

张天航,迟茜元,张碧辉,等,2020. 全国网格化多模式集成空气质量预报的初步建立[J]. 气象,46(3):381-392. Zhang T H, Chi X Y, Zhang B H, et al, 2020. Development of gridding multi-model ensemble air quality forecast in China[J]. Meteor Mon, 46(3):381-392(in Chinese).

全国网格化多模式集成空气质量预报的初步建立^{*}

张天航 迟茜元 张碧辉 张恒德 江 琪 王继康
饶晓琴 谢 超 吕梦瑶 安林昌 南 洋

国家气象中心,北京 100081

提 要:为降低单个模式预报的不确定性和提高多模式集成空气质量预报系统的精细化程度,利用 Cressman 插值初步建立了我国 $0.25^{\circ}\times 0.25^{\circ}$ 网格化污染物实况。结合 4 套空气质量数值预报模式,通过均值集成、权重集成和多元线性回归集成分别逐格点建立了集成预报。在预报当天各单一模式和集成方法前 50 d 预报效果评估基础上,建立了最优集成预报。对 2018 年 12 月 19—22 日一次重污染过程中集成预报的 $PM_{2.5}$ 浓度评估结果显示:在污染较重时刻,最优集成预报与观测之间的归一化平均偏差(NMB)值在重污染地区保持在 $-20\%\sim 40\%$,对污染程度为良及以上区域的预报范围相较于单个模式更接近观测。整个过程中,最优集成在大部分污染区域与观测之间的 NMB 值为 $-20\%\sim 20\%$,均方根误差(RMSE)值为 $35\sim 75\mu g\cdot m^{-3}$,相关系数(R)值大于 0.4。相较于所有单一模式和其他集成方法,最优集成在全国最多的格点有着较高的总体评分。在污染最重区域的 8 个城市,最优集成预报的污染过程平均开始和结束时间分别比观测时间早 1.8 和 6.9 h。未来需融合卫星反演和地表观测来提高网格化污染物实况的精细化程度,利用降尺度、主客观融合和滚动订正等方法进一步提高网格化多模式集成空气质量预报的准确率。

关键词: 网格化,多模式集成,空气质量预报

中图分类号: P456

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2020.03.009

Development of Gridding Multi-Model Ensemble Air Quality Forecast in China

ZHANG Tianhang CHI Xiyuan ZHANG Bihui ZHANG Hengde JIANG Qi WANG Jikang
RAO Xiaolin XIE Chao LYU Mengyao AN Linchang NAN Yang

National Meteorological Centre, Beijing 100081

Abstract: To decrease the forecast uncertainties of single models and improve the refinement of multi-model ensemble air quality forecast system, the gridding observed pollutant concentration with resolution of $0.25^{\circ}\times 0.25^{\circ}$ was firstly established by using Cressman interpolation method. Then, combined with four numerical air quality forecast models, the mean, weighted and multiple linear regression ensembles were established in each grid, respectively. Finally, based on the evaluation results of single models and ensemble methods in previous 50 days, an optimal ensemble was established. The evaluation results of $PM_{2.5}$ concentrations during a heavy pollution process in 19—22 December 2018 showed that in the case of heavy pollution, the NMB values between the optimal ensemble forecast and observations could also be maintained between -20% and 40% . And the forecast coverage area with good and above pollution by the optimal ensemble was closer to observation than those of single models. During the whole process, the

^{*} 国家重点研发计划(2016YFC0203301)、中国气象局气象预报业务关键技术的发展专项(YBGJXM2018-7A)和国家气象中心青年基金(Q201808)共同资助

2019 年 4 月 29 日收稿; 2020 年 1 月 12 日收修定稿

第一作者:张天航,主要从事空气质量预报和检验研究. Email:sharp@mail.iap.ac.cn

通讯作者:张碧辉,主要从事环境气象预报和检验研究. Email:bihui_zhang@qq.com

NMB, root mean squared error (RMSE) and R values between forecasted $\text{PM}_{2.5}$ concentrations by the optimal ensemble and observation were from -20% to 20% , from 35 to $75 \mu\text{g} \cdot \text{m}^{-3}$ and higher than 0.4, respectively, in most polluted areas. Among all single models and ensemble methods, number of grids over China with high total scores was the largest in optimal ensemble. In the eight cities located in the most polluted region, the average onset and end times of the pollution process by optimal ensemble forecast was 1.8 and 6.9 h earlier than observation, respectively. Therefore, we propose that pollutant concentrations retrieval by satellite and surface observation should be fused to improve the refinement of gridding observed pollutant concentrations. And the methods of scale reduction, subjective and objective fusion and rolling correction should be used to further improve the forecast accuracy of gridding multi-model ensemble air quality forecast.

