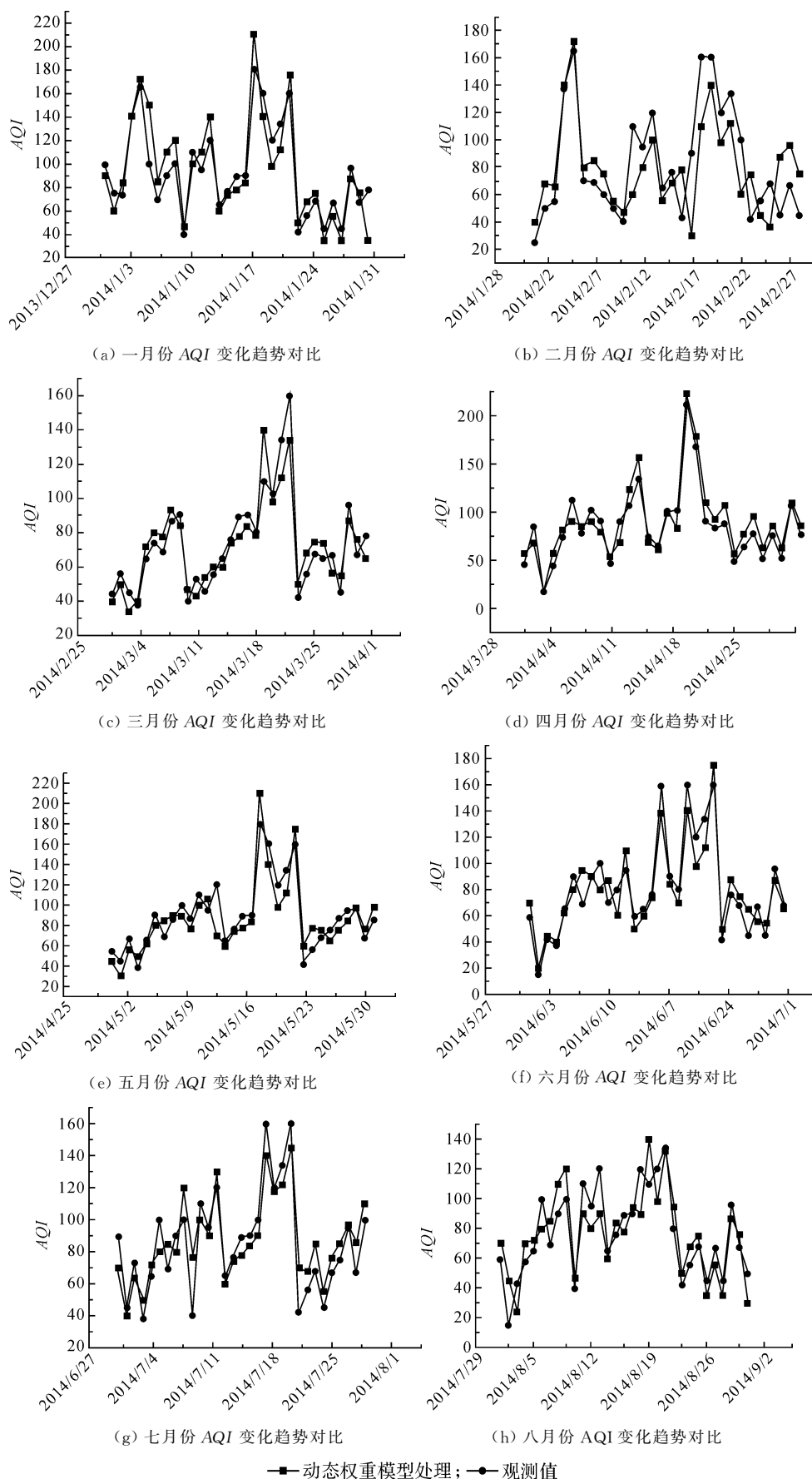


图 3 动态权重分配模型处理后的 AQI 预报结果与观测结果的时间序列对比

由表4可知,将动态权重分配模型与平均集成模型预报结果的评估参数指标进行比较,发现动态权重分配模型的标准化平均偏差(NMB)变小,虽然在整体预报方面还存在偏小的趋势,但相对算术平均值有了一定的提高,标准化平均误差(NME)变小,说明预测误差有所减小,而相关系数(COR)优化后的结果相比平均集成的结果大,说明在预测准确度上有明显的提高。

图3的观测值变化趋势与预测变化趋势,发现动态权重分配模型的预测值与实际观测值具有较好的拟合性与较相同的变化趋势。综上所述可以看出,说明动态权重分配模型预测准确度有所提高,比平均集成模型要好。

