姚雪峰，葛宝珠，王自发，等. 2018. 改进的超级集成预报方法在长江三角洲地区 O_3 预报中的应用 [J]. 大气科学, 42 (6): 1273–1285. Yao Xufeng, Ge Baozhu, Wang Zifa, et al. 2018. Application of improved super ensemble forecast method for O_3 and its performance evaluations over the Yangtze River Delta region [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 42 (6): 1273–1285, doi:10.3878/j.issn.1006-9895.1801.17167.

改进的超级集成预报方法在长江三角洲地区 O_3 预报中的应用

姚雪峰[1,2] 葛宝珠[1] 王自发[1,3,4] 范凡[1,5] 汤莉莉[6] 郝建奇[1,4]
张祥志[6] 晏平仲[1] 张稳定[1] 吴剑斌[7]

1 中国科学院大气物理研究所大气边界层物理与大气化学国家重点试验室，北京 100029
2 解放军 96631 部队，北京 102208
3 中国科学院城市环境研究所城市环境科学卓越中心，厦门 361021
4 中国科学院大学，北京 100049
5 南京信息工程大学，南京 210044
6 江苏省环境监测中心，南京 210036
7 上海市气象局，上海 200030

摘 要 针对当前单模式系统臭氧（O_3）预报的不确定性问题，提出了一种基于活动区间的多模式超级集成的、高效的预报方法。本研究基于长江三角洲（长三角）地区多模式空气质量预报系统，将改进后的超级集成预报方法（AR-SUP）运用到 2015 年长三角地区的 O_3 预报中，并与滑动训练期的超级集成预报（R-SUP）、多模式集成平均预报（EMN）、消除偏差的集成平均预报（BREM）对比，结果表明 AR-SUP 对预报效果的改善最明显，其在暖季和冷季的均方根误差（RMSE）较最优单模式平均下降了 20%和 23%。将 AR-SUP 运用到 48 h 和 72 h 预报中发现，当预报时效增加时该方法依旧保持较高的预报技巧。多项统计数据均证明 AR-SUP 在研究时段内所有站点均能显著减小 O_3 预报误差、提高整体相关性和一致性，有效提高当前短期（三天）预报准确率。

关键词 臭氧 多模式系统 超级集成预报 活动区间
文章编号 1006-9895(2018)06-1273-13 **中图分类号** P456.8 **文献标识码** A
doi:10.3878/j.issn.1006-9895.1801.17167

Application of Improved Super Ensemble Forecast Method for O_3 and Its Performance Evaluation over the Yangtze River Delta Region

YAO Xuefeng[1,2], GE Baozhu[1], WANG Zifa[1,3,4], FAN Fan[1,5], TANG Lili[6], HAO Jianqi[1,4], ZHANG Xiangzhi[6], YAN Pingzhong[1], ZHANG Wending[1], and WU Jianbin[7]

1 *State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences*, *Beijing* 100029
2 96631 *Army*, *People's Liberation Army of China*, *Beijing* 102208
3 *Center for Excellence in Urban Atomsprheric Environment*, *Institute of Urban Environment*, *Chinese Academy of Sciences*, *Xiamen* 361021

收稿日期 2017-05-12；**网络预出版日期** 2018-02-02
作者简介 姚雪峰，女，1986 年出生，博士研究生，主要研究方向为大气污染预报与空气质量区划。E-mail: yaoxuefeng0818@163.com
通讯作者 葛宝珠，E-mail: gebz@mail.iap.ac.cn
资助项目 国家自然科学基金项目41305113、41575123、41620104008、41611540340、91744206，国家科技支撑计划项目2014BAC22B04，中国科学院先导项目XDA19040204，中国科学院重点部署项目ZDRW-CN-2018-1-03
Funded by National Natural Science Foundation of China (Grants 41305115, 41575123, 41620104008, 41611540340, 91744206), National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant 2014BAC22B04), Strategic Priority Research Program (A) of the Chinese Academy of Sciences (Grant XDA19040204), Key Planed Project of the Chinese Academy of Sciences (Grant ZDRW-CN-2018-1-03)

4 *University of Chinese Academy of Sciences*, *Beijing* 100049
5 *Nanjing University of Information Science and Technology*, *Nanjing* 210044
6 *Jiangsu Environmental Monitoring Center*, *Nanjing* 210036
7 *Shanghai Meteorological Service*, *Shanghai* 200030

Abstract Aiming at existing problems in current O_3 single model forecast, an efficient superensemble forecast based on running active range (AR-SUP) is proposed and applied to the EMS-YRD (multi-model ensemble air quality forecast system for the Yangtze River Delta) O_3 forecast during the study period in 2015. The performance of the newly proposed method is compared with those of R-SUP (Running Training Period Superensemble), EMN (Ensemble Mean), and BREM (Bias-Removed Ensemble Mean). The results show that compared with the other three ensemble methods, the AR-SUP exhibits significant improvement in daily O_3 forecast with the RMSE reduced by 20% and 23% from that of the best single model in cool and warm seasons respectively. Further application of the AR-SUP in O_3 ensemble forecast also shows high forecasting skills when the predicting time is extended to 48 h and 72 h. A number of statistical measures (i.e., reduced errors, increased correlation coefficients, and index of agreement) show that the forecasting skill has been improved at all the locations within the study region during all seasons, which indicates this method can be used to help improve the accuracy and reliability of short-term forecasts.

Keywords O_3, Ensemble air quality multi-model forecast system, Superensemble forecast, Running active range

1 引言

随着快速的能源消耗、城市群发展和人口增长，近年来以高浓度臭氧为典型特征的光化学污染在长江三角洲（长三角）地区日益严重（Gao et al.，2011）。研究表明，高浓度臭氧严重影响人体健康、危害生态环境（Zhang et al.，2006）。向公众特别是敏感人群提供准确的臭氧污染的预报和预警服务很有必要（唐晓等，2013）。中国科学院大气物理研究所以其自主研发的嵌套网格空气质量预报系统（NAQPMS）模式为核心、综合集成国际上先进的 CAMx、CMAQ 和 WRF-Chem 等空气质量模式，建立了长三角地区区域空气质量多模式预报系统。该系统在江苏省环境监测中心已投入运行，能够提供长三角地区五天以内的逐小时精细化臭氧预报。

由于数值解法、排放清单、边界条件以及对大气中的物理化学过程的有限认知，臭氧的数值预报不可避免的存在误差和不确定性（Kang et al., 2010，唐晓等，2013）。而对多模式系统而言，如何将多个数值模式结果综合集成一个准确、可靠、直观的预报结果是目前面临的重要问题。多模式集成方法是一种提高模式准确率的非常有效的后处理方法（杨学胜，2001）。Danard et al.（1968）和 Thompson（1977）发现，通过将两个相互独立的预测结果进行特定的线性组合，其预测均方根误差可以小于单个预测的均方根误差。基于此，多模式集成平均（Ensemble Mean, EMN）、消除偏差的集成平均（Bias-removed Ensemble Mean, BREM）、回归改进集成平均（Regression-improved Ensemble Mean）、超级集成法（Superensemble Prediction, SUP）（Krishnamurti et al., 1999, 2000）等集成方法被相继提出，且广泛运用到气象上的中短期预报、延伸期预报及气候预测中（智协飞等，2009, 2014, 2015, 2016；崔慧慧和智协飞，2013; 张涵斌等，2015）。以上研究皆认为，在众多集成方法中，超级集成法可以更有效的减小均方根误差，提高预报技巧。其思想是将时间序列分为训练期和预报期两段，通过在训练期建立模式距平与实况距平的回归方程来确定各个模式权重信息，并将这些权重信息运用于预报期内多模式集成预报。

近年来，多模式集成平均、回归改进集成平均等具有代表性的集成方法也被拓展运用到空气质量预报领域（王自发等，2009；黄思等，2015），但将超级集成法应用到环境空气质量要素的研究却鲜有案例；另外，以往超级集成法的研究多为针对某特定时间序列展开的预报试验（智协飞等，2009；刘建国等，2013），最佳训练期长度也多通过固定的培训期数据来确定，并未随时间变化；且未提出一套适用于业务预报的集成方法（崔慧慧和智协飞，2013；黄思等，2015; 智协飞等，2016）。

针对上述问题，本文尝试将多模式超级集成的方法进行改进，使其适用于实时业务化预报。基于长三角地区区域空气质量多模式预报系统 2015 年的预报结果，将改进后的超级集成法、超级集成法、多模式集成平均和消除偏差的集成平均等多模式集成方法运用到长三角地区 O_3 的 24 h 预报中，比较四种多模式集成方案的预报效果；然后将最优的集成方案运用到 48 h 和 72 h 预报中，考察其短期

（三天）预报水平。

2 资料和方法

2.1 长三角地区多模式空气质量预报系统

长三角地区多模式空气质量预报系统中，区域空气质量模式包括中国科学院大气物理研究所自主开发的嵌套网格空气质量预报模式 NAQPMS（王自发等，2006）、美国环保署（EPA）的 CMAQ 模式（Dennis et al., 1996）、美国 Environ 公司的空气质量综合模拟系统 CAMx 模式（ENVIRON International Corporation, 2002）及由美国 NOAA 预报系统实验室（FSL）开发的气象模式（WRF）和化学模式（CHEM）在线耦合的 WRF-CHEM 模式（Grell et al., 2005）。所有化学传输模式均采用三层嵌套，区域设置如图 1a 所示：第一层（D1）为东亚地区（170×160 格点），第二层（D2）覆盖中国东部地区（160×154 格点），第三层（D3）涵盖江苏省全部、山东南部、安徽大部及浙江北部（199×181 格点）。三层嵌套区域的水平空间分辨分别为 27、9、3 km。各个模式垂直方向上从地面到 20 km 分为 20 层。系统所用气象模式为 WRF 模式（3.5.1 版本）。WRF 模式水平网格设置与化学传输模式一致，采用地形追随坐标，垂直方向分为 30 层，模式层顶为 50 hPa。外层选择较大的区域可以保证长期模拟过程中内层区域具有较为合理的侧边界条件，并可以充分考虑大尺度天气系统对长三角地区的影响。