Key words: gridding, multi-model ensemble, air quality forecast

引 言

空气质量的好坏反映了空气污染的程度,与人体健康息息相关(Cheng et al, 2013; 周骥等, 2018)。同时,作为空气质量重要组成的 $\text{PM}_{2.5}$ 和 PM_{10} 浓度是雾、霾和沙尘等天气现象和等级的重要判据(张小曳等, 2006; 吴兑等, 2010)。当前,多数机构对外发布的空气质量预报的时空分辨率较低,无法满足人们对于精细化定制服务的需求。因此,发展更准确、更高时空分辨率的网格化空气质量预报产品是国家级环境气象业务单位的发展趋势。

空气质量数值预报模式可以提供网格化空气质量预报,但由于现阶段模式中初始场、驱动气象场、气溶胶的排放和生命周期中物理化学过程的不确定性等因素,单个模式预报的污染物浓度和观测之间存在较大误差。相关的评估结果显示,我国各级气象局发展的空气质量数值预报系统预报的 $\text{PM}_{2.5}$ 浓度和观测之间的偏差为 $-30.2\% \sim 49\%$ (赵秀娟等, 2016; 杨关盈等, 2017; Zhou et al, 2017; 刘琳等, 2018)。

多模式集成是降低单个模式预报不确定性的有效手段。瞿元昊等(2018)基于中国和欧洲 7 个空气质量数值预报模式提出了优化的权重集成预报。相较于预报效果最好的单个模式,其预报的 2016 年 6 月至 2017 年 5 月上海市日均 $\text{PM}_{2.5}$ 浓度的污染的预兆评分(TS)提高了 0.28。张天航等(2019b)将多种集成方法相结合,提出了一种最优集成预报方法,其预报的 2015—2016 年全国 $\text{PM}_{2.5}$ 浓度和观测

之间的归一化平均偏差(NMB)在 $\pm 10\%$ 之内。此外,也有一些研究将多模式集成方法应用至 O_3 预报中。吴剑斌等(2017)利用和瞿元昊等(2018)类似的方法对 2015 年 9 月至 2016 年 5 月全国 1529 个国家级环境监测站点的 O_3 浓度进行了多模式集成预报。结果显示集成预报的 O_3 浓度和观测之间的相关系数相较于单个模式提高 0.11 以上。姚雪峰等(2018)将基于活动区间的多模式超级集成应用至 2015 年长三角 26 个站点 O_3 预报中。结果表明,与均值集成、消除偏差的均值集成和滑动训练期的超级集成相比,该集成方法使预报和观测之间的均方根误差(RMSE)降低了 $20\% \sim 23\%$ 。现阶段由于网格化污染物实况匮乏和计算能力的限制,多模式集成空气质量预报多针对站点进行。

网格化污染物实况是网格化多模式集成空气质量预报的基础。利用卫星反演的气溶胶特性来估算地面颗粒物实况已成为未来的发展趋势(Ma et al, 2016; Zhang and Li, 2015; Fang et al, 2016)。但由于估算方法中关键变量的不确定性较大,较难获取夜间浓度,现阶段利用其形成网格化颗粒物实况业务产品的难度较高。此外,利用卫星反演资料估算地表气体浓度的研究也较少。那么在网格化空气质量集成预报发展的初期,将地面观测的站点污染物浓度插值为网格化实况不失为一种高效的方法。

所以,本文利用 Cressman 插值方法,初步建立了网格化污染物实况。再利用多种集成方法分别建立了集成预报。最终在实时预报效果评估基础上,初步得到了多模式最优集成空气质量预报。旨在探索多模式集成预报方法在网格预报上的可行性,为