4 结 语

本文根据空气质量数值预报技术,建立了多模式集成预报模型,以WRF-Chem、WRF-CMAQ、MM5-CMAQ和MM5-CAMx 4种数值预报模式为基础,提出了一个动态权重更新模型,对区域空气质量状况及污染物浓度进行预测预报。以浙江省47个国家站点8个月的监测数据为测试数据,分别用WRF-Chem、WRF-CMAQ、MM5-CMAQ、MM5-CAMx、平均集成模型及动态权重更新模型对浙江省空气质量状况及污染物浓度进行预测预报,并对它们的预报结果进行分析。实验结果表明,动态权重更新模型具有较高的准确度与较小的误差,可以准确预测空气质量状况与污染物浓度分布情况,为浙江省的区域空气质量预测预报业务提供了有力的技术支持与保障。

参考文献:

[1] 谭成好,陈昕,赵天良,等. 空气质量数值模型的构建及应用研究进展[J]. 环境监控与预警,2014(6):1-7.

- [2] 陈彬彬,林长城. 基于CMAQ模式产品的福州市空气质量预报系统[J]. 中国环境科学,2012,32(10):1744-1752.
- [3] 程兴宏,徐祥德,丁国安. CMAQ模式卫星产品源同化模型及其在空气质量预报中的应用研究[J]. 中国科学:地球科学,2010,40(4):511-522.
- [4] 陈焕盛,王自发,吴其重,等. 亚运时段广州大气污染物来源数值模拟研究[J]. 环境科学学报,2010,30(11):2145-2153.
- [5] 黄蕊珠,陈焕盛,葛宝珠,等. 京津冀重霾期间PM_{2.5}来源数值模拟研究[J]. 环境科学学报,2015,35(9):2670-2680.
- [6] CHENG N, HONGXIA L I, FAN M, et al. Overview of urban PM_{2.5} numerical forecast models in China[J]. Asian Agricultural Research, 2015, 7(10):47-53.
- [7] 王自发,李杰,王哲,等. 2013年1月我国中东部强霾污染的数值模拟和防控对策[J]. 中国科学:地球科学,2014,44(11):3-14.
- [8] 王自发,吴其重,晏平仲,等. 北京空气质量多模式集成预报系统的建立及初步应用[J]. 南京信息工程大学学报(自然科学版),2009(3):19-26.
- [9] 王占山,李晓倩,王宗爽,等. 空气质量模型CMAQ的国内外研究现状[J]. 环境科学与技术,2013,36(s):386-391.
- [10] 王茜,伏晴艳,王自发,等. 集合数值预报系统在上海市空气质量预测预报中的应用研究[J]. 环境监控与预警,2010,2(4):1-6.
- [11] 谢涛,姚新,孙世友. 基于GIS的空气质量多模式预报系统[J]. 地理信息世界,2011,9(1):11-15.
- [12] 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[J]. 环境科学学报,2015,35(1):56-64.
- [13] SHRESTHA K L, KONDO A, AKIKAZU K, et al. High-resolution modeling and evaluation of ozone air quality of Osaka using MM5-CMAQ system [J]. Journal of Environmental Sciences, 2009, 21(6):782-789.

Research on Multi-Model Air Quality Integrated Forecasting Model

YAO Wenqiang¹, WANG Zhaoqing¹, TIE Zhixin¹, DING Chengfu², TANG Min¹

(1. School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Focused Photonics (Hangzhou) Inc., Hangzhou 310052, China)

Abstract: Aiming at the problems existing in the field of air quality forecast, such as inaccurate prediction, larger error and missed to report the heavy pollution weather. In this paper, we used the technology named numerical forecast in the field of air quality forecast, and proposed to use multi-model ensemble forecasting method to forecast the regional air quality and pollutant concentration. The experiments used WRF-Chem, WRF-CMAQ, MM5-CMAQ and MM5-CAMx to do integration. And the dynamic weight updating model is proposed, according to the prediction accuracy of these four kinds of numerical model in a period of time and under certain conditions, weights are dynamically assigned to these models' prediction value in the new model. The predictive value of the new model is obtained by the weighted average of the predictive value of these four existing prediction model. And the model was used to forecast the air quality of Zhejiang province, eight months for monitoring data of 47 national stations which are of air quality monitoring network in this province was collected. And the forecast results obtained in the multi-model forecasting system were evaluated and analyzed by parameters specified in paper. The experiment show that the Dynamic Weight updating Model has higher prediction accuracy and less error.

Key words: air quality; multi-model; numerical forecast; dynamic weight updating

(责任编辑: 陈和榜)