系统排放源输入是由清华大学基于中国多尺度排放清单模型 MEIC（Multi-resolution Emission Inventory for China，http://meicmodel.org/[2017-04-20]）制作的实时更新的排放源清单。大气污染物排放源分类是在参考《区域高分辨率大气排放源清单建立的技术方法与应用》及国内已有较为成熟的排放源分类体系的基础上，结合目前我国统计年鉴活动数据分类特征、国民经济行业分类情况以及江苏省的实际情况，针对 SO_2、NO_x、CO、PM_{10}、$PM_{2.5}$ 和 NH_3 等九种污染物，将排放源主要分为农业源、生物源、工业源、电厂源、居民源、交通源及其他源共七类。排放源的空间分辨率和化学传输模式及气象模式匹配，时间分辨率为 1 小时。

2.2 站点选取及观测资料

$$S_{\text{picked}} = \max(D(S_{\text{refer}}, S_{\text{candidate}})), \qquad (1)$$

长三角地区大部分空气质量监测站点的空间位置存在向城市集中叠加的现象，故按照测站之间的距离远近，挑选了有代表性的 26 个站点（图 1b）。具体算法是：任选一个站点作为基准站 S_{refer}，然后按（1）式从余下站点中选取代表站点。从备选站点 $S_{\text{candidate}}$ 中找出距离基准站 S_{refer} 最远的站点作为已选站点 S_{picked}，然后将已选站点从备选站点中剔除，重复（1）式表述的挑选过程，直到已选站点个数满足研究需求。

选取的空气质量数据来自中国环境监测总站，包括臭氧（O_3）和二氧化氮（NO_2）。研究时段为 2015 年 1～12 月，其中，1 月 1 日至 3 月 5 日为准备期，评估时段为 3 月 6 日至 12 月 31 日，按月分为冷暖两季，冷季为 3～4 月和 10～12 月，暖季为 5～9 月。模式数据为相应时段系统包含的 NAQPMS、CMAQ、CAMx 及 WRFCHEM 模式的 24、48 及 72 h 预报产品。观测及模式数据分辨率均为 1 h。

O_3 因极易受气象及城市排放条件影响，短时及

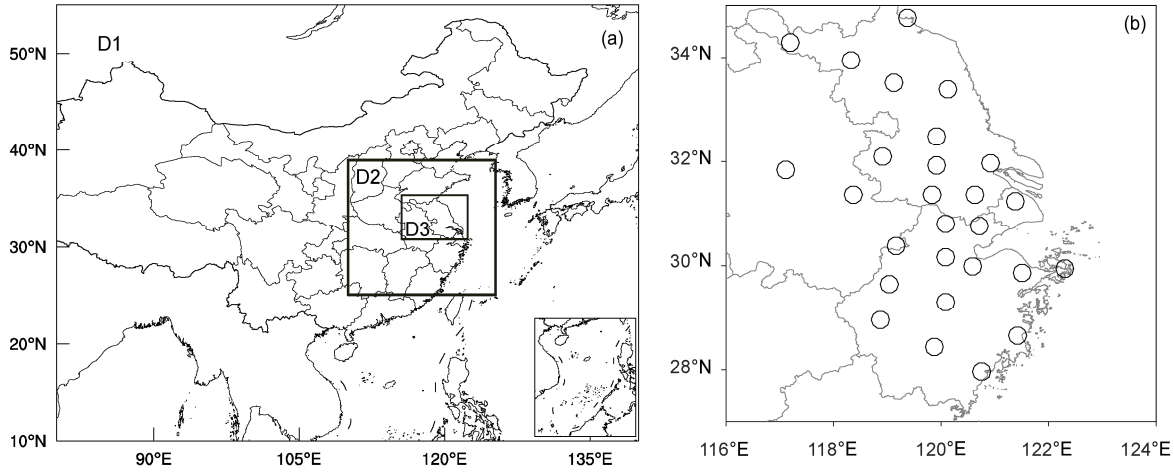

图 1 长三角地区（a）多模式系统区域设置，（b）代表空气质量监测站的分布

Fig. 1 (a) Domain setting of the multi-model ensemble air quality forecast system, (b) representative air quality monitoring stations for the Yangtze River Delta

日变化显著（殷永泉等，2004）；同时由于 O_3 关注短期急性健康效应，我国于 2012 年将日最大 8 h 浓度作为 O_3 日评估标准纳入到新的《环境空气质量标准》中（中华人民共和国环境保护部，2016）。故文中 O_3 的评估考察日最大 8 h 浓度，具体算法参考了美国环保署公布的 8 h O_3 数据处理指南，即先对小时分辨率资料进行 8 h 滑动平均，然后逐日对滑动平均结果求最大值得到（Park, 1998）。

2.3 集成方法

（1）多模式集成平均（EMN）

将多个模式的结果通过求平均转化为一个预报结果，即多模式集成平均：

$$V_{\mathrm{EMN}}(j,\ t)=\frac{1}{N}\sum_{i=1}^{N}F_i(j,\ t), \qquad (2)$$

其中，$V_{\mathrm{EMN}}(j, t)$为第 j 个测站 t 时刻的多模式集成平均的预报值，N 为参加集成的模式总数，$F_i(j, t)$ 为第 i 个模式在第 j 个测站 t 时刻的预报值。

（2）消除偏差的集成平均（BREM）

模式通常存在系统偏差，通过引入观测平均态，并在其基础上叠加模式预报距平，可得到无偏的集成平均：

$$V_{\mathrm{BREM}}(j,\ t)=\overline{O}(j,\ t-1)+\frac{1}{N}\sum_{i=1}^{N}\left[F_i(j,\ t)-\overline{F_i}(j,\ t)\right], \qquad (3)$$

其中，$V_{\mathrm{BREM}}(j, t)$为第 j 个测站 t 时刻的消除偏差的集成预报值，$\overline{O}(j,\ t-1)$ 为从初始时刻到 $t-1$ 时刻所有观测值的平均，$\overline{F_i}(j,\ t)$ 为第 i 个模式在第 j 个测站从初始时刻到 t 时刻预报值的平均。该方法与多模式集成平均的差异在于参与集成模式结果是无偏的，但此刻的集成仍是等权相加。

（3）超级集成法（SUP）及滑动训练期的超级集成（R-SUP）

Krishnamurti et al.（1999）指出，系统偏差信息还可通过模式在训练期的表现提取得到，并在预报期予以消除。通过选定一段训练期，在训练期获取各个模式的权重系数，并在训练期内得到观测平均态和模式距平，综合考虑模式的不等权重和消除偏差即得到超级集成预报（SUP）：

$$V_{\mathrm{SUP}}(j,\ t)=\overline{O}(j,\ t-1)+\sum_{i=1}^{N}a_i(j,\ t)\left[F_i(j,\ t)-\overline{F_i}(j,\ t)\right], \qquad (4)$$

其中，$V_{\mathrm{SUP}}(j, t)$ 为超级集成预报值，$a_i(j, t)$ 为权重系数。

$$G=\sum_{t=1}^{N_{\mathrm{tr}}}(V'_{\mathrm{SUP_t}}-O'_{\mathrm{t}})^2, \qquad (5)$$

根据最小二乘法原理建立训练期回归方程使得训练期中误差 G 最小，以此确定最佳的训练期长度及参与超级集合的各模式权重系数 $a_i(j, t)$，如（5）式所示。其中，V'_{SUPt} 为训练期超级集合预报值，O'_t 为实测值，N_{tr} 为训练期长度。$a_i(j, t)$ 的获取既可以采用线性回归，也可采用非线性神经网络技术（智协飞等，2009，2013）。将 $a_i(j, t)$ 代入公式（4），即可获得预报期的超级集成预报值。由于训练期的选取对集成预报结果有重要影响，一般需要通过试验来确定。以上的超级集成预报模型中，权重 $a_i(j, t)$ 不随时间变化，并将运用于整个训练期的预报。智协飞等（2009）指出，随着预报期的增长固定的权重系数会逐渐失效，并提出了滑动训练期的超级集成预报（R-SUP）方案，即最佳训练长度 N_{best} 随着预报期逐日滚动向前，由离预报日最近的前 N_{best} 天作为训练期，$a_i(j, t)$ 即可随着预报日的变化逐日更新。已有的研究表明，R-SUP 较 SUP 能够进一步改善预报效果（智协飞等，2009；刘建国等，2013）。本文中，将 2015 年 1 月 1 日至 3 月 16 日共计 75 天作为训练期，2015 年 5 月 1 日至 6 月 30 日共计 61 日作为预报期，根据（5）式采用线性回归的方法确定权重系数，考察训练期长度从 2 天逐日延长至 75 天时预报期预报效果。结果表明，在训练期为 31 天，预报期平均误差最小，故选取 31 天为最佳训练长度，并按上述滑动训练期的超级集成法制作了集成预报。