未来精细化环境气象智能网格预报打下基础。

1 资料介绍

1.1 网格化污染物实况

网格化污染物实况是网格化多模式集成空气质量预报的基础,但目前没有形成一套标准的方法来制作网格化实况。基于后续网格化多模式集成计算机时的考虑,本文利用常用的 Cressman 插值方法,将中国环境监测总站发布的全国约 1500 多个站点观测的 6 种污染物($\text{PM}_{2.5}$ 、 PM_{10} 、 O_3 、 SO_2 、 NO_2 和 CO)逐小时浓度插值为 $0.25^\circ \times 0.25^\circ$ 网格化实况。基于保证插值之后全国范围无缺省值的原则,将插值半径设置为 7.5° 、 5° 、 3° 、 1° 和 0.25° 。图 1 显示了重污染时刻站点观测和插值的网格化 $\text{PM}_{2.5}$ 质量浓度。可以看出插值的网格化 $\text{PM}_{2.5}$ 质量浓度从空间分布和量级上均反映出了站点观测在汾渭平原、河北南部、河南、山东西部、湖北北部、四川盆地、吉林北部等地的中度及以上污染。但网格化实况对重污染区个别数值偏低的站点 $\text{PM}_{2.5}$ 质量浓度高估程度较重,例如河南北部个别站点观测浓度为 $0 \sim 35 \mu\text{g} \cdot \text{m}^{-3}$,插值后为 $150 \sim 250 \mu\text{g} \cdot \text{m}^{-3}$,精细化程度有待改进。从全国站点插值和观测的散点图(图 2)可以看出,绝大多数站点的插值和直接观测的 $\text{PM}_{2.5}$ 浓度差异较小,但由于插值精细化程度问题,

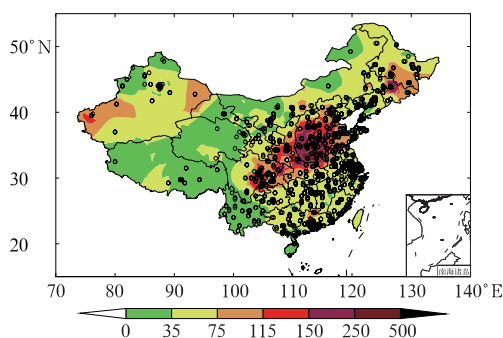


图 1 2018 年 12 月 21 日 01 时插值的网格化(填色图)和站点观测(圆形点)的 $\text{PM}_{2.5}$ 质量浓度(单位: $\mu\text{g} \cdot \text{m}^{-3}$)

Fig. 1 Interpolated gridding (colouring map) and site observed (circular points) $\text{PM}_{2.5}$ mass concentrations (unit: $\mu\text{g} \cdot \text{m}^{-3}$) at 01:00 BT 21 December 2018

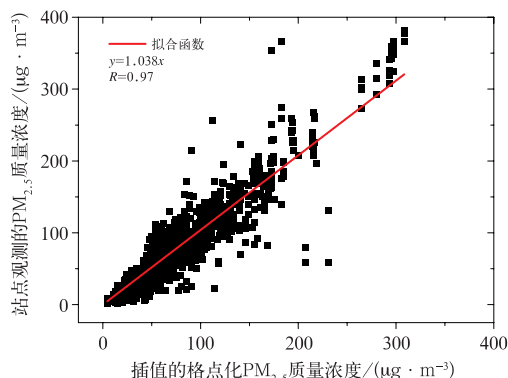


图 2 2018 年 12 月 21 日 01 时插值的网格化和站点观测的 $\text{PM}_{2.5}$ 质量浓度散点图

Fig. 2 Scatter plot of interpolated gridding and site observed $\text{PM}_{2.5}$ mass concentrations at 01:00 BT 21 December 2018

少数站点上存在较大差异。整体来说二者之间相关系数为 0.97,拟合函数斜率为 1.04,说明利用 Cressman 插值可以初步提供网格化污染物实况。

1.2 空气质量数值预报模式

本文所用空气质量数值模式预报资料分别来自中国气象局 4 个国家级或区域级环境气象业务中心,包括国家级中国化学天气预测平台系统(CUACE)、华北区域环境气象预报系统(BREMPS)、华东区域大气环境模式系统(RAEMS)、华南区域中心发展的由全球以及区域同化和预报增强系统(GRAPES)驱动的交互式多尺度空气质量模式(CMAQ),以下分别简称为 CUACE、NNC、ENC 和 SNC 模式。各模式预报区域和各参数设置见表 1。本文选用 2018 年 12 月 19—22 日过程期间各模式 0~72 h 时效预报资料。

为了和网格化实况空间分辨率保持一致,利用双线性插值将所有数值模式预报结果插值为 $0.25^\circ \times 0.25^\circ$ 分辨率。

2 集成方法

针对每种污染物、每个格点和每个预报时次,整体集成方法分为三个阶段(图 3):(1)多方法集成期。利用预报时刻前 50 d 观测资料和各模式预报

表 1 各空气质量数值预报模式参数

Table 1 Parameters of numerical air quality forecast models

模式	预报区域	水平分辨率/km	预报时效/h	排放清单(张恒德等,2018)
CUACE	15.00°~53.85°N、70.00°~134.35°E	15	216	HTAP(2012) EDGARV4.2(2012)
NNC	30.26°~48.22°N、100.73°~128.82°E	9	96	INDEX-B(2012)
ENC	20.30°~43.62°N、103.30°~132.65°E	6	96	INDEX-B(2012)
SNC	17.04°~33.04°N、101.04°~125.04°E	8	96	INDEX-B(2012)

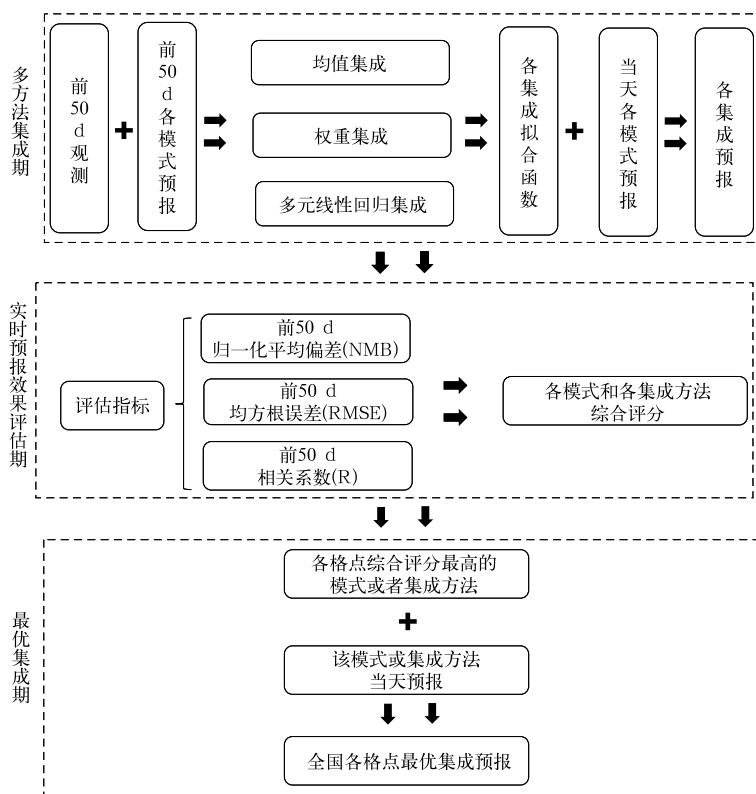


图 3 多模式集成空气质量预报系统流程图

Fig. 3 Flow chart of multi-model ensemble air quality forecast system

资料分别使用均值集成(mean)、权重集成(weighted)和多元线性回归集成(MLR)建立集成预报;(2)实时预报效果评估期。利用预报当天前 50 d 观测资料和各单一模式与集成方法的预报结果,基于 3 种评估指标(NMB、RMSE 和 R),得到各单一模式和集成方法的综合评分;(3)最优集成期。将各格点综合评分最高方法的当天预报结果作为最优集成(best)预报,最后组成全国最优集成预报。

2.1 均值集成

该集成为各单一模式的简单平均,公式如下:

$$F_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n F_i \quad (1)$$

式中: n 为区域模式的数量,取值为 4; i 为第 i 个区域模式; F_{mean} 和 F_i 分别为均值集成和第 i 个模式预报的污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$ 。

2.2 权重集成

权重集成将各单一模式预报时刻前 50 d 预报和观测值间的偏差作为权重系数,公式如下:

$$F_{\text{wighted}} = \sum_{i=1}^n W_i F_i = \sum_{i=1}^n \frac{\frac{1}{|Avg_i - Avg_{\text{obs}}|}}{\sum_{i=1}^n \frac{1}{|Avg_i - Avg_{\text{obs}}|}} \times F_i \quad (2)$$