（4）改进的超级集成法（AR-SUP）

已有的超级集成法研究中，一般是将预报期固定后，逐渐增加训练期长度来获取相应的权重系数，通过考察不同权重系数下的集成预报结果优劣来选定最佳的训练期长度。这样得到的最佳训练 期长度和固定预报期的选取有关，一旦预报期变化，在调试阶段获取的最佳训练期长度势必不再是“最佳”。而在实际的业务预报中，预报期逐日更新，故最佳训练期长度应随预报期变化，最佳训练期也应合理涵盖历史数据，不断纳入新的观测信息。

基于此，设计一个用于挑选最佳训练期且长度可变的活动区间，区间右端为靠近某一预报日的相对固定端，左端则为远离预报日的活动端，在此活动区间内，左端不断向远离预报日的方向推进，活动区间的长度由此不断增加。另外，考虑到用活动区间确定最佳训练期势必需要活动的预报期来与

之配合，本文中，在预报日活动区间之间紧邻活动区间的右侧设计一个“试预测期”，用试预测期的预报效果来选定针对某一预报日的最佳训练期长度和与之对应的最优权重系数。随着预报日的更新变化，活动区间和试预测期向前逐日滚动推进。活动区间长度的变化范围最短定为 2 d，参考相关文献，将最大长度定为 60 d（智协飞等，2009；崔慧慧和智协飞，2013；黄思等，2015）。试预测期长度最优为 4 d，具体的确定方法后面再详细讨论。

按上述思想，将首个最大活动区间及其临近试预测期作为准备期（1 月 1 日至 3 月 5 日）。在首个最大活动区间及其临近试预测期以后（即自 3 月 6 日始），$a_i(j, t)$ 按照上述活动区间的方法获取。本文中按如下流程进行逐日的超级集成预报：

（1）在活动区间内，训练期由 2 d 到 60 d 逐渐增加。在每一个训练期内，按（5）式建立回归方程，由最小二乘法原理使训练期误差 G 最小来得到不同训练期条件下的 $a_i(j, t)$ 组合。

$$L=\min\left(\sum_{m=1}^{N_{\mathrm{tf}}}\left[V_{km}(j,t)-O_m(j,t)\right]^2\right),\ k=1,\cdots,N_{\mathrm{w}}. \tag{6}$$

（2）将 $a_i(j, t)$ 组合代入（4）式得到试预测期超级集成预报值 $V_{km}(j,t)$ 试预测期内绝对误差和最小即（6）式为判断标准，寻找最佳的 $a_i(j, t)$ 组合，代入（4）式即得到准备期后第一天的预报。其中，$V_{km}(j,t)$ 第 k 种权重系数组合在第 m 个试预测天的集成预报值，$O_m(j,t)$ 为第 m 个试预测天的观测值，N_{tf} 为试预测期长度，此处为 4 d；N_{w} 为训练期个数即权重组合数，此处为 59。

（3）按此方法，训练长度挑选区间及试预测期向前逐日滚动推进来获取此后每一日的超级集成预报值。

2.4　评估参数

为评估改进后的超级集合算法的预报能力，本文中用到的常用预报能力评估参数包括：均方根误差（RMSE）、标准化平均误差（NME）、平均偏差（MB）、标准化平均偏差（NMB）、相关系数（r）、IOA（一致性指数）。其中，前四项为误差评估参数；相关系数考察观测和预报时间序列的相关程度，一定意义上反应了模式趋势预报的能力；一致性指数综合体现了模式与观测的一致程度（Kang et al., 2010）。各参数表达式见表 1。

表 1　评估参数定义

Table 1　Definitions of evaluation parameters

参　数	表　达　式
均方根误差（RMSE）	$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_1^N(\mathrm{Model}-\mathrm{Obs})^2}$
平均偏差（MB）	$\mathrm{MB}=\frac{1}{N}\sum_1^N(\mathrm{Model}-\mathrm{Obs})$
标准化平均偏差（NMB）	$\mathrm{NMB}=\frac{\sum_1^N(\mathrm{Model}-\mathrm{Obs})}{\sum_1^N\mathrm{Obs}}\times100\%$
标准化平均误差（NME）	$\mathrm{NME}=\frac{\sum_1^N\lvert\mathrm{Model}-\mathrm{Obs}\rvert}{\sum_1^N\mathrm{Obs}}\times100\%$
相关系数（r）	$t_{\mathrm{mp}}=\mathrm{Cov}(\mathrm{Model}-\mathrm{Obs}),$ $r=t_{\mathrm{mp}}(1,2)/\left(t_{\mathrm{mp}}(1,1)\cdot t_{\mathrm{mp}}(2,2)\right)$
一致性指数（IOA）	$\mathrm{IOA}=1-\frac{\sum_1^N(\mathrm{Model}-\mathrm{Obs})^2}{\sum_1^N\left(\lvert\mathrm{Model}-\mathrm{Avg(Obs)}\rvert+\lvert\mathrm{Obs}-\mathrm{Avg(Obs)}\rvert\right)^2}$

注：Cov 表示协方差；t_{mp} 为中间变量，表示模式值与观测值的协方差（Cov），是一个 2×2 的矩阵；Avg 表示平均值。

3　结果与讨论

3.1　单个模式预报效果评估

计算各评估参数在冷暖季的区域平均（各代表站平均）值，如表 2 所示。整体来看，暖季各模式 RMSE 在 42 μg m^{-3} 以上，冷季在 33 μg m^{-3} 以上，结合 RMSE 和 NME 来看，CAMx 模式误差最小。MB 的结果各个模式差异较大：NAQPMS 和 WRFCHEM 模式存在低估，且暖季尤为显著，分别可达−26.84 μg m^{-3} 和−22.21 μg m^{-3}，而 CMAQ 和 CAMx 模式在冷季的正误差明显，在 6～8 μg m^{-3} 左右。NMB 的结果与 MB 的结果类似。IOA 与相关系数的结果较一致，CMAQ 和 CAMx 模式高于 NAQPMS 和 WRFCHEM 模式，冷季略高于暖季，且四者比较而言，CAMx 模式最高。

$$\mathrm{NO_2}+h_{\mathrm{v}}\rightarrow\mathrm{NO}+\mathrm{O}(^3P), \tag{7}$$

表 2　单模式预报评估参数的各代表测站平均值对比

Table 2　Comparison of evaluation parameters for single models averaged at all representative stations

	RMSE/μg m^{-3}		NME		MB/μg m^{-3}		NMB		r		IOA	
	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季
NAQPMS 模式	52.95	41.05	46.08%	42.99%	−26.84	−18.40	−10.21%	−14.72%	0.41	0.53	0.55	0.63
CMAQ 模式	44.74	36.52	51.43%	43.11%	−3.35	6.24	19.36%	17.56%	0.53	0.59	0.66	0.70
CAMx 模式	42.40	33.55	48.30%	41.80%	1.33	7.70	22.13%	19.25%	0.52	0.61	0.67	0.71
WRFCHEM 模式	53.09	41.58	48.77%	41.97%	−22.21	−15.83	−1.74%	−9.80%	0.32	0.48	0.49	0.55

$$O(^3P)+O_2+M\rightarrow O_3+M\ , \qquad (8)$$

$$O_3+NO\rightarrow NO_2+O_2. \qquad (9)$$

Haggen-Smit（1952）指出，氮氧化物（NO_x）是对流层 O_3 的主要前体物之一。在太阳辐射强度稳定的理想状态下，NO、NO_2、O_3 三者处于（7）～（9）式所示的化学平衡状态。其中，h_v 为光能；$O(^3P)$ 为激发态的氧原子；M 为另一物种，如 N_2 或者 O_2 的一个分子，因臭氧的稳定性极差，M 能够吸收反应热量，使得 O_3 不至于温度过高而分解。换言之，氮氧化物浓度对 O_3 浓度有重要影响，故模式前体物预报偏差可能是导致 O_3 的预报偏差的原因之一。为验证这一想法，将 O_3 与 NO_2 浓度之和作为考察指标，将 2015 年各单模式 O_3 与 NO_2 逐小时浓度相加，计算区域评估参数并与同时段单项 O_3 对比（图 2）。和单项 O_3 预报相比，各模式的 O_3+NO_2 的 RMSE 平均下降了 1.5 μg m^{-3}，其中 CAMx 模式下降了 2.7 μg m^{-3}。NMB 的结果则表明，各模式对 O_3+NO_2 预报的相对误差较 O_3 显著减小，其中 NAQPMS 和 WRFCHEM 模式的负偏差平均减小了 20%。

总体而言，各模式仍存在较大的预报误差，且暖季的预报误差大于冷季，CMAQ 和 CAMx 模式在冷季存在明显高估，而 NAQPMS 和 WRFCHEM 模式在暖季存在明显低估。各模式对氮氧化物的预报偏差是导致 O_3 预报误差的重要原因之一，模式在前体物与 O_3 化学转化方面还有待进一步研究。各模式综合比较而言，CAMx 模式预报误差最小，且与观测保持了较好的相关性和一致性。

3.2 不同集成方法的对比评估

3.2.1 整体效果对比评估

用改进后的超级集成方法（AR-SUP）、滑动训练期的超级集成方法（R-SUP）、多模式集成平均方法（EMN）和消除偏差的集成平均方法（BREM）四种不同集成方法对多模式系统的 O_3 预报进行集成，并计算评估参数的区域平均值（表 3）。与最优单模式的结果对比：四种多模式集成方法在冷、暖两季均降低了预报误差，具体来看：AR-SUP 和 RUP 方法误差降低较显著，二者相较 AR-SUP 方法更优。RMSE 较最优单模式在暖季和冷季分别降低了 8.67 μg m^{-3} 和 7.71 μg m^{-3}，NME 则分别降低了 24.94% 和 17.39%；EMN 方法降幅最不明显，RMSE 与 NME 与最优单模式的结果接近。从 MB 和 NMB 可以看出，采用 R-SUP 方法后，MB 控制在±1.5 μg m^{-3} 以内，NMB 亦在±1%以内，而 AR-SUP 方法则在此基础上进一步将 MB 控制在±0.2 μg m^{-3} 以内，NMB 控制在±0.5%以内，较最优单模式分别下降超过 7 μg m^{-3} 和 20%，改善明显；EMN 和 BREM 方法改善了单模式在冷季的高估现象，但却整体出现了负偏差，在暖季尤为明显。这是由于 EMN 和 BREM 方法虽有差异，但均是基于等权的多模式集成，而从单模式评估结果可知，NAQPMS 和

表 3 四种集成预报的评估参数的各代表测站平均值对比

Table 3 Comparison of evaluation parameters averaged at all representative stations for four ensemble forecasts

	RMSE/μg m^{-3}		NME		MB/μg m^{-3}		NMB		r		IOA	
	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季	暖季	冷季
AR-SUP	33.73	25.84	23.36%	24.41%	−0.14	0.06	0.37%	0.29%	0.61	0.75	0.78	0.86
R-SUP	36.12	28.56	25.76%	26.50%	0.76	−1.15	0.73%	−0.52%	0.50	0.67	0.70	0.78
EMN	44.33	35.19	46.86%	39.14%	−12.57	−5.41	−7.54%	−2.69%	0.55	0.61	0.62	0.68
BREM	36.79	32.84	31.62%	35.33%	−10.69	−4.32	−3.87%	−5.68%	0.53	0.58	0.64	0.71

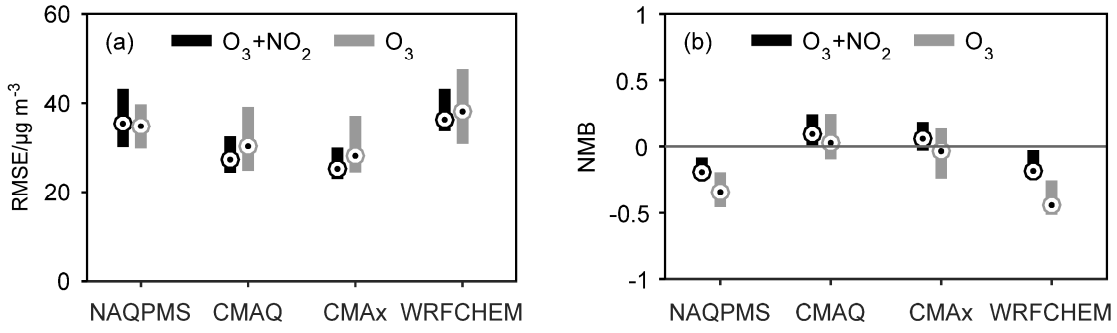

图 2 2015 年 3 月 6 日至 12 月 3 日各代表测站平均的 O_3+NO_2 与 O_3 浓度的（a）RMSE（单位：μg m^{-3}）、（b）NMB

Fig. 2 (a) RMSE (Root Mean Square Error, units: μg m^{-3}) and (b) NMB (Normalized Mean Bias) for O_3+NO_2 and O_3 concentrations averaged at all representative stations from 6 March 2015 to 3 December 2015

WRFCHEM 模式负偏差显著，导致了 EMN 和 BREM 方法集成出现负偏差。采用 AR-SUP 方法集成预报后相关系数也得到明显提升，两季分别提升了 0.06 和 0.12，而其余方法和最优单模式接近。从区域平均的角度结合一致性指数 IOA 来看不同集成方法对预报技巧的改善情况：AR-SUP 方法的 IOA 指数较最优单模式提高了 0.09 和 0.14；R-SUP 方法次之，提升值分别为 0.03 和 0.06；BREM 方法和最优单模式接近；EMN 方法则较最优单模式下降了 0.05 和 0.03。这说明，参与集成的单模式预报技巧差异较大时，对各模式结果进行算术平均或消除偏差的算术平均并不能明显提升预报技巧，算术集成平均甚至会出现集成预报技巧低于最优单模式的情况。相较于前两者，R-SUP 方法能够明显减小单模式预报误差，而 AR-SUP 方法在此基础上能够进一步提高预报技巧。

将 2015 年 3 月 6 日至 12 月 3 日日最大 8 小时 O_3 集成预报与观测值的区域平均值绘成时间序列（图 3）。从 EMN 和 BREM 方法预报结果来看，3～4 月中旬，二者（蓝色、绿色线）略高于观测值（黑色圆圈）；4 月中旬进入暖季后高污染事件出现明显增多，预报低估逐渐凸显；11 月，预报和观测均吻合较好；2 月以后，预报整体偏高。这说明 EMN 和 BREM 方法在冷季的部分时段（3 月、12 月）存在高估现象，在冷暖过渡的季节（4 月、11 月）与观测时间序列同步较好，而在暖季（5～9 月）和 10 月则存在出现明显低估。这可能与 CAMx 模式与 CMAQ 模式正偏差明显、NAQPMS 与 WRFCHEM 模式负偏差明显有关。从 R-SUP 方法得到的时间序列（紫色线）与观测值的对比来看，R-SUP 方法能够改善 EMN 和 BREM 方法在暖季明显低估的情况，但对暖季部分时段的高污染过程的谷值预报整体偏高（5～7 月上旬、9 月中下旬），在冷季的部分时段（3 月中下旬、10 月中下旬至 11 月上旬）出现低估。AR-SUP 方法（红色线）相较于 R-SUP 方法与观测值保持了更好的一致性，其对 4 月下旬、5 月下旬、6 月中旬、8 月上旬及下旬、9 月上旬和 10 月下旬的高污染事件（峰值）和污染清除的转折阶段（谷值）的出现时间和浓度值预报较准确。究其原因在于，R-SUP 方法是历史观测平均态基础上的模式距平加权集成，而历史观测平均态决定于训练期内的历史观测值，R-SUP 方法训练期长度固定，实际预报值易受训练期内观测平均态的影响而出现谷值偏高、峰值错位等现象；而 AR-SUP 方法通过试预测期的预报效果来不断调整训练期长度，合理选取观测平均态，从而能够进一步逼近观测，减小误差。

为进一步考察不同集成方法对实况观测动态变化的预报能力及在不同观测浓度水平下不同集成方法的预报偏差变化趋势，按 Mathur et al.（2008）提出的方法，计算各代表测站不同集成预报值及观测值在冷、暖季的浓度频率分布函数，截取集成预报值的 5%、25%、50%、75%和 95%百分位点与相应百分位点观测值对比（百分位点大小表示浓度的高低）。为了便于分析，在图 4 中增加了 1:2、1:1 以及 2:1 三条特征线。受各单模式预报水平影响，EMN 和 BREM 方法在暖、冷两季，存在低浓度预报整体偏高（5%、25%百分位点偏高），高浓度预报整体偏低（75%、95%百分位点偏低）的现象，且冷季更为明显，二者与 1:1 特征线拟合优度系数在 0.85 以下。R-SUP 方法在暖季的预报与观测值散点基本沿着 1:1 特征线集中分布，拟合优度系数为 0.9；而在冷季，预报与观测值散点沿 1:1 特征线的集中度降低，较低浓度（5%、25%百分位点）沿 1:2 和 1:1 特征线之间的区域散布，较高浓度在 1:1 和 2:1 特征线之间的区域散布，拟合优度系数降至 0.79；R-SUP 方法在冷季各百分位点预报偏差增大的原因可能是由于冷季 O_3 浓度较低使得历史平均态影响增加，进而出现低浓度预报偏高，高浓度预报偏低的现象。AR-SUP 方法则一定程度上修正了上述偏差，与 1:1 特征线拟合优度系数保持在 0.91。

3.2.2 时空表征能力评估

为考察集成预报法的空间表征能力，图 5 给出了不同集成预报绝对误差（MB）的年均值在长三角地区的空间分布。从图中可以看出，EMN 方法的结果在暖季长三角的整个中东部地区明显低估（<-5 μg m^{-3}）m，西部地区则为明显高估（>5 μg m^{-3}）；在冷季，低估的范围和强度有所减缓，但在中部地区仍为明显低估（<-5 μg m^{-3}），且除西部地区外，北部和南部的部分站点也出现了高估的现象。从 BREM 方法的预报结果来看，在暖季，MB 的分布与 EMN 方法的结果近似，但就误差的范围和强度而言，优于多模式集成平均；在冷季，除北部的个别站点及中部的部分站点预报偏低（<-5 μg m^{-3}）外，大部分站点的 MB 在 ± 5 μg m^{-3} 以内。从空间分布上看，R-SUP 方法较前两者能够明显减小误

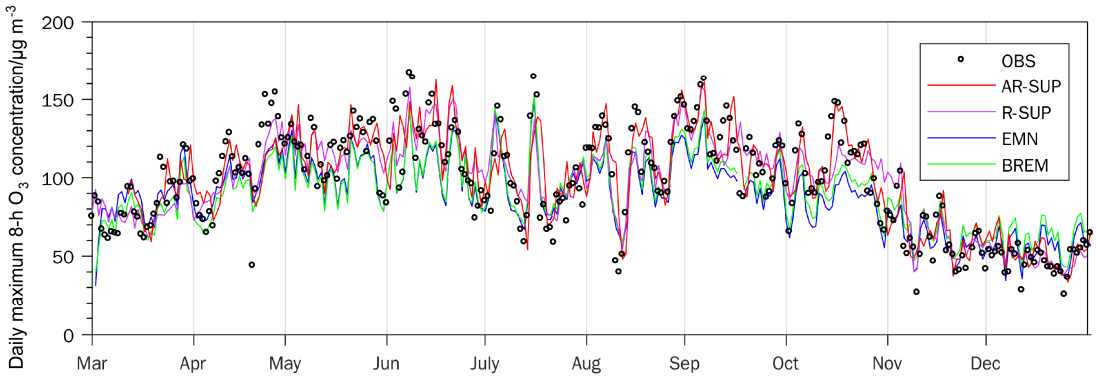

图 3 2015 年 3 月 6 日至 12 月 3 日观测的与集成预报的日最大 8 小时 O_3 浓度区域平均时间序列

Fig. 3 Time series of regional averages of daily maximum 8-h O_3 concentrations from observations and ensemble forecasts from 6 March 2015 to 3 December 2015