式中: n 为区域模式的数量,取值为 4; F_{wighted} 为权重集成预报的污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$; W_i 为第 i 个模式的权重系数; Avg_{obs} 和 Avg_i 分别为预报时刻前 50 d 观测值和第 i 个模式预报的 $\text{PM}_{2.5}$ 质量浓度均值,单位: $\mu\text{g} \cdot \text{m}^{-3}$ 。

2.3 多元线性回归集成

多元线性回归集成中,观测值被设置为因变量,各单一模式预报值被设置为自变量,预报时刻前 50 d 样本用于建模,公式如下:

$$F_{\text{MLR}} = \sum_{i=1}^n a_i F_i + a_0 \quad (3)$$

式中: n 为区域模式的数量,取值为 4; F_{MLR} 为多元线性回归集成预报的污染物浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$; a_i 为训练阶段回归求解的第 i 个模式的权重系数; a_0 为常数项。

2.4 最优集成

将最优集成作为本集成系统最终出口的目的是为预报员提供唯一稳定客观参考。在各单一模式和集成方法短期内预报效果波动不大的前提下,对其预报当天前 50 天预报效果进行实时评估,将综合评分最高的方法作为当天的最优集成预报。所采用的评估指标包括 $|NMB|$ 、 $RMSE$ 和 R ,公式如下:

$$|NMB_j| = \left| \frac{\sum_{k=1}^m (F_{j,k} - OBS_k)}{\sum_{k=1}^m OBS_k} \times 100\% \right| \quad (4)$$

$$RMSE_j = \sqrt{\frac{1}{m} \sum_{k=1}^m (F_{j,k} - OBS_k)^2} \quad (5)$$

$$R_j = \frac{\sum_{k=1}^m (F_{j,k} - \bar{F})(OBS_k - \overline{OBS})}{\sqrt{\sum_{k=1}^m (F_{j,k} - \bar{F})^2 (OBS_k - \overline{OBS})^2}} \quad (6)$$

式中: j 为被评估的第 j 种方法,共 7 种,包括 CUACE、NNC、ENC、SNC、Mean、Weighted 和 MLR; m 为评估所用历史数据天数,取值为 50 d, k 为第 k 天; $F_{j,k}$ 为第 j 种方法第 k 天预报的污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$; OBS_k 为第 k 天观测的污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$; $\bar{F}$ 和 $\overline{OBS}$ 分别为预报和观测的平均污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$ 。

针对每个评估指标,把所有 7 种方法按照指标好坏程度排序,最差至最好分别得 0~6 分,如 $|NMB|$ 越小表示偏差越小,预报效果也越好,所以将所有方法中 $|NMB|$ 最小的记为 6 分, $|NMB|$ 最大的记为 0 分。每种方法各指标得分如下:

$$Point_{|NMB_j|} = \text{Sort}(|NMB_{j=1,2,\dots,7}|)_j \quad (7)$$

$$Point_{RMSE_j} = \text{Sort}(RMSE_{j=1,2,\dots,7})_j \quad (8)$$

$$Point_{R_j} = \text{Sort}(R_{j=1,2,\dots,7})_j \quad (9)$$

式中: Sort 和 $Point$ 分别为各方法在每个评估指标的排位和得分。

将所有单一模式和集成方法的 3 个评估指标得分相加,选择总体评分最高方法作为当天最优集成预报,其计算方法如下:

$$Point_{\text{total}_j} = Point_{|NMB_j|} + Point_{RMSE_j} + Point_{R_j} \quad (10)$$

$$F_{\text{Best}} = F_{\text{Max}(Point_{\text{total}_j=1,2,\dots,7})} \quad (11)$$

式中: $Point_{\text{total}_j}$ 为第 j 种方法的总体评分; F_{Best} 为最优集成预报的污染物质量浓度,单位: $\mu\text{g} \cdot \text{m}^{-3}$; Max 是取所有方法总体评分最高的函数。

3 效果评估

张天航等(2019b)已评估了类似多模式集成方法的长期表现,针对 2018 年 12 月 19—22 日发生在华北中南部、汾渭平原和四川盆地等地的一次重污染过程评估全国网格化多模式集成空气质量预报的表现,并着重分析过程中主导 AQI 值的 $\text{PM}_{2.5}$ 浓度的预报效果。