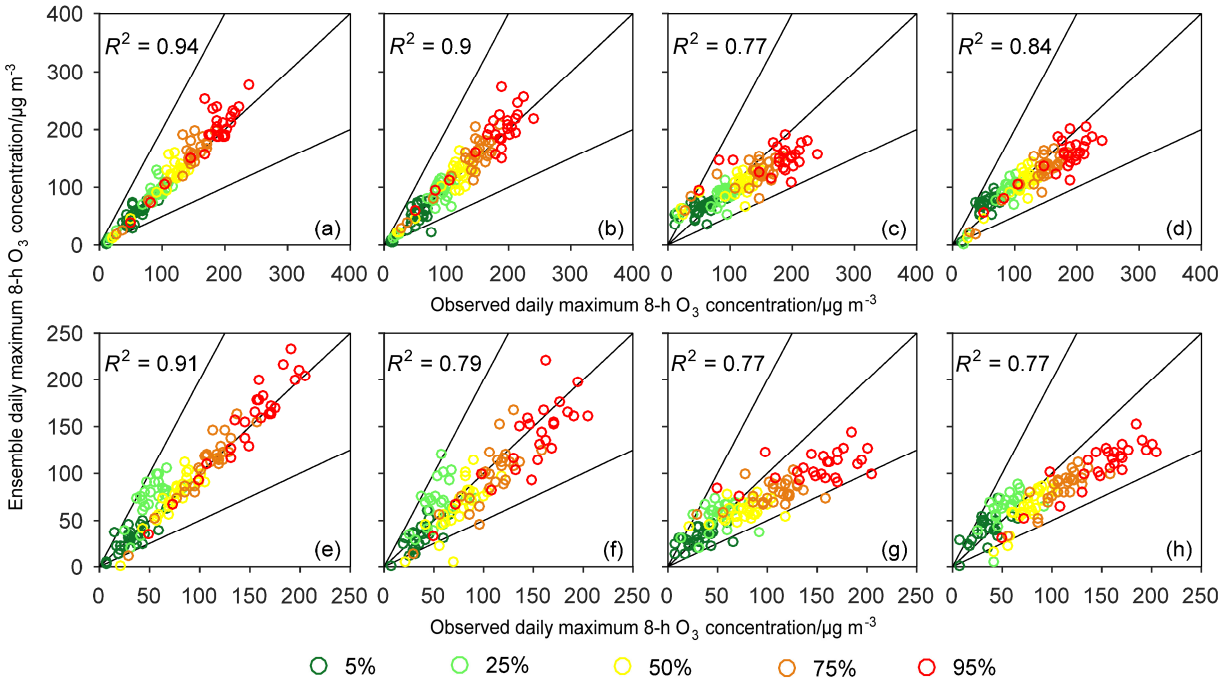

图 4 2015 年 3 月 6 日至 12 月 3 日各代表测站平均的观测值与多模式集成预报的日最大 8 小时 O_3 浓度在选定百分位点的散点图：（a、d）改进的超级集成方法 AR-SUP；（b、f）滑动训练期的超级集成方法 R-SUP；（c、g）多模式集成平均方法 EMN；（d、h）消除偏差的集成平均方法 BREM。上排为暖季，下排为冷季

Fig. 4 Scatterplots of daily maximum 8-h O_3 concentrations from observations and multi-model ensemble forecasts averaged at all representative stations for selected percentiles from 6 March 2015 to 3 December 2015: (a, d) AR-SUP (Advanced Running-Super ensemble Prediction); (b, e) R-SUP (Running-Super ensemble Prediction); (c, f) EMN (Ensemble Mean); (d, h) BREM (Bias-Removed Ensemble Mean). Upper panels indicate warm season, bottom panels indicate cool season

差，在暖季，长三角中部地区大部分站点的 MB 均控制在±2 μg m^{-3} 以内，但在北部及东南部的部分站点略偏高（>2 μg m^{-3}），个别站点存在明显低估的现象（<−5 μg m^{-3}）；在冷季，误差为±2 μg m^{-3} 站点范围缩小，在北部和东部低估站点（<−2 μg m^{-3}）范围扩大，西部则有高估站点（>2 μg m^{-3}）出现。与 R-SUP 方法比较，AR-SUP 方法在暖、冷两季 MB 在±2 μg m^{-3} 以内的站点范围进一步扩大；但在中部有一个站点在暖季出现明显低估（<−5 μg m^{-3}）、冷季明显高估（>5 μg m^{-3}）。查找原因后发现，该站点的观测资料自 2015 年 10 月下旬开始缺测，而本文中规定，当活动区间内缺测值超过半数，基于活动区间超级集成失效，以消除偏差的集成预报（BREM）代替，导致 AR-SUP 方法在该站点结果

出现较大偏差。

为进一步比较不同集成方法在全年的时空表征能力，将各个测站的集成预报值和观测值按照表 1 逐月计算评估参数，绘制成盒状图（图 6）。逐月比较来看，AR-SUP 方法的 RMSE 总体最低，R-SUP 方法与 AR-SUP 方法接近，BREM 方法次之，EMN 方法最大。MB 的结果显示 AR-SUP 方法的优势明显，各月中位数在 0 左右，百分位之差亦控制在 5 μg m^{-3}，无明显季节差异；R-SUP 方法结果略高于 AR-SUP 方法，各月中位数绝对值在 5 μg m^{-3} 左右，百分位之差在 10 μg m^{-3} 左右；EMN 和 BREM 方法的 MB 要明显大于前两者，二者在暖季的 MB 达到或超过了−20 μg m^{-3}，冷季的 MB 相对减小，但百分位之差超过 10 μg m^{-3}。EMN 和 BREM 方法的相关系数保持了较高水平，二者在暖季可达到或超过 0.6，冷季亦维持在 0.4 以上；R-SUP 方法的相关系

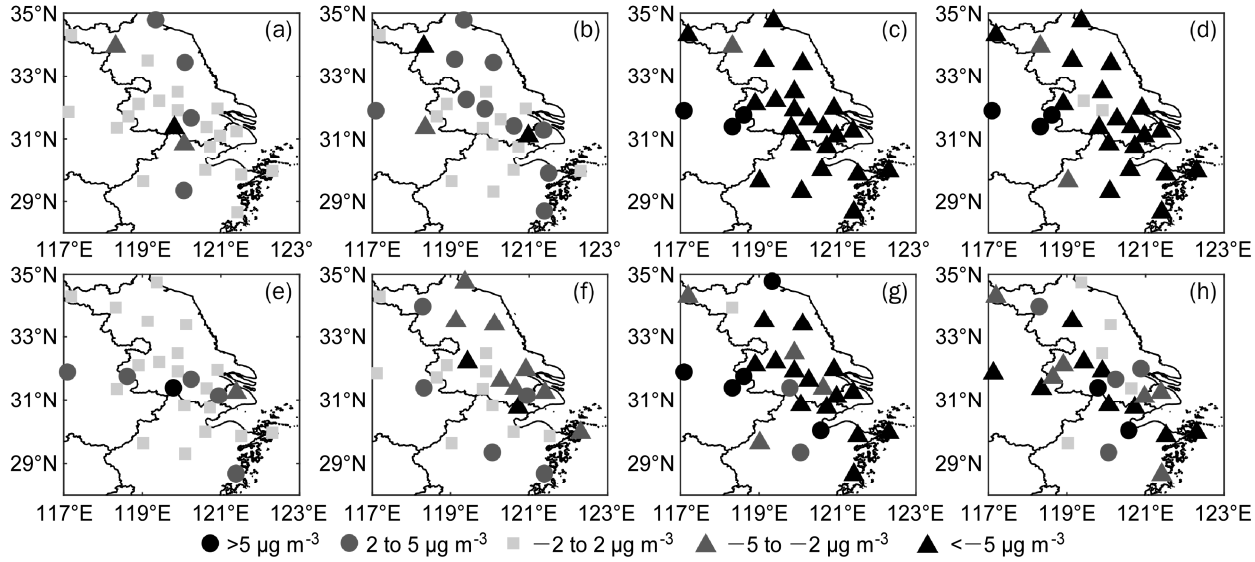

图 5 2015 年 3 月 6 日至 12 月 3 日观测值的与多模式集成预报的日最大 8 小时 O_3 浓度绝对误差（MB）平均值在代表站点的空间分布：（a、e）改进的超级集成方法 AR-SUP；（b、f）滑动训练期的超级集成方法 R-SUP；（c、g）多模式集成平均方法 EMN；（d、h）消除偏差的集成平均方法 BREM。上排为暖季，下排为冷季

Fig. 5 Mean biases (MB) of daily maximum 8-h O_3 concentrations from observations and multi-model ensemble forecasts averaged at all representative stations from 6 March 2015 to 3 December 2015: (a, e) AR-SUP; (b, f) R-SUP; (c, g) EMN; (d, h) BREM. Upper panels indicate. Upper panels indicate warm season, bottom panels indicate cool season