3.1 预报当天前 50 d

预报当天前 50 d 预报效果评估是当天最优集成预报的基础。以 2018 年 12 月 19 日 20 时前 50 d 预报效果评估为例(图 4)。在华北地区, CUACE、NNC 和 ENC 模式有预报范围重合区域。ENC 模式预报的 $\text{PM}_{2.5}$ 质量浓度与观测值最为接近, NNC

模式在局部地区高估了一个污染等级,而 CAUCE 模式的高估程度更大、范围更广。在华中和华南地区,CUACE、ENC 和 SNC 模式有预报范围重合区域。ENC 模式的预报效果最佳,SNC 模式对 $\text{PM}_{2.5}$ 质量浓度略有高估,CAUCE 模式对华中地区 $\text{PM}_{2.5}$ 质量浓度高估了 1~2 个污染等级,幅度较大。总的来说,在我国东部地区,CAUCE 模式高估了观测的浓度,ENC 模式则表现良好。由各单一模式和集成方法预报和观测的 $\text{PM}_{2.5}$ 质量浓度均值对比可知,类似 CUACE 模式,Mean 集成较大幅度高估了华北、华中等地的 $\text{PM}_{2.5}$ 质量浓度,高估程度为 1~2 个污染等级。Weighted 集成结果则与 ENC、NNC 模式相近,对我国东部 $\text{PM}_{2.5}$ 浓度量级的预报较好,但对空间分布的预报不够准确。单个模式中仅有 CUACE 模式覆盖到了我国西部地区。在新疆和内蒙古中西部等地,CUACE 模式较大地低估了颗粒物浓度,但在西藏地区与观测结果差异较小。这是由于目前业务运行的 CUACE/Haze-fog 模式并未考虑自然沙尘气溶胶,因此该模式的预报结果在新疆和内蒙古中西部等沙尘源区有所低估,而对于较少受沙尘影响的西藏地区的预报则更加准确。

MLR 集成对我国中东部污染等级为良的预报的空间分布在所有方法中最接近观测值,但对轻度污染的覆盖范围的预报偏小。同时 MLR 集成还预报出了新疆地区良及以上的 $\text{PM}_{2.5}$ 质量浓度,虽与观测值仍有一定差距,但较 CUACE 模式和其他集成方法已大幅改善。

从各方法与观测之间的 $|\text{NMB}|$ 值来看。CUACE 模式和 SNC 模式在华中地区与观测之间的偏差较大, $|\text{NMB}|$ 值在部分地区超过了 80%。Mean 集成和 Weighted 集成的 $|\text{NMB}|$ 在我国中东部地区较为接近,但 Weighted 集成在华中地区偏差小于 Mean 集成。MLR 集成的预报偏差在全国范围内都控制在 0~40% 内,且大部地区要低于 20%,还有效降低了 CUACE 模式、Mean 集成和 Weighted 集成在西部地区的偏差。此外,在西藏地区,MLR 集成相较于 CUACE 模式对预报偏差的改善幅度较大,预报和观测之间的 $|\text{NMB}|$ 值从 CUACE 模式的 80%~100%,下降至了 MLR 集成的 0~40%。

就预报和观测间的 RMSE 值而言,CUACE 模式预报的 RMSE 值最高,在华北、华中及新疆的部

分地区超过了 $115 \mu\text{g} \cdot \text{m}^{-3}$ 。ENC 模式、NNC 模式、SNC 模式、Mean 集成和 Weighted 集成在东部地区的 RMSE 值都小于 $75 \mu\text{g} \cdot \text{m}^{-3}$,但在新疆地区 RMSE 值较高。MLR 集成在全国大部地区的 RMSE 要低于 $35 \mu\text{g} \cdot \text{m}^{-3}$,仅在华北、华中和新疆的少数地区为 $35 \sim 75 \mu\text{g} \cdot \text{m}^{-3}$,是所有方法中最低的。

从预报和观测值间的 R 可以看出,Mean 集成、Weighted 集成和 MLR 集成与观测值间的 R 在华北、华中和华东的大部地区均大于 0.6,表现出较强正相关,在我国其余大部分地区为 0.2~0.6。此外,MLR 集成在西部地区的 R 值要高于其他模式和集成方法。