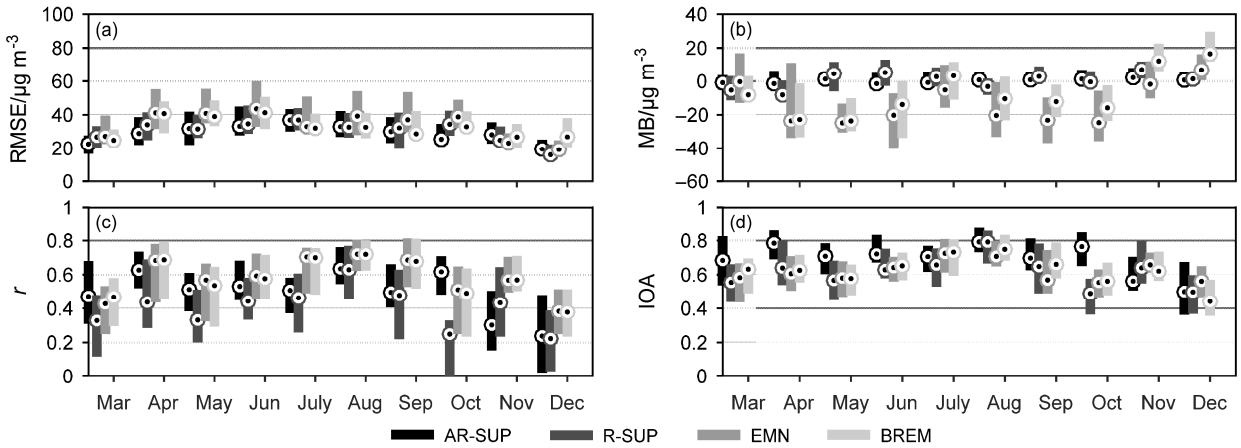

图 6 2015 年 3 月 6 日至 12 月 3 日各代表测站平均的集成预报的日最大 8 小时 O_3 浓度逐月的（a）RMSE（单位：μg m^{-3}）、（b）MB（单位：μg m^{-3}）、（c）相关系数 r、（d）IOA 箱式图。仅显示第一四分位、第三四分位数据及中位数

Fig. 6 Box plots of evaluation parameters of monthly mean daily maximum 8-h O_3 concentrations from ensemble forecasts averaged at all representative stations from 6 March 2015 to 3 December 2015: (a) RMSE (units: μg m^{-3}), (b) MB (units: μg m^{-3}), (c) correlation coefficient r, (d) IOA (Index of Agreement). Only 25th, 75th percentiles, and median values are shown

数最低，暖季在 0.4 左右，冷季跌至 0.2 左右。AR-SUP 方法则在 R-SUP 方法基础上整体将相关系数提升了 0.2 左右，在大部分月份达到或超过了 EMN 和 BREM 方法的相关系数水平。分析其原因在于：BREM 和 EMN 方法基于单模式的等权集成，能较好保留单模式良好的趋势预报能力。R-SUP 方法从减小误差的角度来挑选权重系数，在有限逼近观测的过程中势必导致模式预测时间序列变化趋势改变，这种改变在短时间内、观测浓度较低时相对明显，其表现为冷季的月尺度相关性显著降低。而 AR-SUP 方法能够较 R-SUP 方法进一步逼近观测，这在一定程度上修正了时间序列变化趋势的改变，从而提高与观测序列的相关性。从 IOA 指数的结果来看，AR-SUP 方法明显优于其余三种方法，其 IOA 指数在绝大部分月份超过了 0.6，部分月份（4、8、9 月）达到甚至超过 0.8，整体较其余三种方法高出约 0.1；其余三种方法比较而言，R-SUP 和 BREM 方法接近，R-SUP 方法略高，EMN 方法最低。

从上述评估结果可知，四种方法比较而言，AR-SUP 方法能够有效的减小误差，在冷、暖季均能保持高水平且稳定的预报技巧，预报效果要明显优于其余三种集成方法。

3.3 改进的超级集成法在短期预报中的应用

在日常的业务预报中，未来三天的空气质量均是关注重点。将 AR-SUP 方法进一步运用到 48 h 和 72 h 预报中，并与最优单模式进行对比评估，检验该方法短期预报效果。

图 7 为 72 h 内预报评估参数箱式图对比，从图中可以看出，CAMx 模式暖季的 RMSE 随着预报时效的增加而增长，其中位数从 37.7 μg m^{-3} 升至 41.4 μg m^{-3}；而 AR-SUP 方法预报则稳定在 36 μg m^{-3} 左右，二者差距随时间增长，冷季与暖季结果类似。这从二者 MB 的对比中有更加直观的体现：CAMx 模式在暖季为较大的负偏差，且偏差随预报时效的增加表现愈加显著；而 AR-SUP 方法的 MB 则稳定在±5 μg m^{-3} 以内。从相关系数（*r*）的对比结果看：CAMx 模式暖季 24 h 预报的 *r* 中位数高于 AR-SUP 方法，随着预报时效的增加，CAMx 模式的 *r* 中位数下降且百分位之差显著增大，至 72 h 与 AR-SUP 方法虽中位数接近，整体水平已显著低于 AR-SUP 方法。在冷季，AR-SUP 方法的 *r* 中位数稳定在 0.78 左右，且百分位之差亦稳定在 0.3 左右；而 CAMx 模式则整体随时间下降。各项评估参数的结果都表明 AR-SUP 方法的预报技巧并未随着预报时效的增加出现明显降低。

在日常的业务预报中，环境空气质量指数（Air Quality Index, AQI）等级准确率亦是考察预报能力的一项重要指标。按照环境空气质量指数（AQI）技术规定中关于空气质量分指数（Individual Air Quality Index, IAQI）的相关定义及其等级划分规定（中华人民共和国环境保护部，2016），计算了 O_3

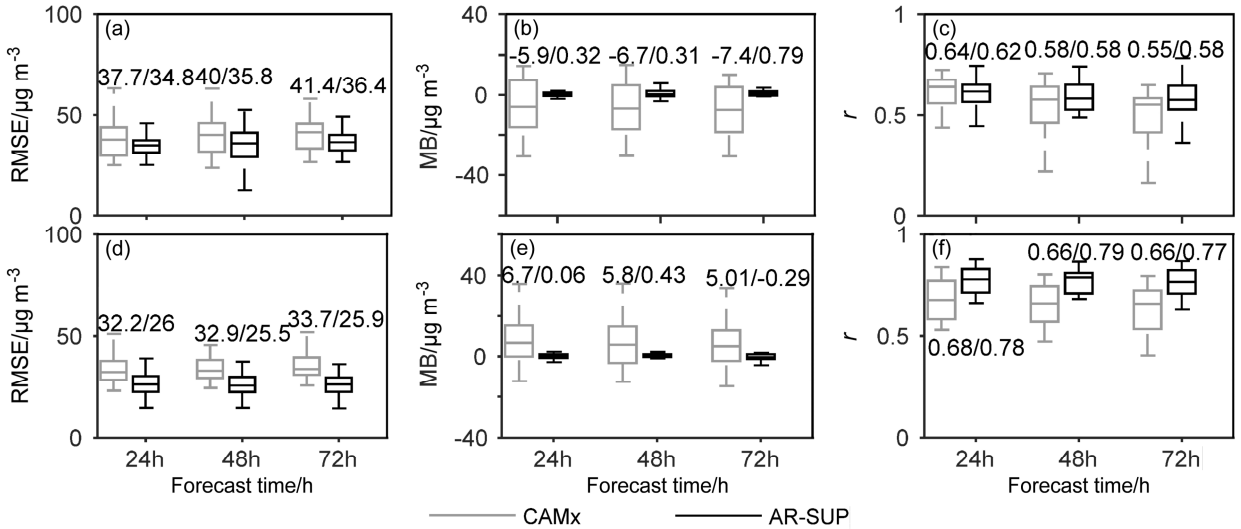

图 7 2015 年 3 月 6 日至 12 月 3 日改进的超级集成方法 AR-SUP 预报的与最优单模式 CAMx 模拟的各代表测站平均的日最大 8 小时 O_3 浓度在 72 h 内预报的（a、d）RMSE（单位：μg m^{-3}）、（b、e）MB（单位：μg m^{-3}）、（c、f）相关系数 *r* 的箱式图。上排为暖季，下排为冷季

Fig. 7 Box plots of evaluation parameters of daily maximum 8-h O_3 concentration from CAMx model and AR-SUP averaged at all representative stations within 72 hours from 6 March 2015 to 3 December 2015: (a, d) RMSE (units: μg m^{-3}); (b, e) MB (units: μg m^{-3}); (c, f) correlation coefficient *r*. Upper panels indicate warm season, bottom panels indicate cool season

的环境空气质量分指数，并对 72 h 内 CAMx 模式和 AR-SUP 方法的 IAQI 预报的等级准确率进行了对比分析（图 8）。其中，灰色和黑色盒状图分别对应 CAMx 模式与 AR-SUP 的结果。等级准确率（c_{H}）（Eder et al.，2009）定义为：$c_{\mathrm{H}i} = N_f^i / N_{\mathrm{obs}}^i$，其中，$i$ 为分指数等级（1，…，6），$c_{\mathrm{H}i}$ 表示第 i 等级准确率，N_f^i 为预报发生第 i 等级的事件个数，N_{obs}^i 为观测出现第 i 等级的事件个数。本文中，所有测站并未出现六级以上的 O_3 污染事件，故图中仅列出五级以内的准确率对比。从 24 h 预报等级准确率对比可知，相较于 CAMx 模式，AR-SUP 方法对各等级准确率均有所提高，其中，一级和二级污染准确率均超过 70%，较 CAMx 模式提高了 10%和 6%，三级和四级污染准确率则分别提升了 20%和 15%，五级污染时 CAMx 模式准确率降至 0，而此时 AR-SUP 方法仍具备一定的预报能力。在此基础上考察 48 h 和 72 h 预报可知，随着预报时效的增加，CAMx 模式各等级准确率均有不同程度的下降，污染等级越高下降愈加明显，至 72 h 三级和四级污染准确率分别下降了 8%和 4.1%，而反观 AR-SUP 方法，一至四级污染准确率并未出现明显一致下跌，准确率水平分别稳定在 80%、69%、40%和 20%，因五级污染仅在少数站点零星出现，且受 CAMx 等单模式准确率持续为 0 的影响，AR-SUP 方法准确率出现较大波动，48 h 升至 17%，72 h 跌至 0。

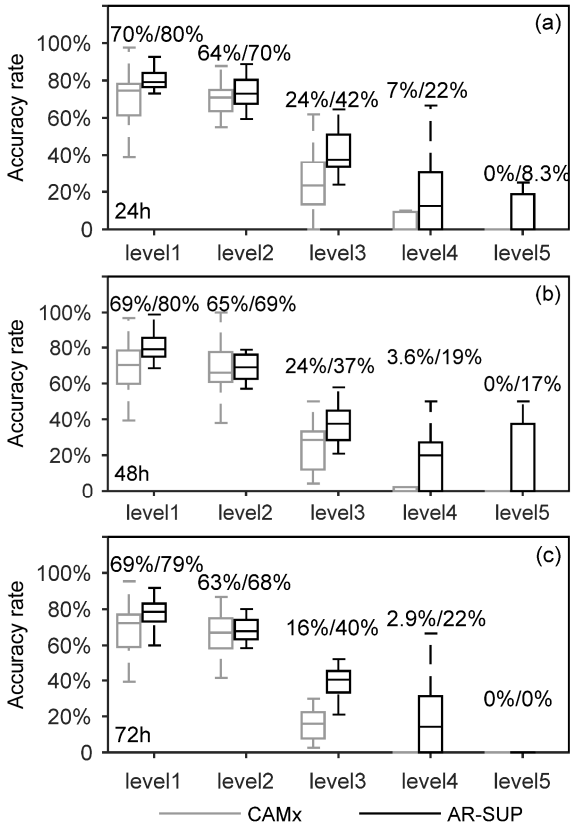

图 8　改进的超级集成方法 AR-SUP 与最优单模式 CAMx 的各代表测站平均的 O_3 空气质量分指数的等级准确率箱式图。预报时间分别为：（a）24 h；（b）48 h；（c）72 h。仅显示第一四分位、第三四分位数据及中位数

Fig. 8　Box plots of level accuracy of O_3 IAQI (Individual Air Quality Index) averaged at all representative stations from CAMx model and AR-SUP. Forecast time: (a) 24 h; (b) 48 h; (c) 72 h. Only 25th, 75th percentiles, and median values are shown

3.4　最优试预测期的选取

在改进的超级集成方法（AR-SUP）中，试预测期长度也是影响预报效果的关键因子。将 2015 年 8～9 月、11～12 月分别作为暖季和冷季试验期，均方根误差 RMSE 和一致性指数 IOA 作为考察预报准确率的两项评估参数。将试预测期从 1～30 天逐渐增加，计算在不同试预测期下代表测站 RMSE 和 IOA 的区域平均值变化情况，并与同期消除偏差的集成平均预报（BREM）的 RMSE 和 IOA 对比（图 9）。

最优试预测期长度的选取应遵循 RMSE 较低、IOA 较高的原则，同时兼顾考虑计算效率。从图 9 中可知，在暖季，随试预测天数增加 AR-SUP 方法的 RMSE 呈一致下降的趋势；而在冷季，RMSE 则呈现先降后升的趋势；IOA 指数在冷暖季皆表现为先升后降的趋势。AR-SUP 方法的 RMSE 和 IOA 指

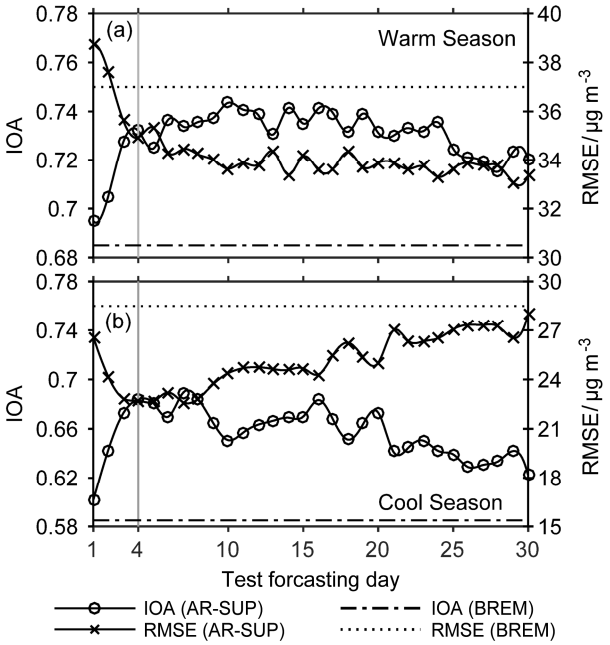

图 9　试验期内代表测站平均的 IOA 和 RMSE 随试预测期变化：（a）暖季；（b）冷季

Fig. 9　Changes of IOA and RMSE with the number of forecasting days averaged at all representative station during the trial periods: (a) Warm season; (b) cool season

数在冷、暖季节的变化虽各不相同，但在试预测期天数较短时，均存在一个显著变化期：当试预测期小于 3 天时，暖季 AR-SUP 方法的 RMSE 大于 BREM 方法的，说明试预测期天数过短并不能有效的减小误差、提高准确率。这是由于在实际业务预报中，预测日当天的观测值未知，故预测日的最佳训练期和模式权重并不可得，而此时经由活动区间选取的最佳训练期长度及最佳权重，只是针对试预测预报的“最佳”。试预测期过短，会使得最佳训练期长度及权重获取的偶然性和不确定性增大，导致预报准确率降低，相应 RMSE 较大、IOA 指数较低。所以，需要适当延长试预测期，使得最佳训练长度及权重更具普适性，得到针对预测日的“相对最佳”训练期及模式权重。但试预测期亦不可过长，过长的试预测期会导致计算效率降低。所以需合理挑选试预测期长度。可以看到，试预测期由 1 天延长至 4 天时，RMSE 迅速下降，IOA 指数则明显升高；而 4 天以后，RMSE 的降幅明显减缓，在冷季则呈现出缓慢上升的趋势，而 IOA 指数则开始缓慢上升，说明此时延长试预测期并不能有效提高准确率且 RMSE 和 IOA 指数对试预测期的增加不再敏感。故从 RMSE 和 IOA 指数的变化趋势及计算效率综合考虑，选择 4 天为最佳的试预测期。

4 结论

本文将超级集成方法进行改进，使其适用于日常的业务预报。将改进后的超级集成方法、滑动训练期的超级集成法、多模式集成平均的方法、消除偏差的集成平均法与长三角地区多模式空气质量预报系统的预报结果相结合制作了 2015 年长三角地区的 O_3 集成预报，并对四种集成方法得到的预报结果从多方面进行了对比评估；在此基础上用最优的改进超级集成方法对 48 h 和 72 h 预报结果进行集成，并与最优单模式的结果进行比较；最后对改进的超级集成方法中最佳试预测期进行了调试。得到以下结论：

（1）从单模式对比评估结果来看，CAMx 模式与 CMAQ 模式误差小于 NAQPMS 模式与 WRFCHEM 模式，前两者相较而言，CAMx 模式更优；前两者在冷季存在高估，后两者在冷暖两季均存在低估，且暖季更为明显。模式对氮氧化物预报偏差是导致 O_3 预报误差的重要原因之一。

（2）改进的超级集成方法（AR-SUP）、超级集成方法（R-SUP）、多模式集成平均方法（EMN）以及消除偏差的集成平均方法（BREM）均能在一定程度上减小单模式的系统误差，但效果有差异。EMN 方法为多个单模式的算术平均，实际也包含对单模式系统误差的算术平均。在时间上表现为暖季预报偏低、冷季部分时段预报偏高，在空间上表现为东部低估、西部高估。BREM 方法在多模式集成平均的基础上引入观测平均态，预报效果优于 EMN 方法，其在冷季的高估现象不明显，暖季低估的现象也得到一定的改善，低估的空间范围也有所减小。相较于 EMN 和 BREM 方法，R-SUP 方法能够更为显著的减小系统误差和预报偏差，但仍存在观测高浓度时预报偏低，低浓度偏高的现象，且该方法在减小误差的同时损失了局部相关性，这在冷季时尤为显著。AR-SUP 方法通过活动区间与试预测期配合来合理选取观测平均态的取值范围，不仅能够在 R-SUP 方法基础上进一步逼近观测，且在一定程度上改善了 R-SUP 方法相关性较低的局限性。

（3）和最优单模式相比，AR-SUP 方法减小了误差和偏差、提高了相关性和一致性，且 48 h 和 72 h 预报和 24 h 预报相比，误差并无明显增长，预报技巧稳定在较高水平。进一步说明了该方法的稳健性。

（4）运用 AR-SUP 方法，试预测期也是影响预报效果的重要因子。通过对 AR-SUP 方法试预测期的调试，发现试预测期过短，会导致预报的不确定性增大，预报技巧降低；而试预测期过长则会降低运算效率。综合考虑后将试预测期定为 4 天时，预报效果最佳。

（5）本文中集成方法仅在选定的离散站点中应用，在以后的研究中将进一步应用到所有站点和连续网格点中。

参考文献（References）

崔慧慧, 智协飞. 2013. 基于 TIGGE 资料的地面气温延伸期多模式集成预报 [J]. 大气科学学报, 36: 165–173. Cui H H, Zhi X F. 2013. Multi-model ensemble forecasts of surface air temperature in the extended range using the TIGGE dataset [J]. Transactions of Atmospheric Sciences, 36: 165–173, doi:10.13878/j.cnki.dqkxxb.2013.02.006.