从各方法的总体评分上来看,CUACE 模式在西部地区评分较高,全部达到了 15 分以上,东部评分整体偏低。ENC 模式在华北、华东、华中部分地区表现良好,评分最高可达 15~16 分。NNC 模式和 SNC 模式整体表现欠佳,达到 15 分以上的区域较少。各集成方法在我国西部地区的评分整体较高,尤其是 MLR 集成在西部绝大多数区域的评分都超过了 18 分,高于其他集成方法。而在东部地区,三种集成方法的总体评分较为接近。但是全国整体来说,集成后的总体评分高于单一模式,说明集成预报是提高单一模式预报准确率的有效手段。

图 5 显示了预报时刻前 50 d,每个评估指标得分最高方法的全国分布,其中每个格点总体评分最高的方法即为该格点当天的最优集成预报方法。从 $|\text{NMB}|$ 指标来看,我国西部和北部大部地区得分最高的方法是 MLR 集成,而在东部评分最高的预报方法则各不相同。针对 R 的得分最高预报方法在西部大部分地区仍是 MLR 集成,少部分地区为 CUACE 模式;在我国东部地区,以 Mean 集成和 ENC 模式为主。RMSE 得分最高的方法在全国大部分地区为 MLR 集成,少部分地区为 Mean 集成和 Weighted 集成。总体评分指标得分最高的方法在我国东部和东北地区以三种集成方法为主,河北中南部以 NNC 模式为主,全国其他区域为 MLR 集成。这说明没有任何一种预报方法在全国任意格点均表现出最好的预报效果。这也是建立最优集成预报的意义所在:能综合各方法在不同格点上的优势,以解决不同地区总体评分最高预报方法有所不同的问题,进一步提高预报准确率。