Danard M B, Holl M M, Clark J R. 1968. Fields by correlation assembly—A numerical analysis technique [J]. Mon. Wea. Rev., 96: 141–149, doi:10.1175/1520-0493(1968)096<0141:FBCAAN>2.0.CO;2.

Dennis R L, Byun D W, Novak J H, et al. 1996. The next generation of integrated air quality modeling: EPA’s Models 3 [J]. Atmos. Environ., 30: 1925–1938, doi:10.1016/1352-2310(95)00174-3.

Eder B, Kang D W, Mathur R, et al. 2009. A performance evaluation of the national air quality forecast capability for the summer of 2007 [J]. Atmos. Environ., 43: 2312–2320, doi:10.1016/j.atmosenv.2009.01.033.

Park T. 1998. EPA-454/R-98-017 Guideline on data handling conventions

for the 8-hour ozone NAAQS [S]. United States Environmental Protection Agency's Office of Air Quality Planning and Standards, 4-7.

Gao L N, Jia G S, Zhang R J, et al. 2011. Visual range trends in the Yangtze River Delta region of China, 1981–2005 [J]. Journal of Air & Waste Management Association, 61: 843–849, doi:10.3155/1047-3289.61.8.843.

Grell G A, Peckham S E, Schmitz R, et al. 2005. Fully coupled "online" chemistry within the WRF model [J]. Atmos. Environ., 39: 6957–6975, doi:10.1016/j.atmosenv.2005.04.027.

Haggen-Smit A J. 1952. Chemistry and physiology of Los Angeles smog [J]. Industrial & Engineering Chemistry, 44: 1342–1346, doi:10.1021/ie50510a045.

黄思, 唐晓, 徐文帅, 等. 2015. 利用多模式集合和多元线性回归改进北京 PM_{10} 预报 [J]. 环境科学学报, 35: 56–64. Huang S, Tang X, Xu W S, et al. 2015. Application of ensemble forecast and linear regression method in improving PM_{10} forecast over Beijing areas [J]. Acta Scientiae Circumstantiae, 35: 56–64, doi:10.13671/j.hjkxxb.2014.0728.

Kang D W, Mathur R, Trivikrama Rao S. 2010. Real-time bias-adjusted O_3 and $PM_{2.5}$ air quality index forecasts and their performance evaluations over the continental United States [J]. Atmos. Environ., 44: 2203–2212, doi:10.1016/j.atmosenv.2010.03.017.

Krishnamurti T N, Kishtawal C M, LaRow T E, et al. 1999. Improved weather and seasonal climate forecasts from multi-model super ensemble [J]. Science, 285: 1548–1550, doi:10.1126/science.285.5433.1548.

刘建国, 谢正辉, 赵琳娜, 等. 2013. 基于TIGGE多模式集合的24小时气温BMA概率预报 [J]. 大气科学, 37: 43–53. Liu J G, Xie Z H, Zhao L N, et al. 2013. BMA probabilistic forecasting for the 24-h TIGGE multi-model ensemble forecasts of surface air temperature [J]. Chinese Journal of Atmospheric Sciences, 37: 43–53, doi:10.3878/j.issn.1006-9895.2012.11232.

Mathur R, Yu S C, Kang D W, et al. 2008. Assessment of the wintertime performance of developmental particulate matter forecasts with the Eta-Community Multiscale Air Quality Modeling System [J]. J. Geophys. Res., 113: D02303, doi:10.1029/2007JD008580.

唐晓, 朱江, 王自发, 等. 2013. 基于集合卡尔曼滤波的区域臭氧资料同化试验 [J]. 环境科学学报, 33: 796–805. Tang X, Zhu J, Wang Z F, et al. 2013. Regional ozone data assimilation experiment based on ensemble Kalman filter [J]. Acta Scientiae Circumstantiae, 33: 796–805, doi:10.13671/j.hjkxxb.2013.03.026.

Thompson P D. 1977. How to improve accuracy by combining independent forecasts [J]. Mon. Wea. Rev., 105: 228–229, doi:10.1175/1520-0493(1977)105<0228:HTIABC>2.0.CO;2.

王自发, 谢付莹, 王喜全, 等. 2006. 嵌套网格空气质量预报模式系统的发展与应用 [J]. 大气科学, 30: 778–790. Wang Z F, Xie F Y, Wang X Q, et al. 2006. Development and application of nested air quality prediction modeling system [J]. Chinese Journal of Atmospheric Sciences, 30: 778–790, doi:10.3878/j.issn.1006-9895.2006.05.07.

王自发, 吴其重, Gbaguidi A, 等. 2009. 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), 1: 19–26. Wang Z F, Wu Q Z, Gbaguidi A, et al. 2009. Ensemble air quality multi-model forecast system for Beijing (EMS–Beijing): Model description and preliminary application [J]. Journal of Nanjing University of Information Science & Technology (Nature Science Edition), 1: 19–26, doi:10.13878/j.cnki.jnuist.2009.01.003.

杨学胜. 2001. 业务集合预报系统的现状及展望 [J]. 气象, 27 (6): 3–9. Yang X S. 2001. The new development and the outlook of the operational ensemble prediction system [J]. Meteorological Monthly, 27 (6): 3–9, doi:10.7519/j.issn.1000-0526.2001.6.001.

殷永泉, 李昌梅, 马桂霞, 等. 2004. 城市臭氧浓度分布特征 [J]. 环境科学, 25 (6): 16–20. Yin Y Q, Li C M, Ma G X, et al. 2004. Ozone concentration distribution of urban [J]. Environmental Science, 25 (6): 16–20, doi:10.13227/j.hjkx.2004.06.004.

张涵斌, 智协飞, 王亚男, 等. 2015. 基于 TIGGE 资料的西太平洋热带气旋多模式集成预报方法比较 [J]. 气象, 41: 1058–1067. Zhang H B, Zhi X F, Wang Y N, et al. 2015. Comparison of multi-model ensemble methods for western Pacific tropical cyclone forecast based on TIGGE data [J]. Meteorological Monthly, 41: 1058–1067, doi:10.7519/j.issn.1000-0526.2015.09.002.

Zhang Y H, Huang W, London S J, et al. 2006. Ozone and daily mortality in Shanghai, China [J]. Environmental Health Perspectives, 114: 1227–1232, doi:10.1289/ehp.9014.

智协飞, 林春泽, 白永清, 等. 2009. 北半球中纬度地区地面气温的超级集合预报 [J]. 气象科学, 29: 569–574. Zhi X F, Lin C Z, Bai Y Q, et al. 2009. Super ensemble forecasts of the surface temperature in Northern Hemisphere middle latitudes [J]. Scientia Meteorologica Sinica, 29: 569–574, doi:10.3969/j.issn.1009-0827.2009.05.001.

智协飞, 彭婷, 李刚, 等. 2014. 多模式集成的概率天气预报和气候预测研究进展 [J]. 大气科学学报, 37: 248–256. Zhi X F, Peng T, Li G, et al. 2014. Advances in multi-model ensemble probabilistic prediction [J]. Transactions of Atmospheric Sciences, 37: 248–256, doi:10.13878/j.cnki.dqkxxb.2014.02.013.

智协飞, 季晓东, 张璟,等. 基于 TIGGE 资料的地面气温和降水的多模式集成预报 [J]. 大气科学学报, 2013, 36 (3): 257–266. Zhi X F, Ji X D, Zhang J, et al. Multi-model ensemble forecasts of surface air temperature and precipitation using TIGGE datasets[J]. Transactions of Atmospheric Sciences, 36 (3): 257–266.doi: 10.3969/j.issn.1674-7097.2013.03.001.

智协飞, 张璟, 段晚锁. 2015. ENSO 预测的目标观测敏感区在热带太平洋海温的多模式集合预报中的应用 [J]. 大气科学, 39: 767–776. Zhi X F, Zhang J, Duan W S. 2015. Application of sensitive area for target observation associated with El Niño–Southern Oscillation predictions to multi-model ensemble forecast of the tropical Pacific sea surface temperature [J]. Chinese Journal of Atmospheric Sciences, 39: 767–776, doi:10.3878/j.issn.1006-9895.1408.14181.

智协飞, 赵欢, 朱寿鹏, 等. 2016. 基于CMIP5多模式回报资料的地面气温超级集合研究 [J]. 大气科学学报, 39: 64–71. Zhi X F, Zhao H, Zhu S P, et al. 2016. Super ensemble hindcast of surface air temperature using CMIP5 multi-model data [J]. Transactions of Atmospheric Sciences, 39: 64–71, doi:10.13878/j.cnki.dqkxxb.20150109001.

中华人民共和国环境保护部. 2016. HJ 633-2012 环境空气质量指数(AQI)技术规定(试行) [S]. 北京: 中国环境科学出版社, 2–3. Ministry of Environmental Protection of the People's Republic of China. 2012. HJ 633-2012 Technical Regulation on Ambient Air Quality Index (on trial) [S]. Beijing: China Environmental Science Press, 2–3.