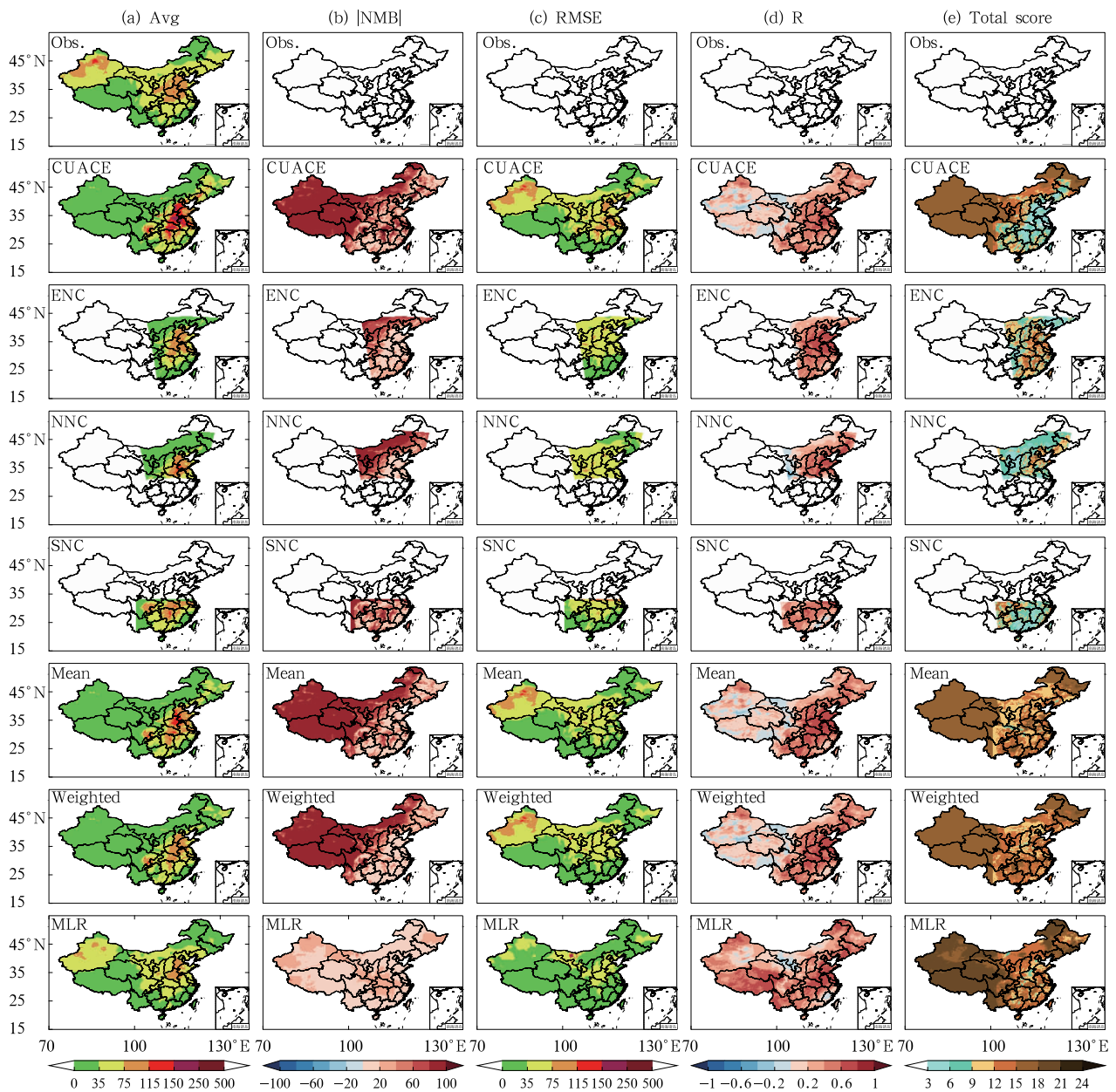


图4 2018年12月19日20时前50 d各模式和各集成预报的 $\text{PM}_{2.5}$ 平均质量浓度(a,单位: $\mu\text{g} \cdot \text{m}^{-3}$)

以及与观测值间的|NMB|(b,单位:%)、RMSE(c,单位: $\mu\text{g} \cdot \text{m}^{-3}$)、R(d)和总体评分(e)

Fig. 4 Observed and forecasted average $\text{PM}_{2.5}$ mass concentrations (a, unit: $\mu\text{g} \cdot \text{m}^{-3}$)

by single models and ensemble methods and the |NMB| (b, unit: %), RMSE (c, unit: $\mu\text{g} \cdot \text{m}^{-3}$), R (d) and total scores between them (e) in previous 50 d before 20:00 BT 19 December 2018

3.2 未来3 d

在3.1节前50 d预报效果评估的基础上,得到了2018年12月19日20时每个格点起报的最优集成预报,本节对比了污染较重时刻和整个污染过程期间各单一模式和各集成方法的预报效果。

3.2.1 污染较重时刻

污染较重时刻的预报效果反映了预报方法对峰值的预报能力。图6展示了2018年12月21日01时观测和各方法预报的 $\text{PM}_{2.5}$ 质量浓度和NMB值。其中,预报值为各方法2018年12月19日20时起报第29小时值。

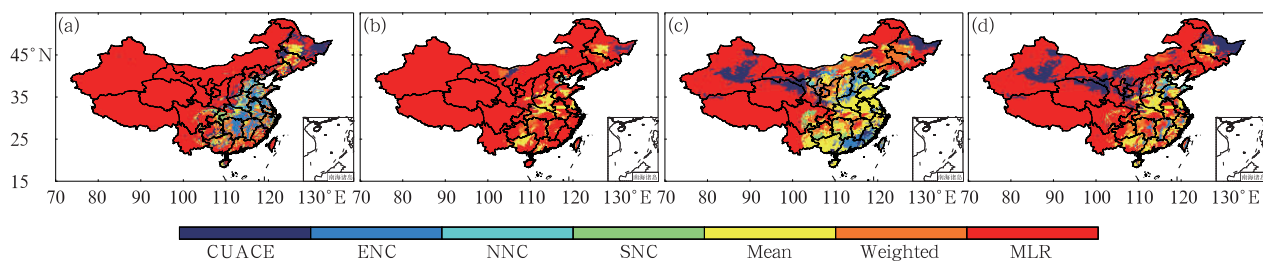


图 5 2018 年 12 月 19 日 20 时前 50 d 每个指标得分最高的方法

(a) $|NMB|$, (b) RMSE, (c) R, (d) 总体评分

Fig. 5 Methods with highest scores at each evaluation index in previous 50 d before 20:00 BT 19 December 2018

(a) $|NMB|$, (b) RMSE, (c) R, (d) total score

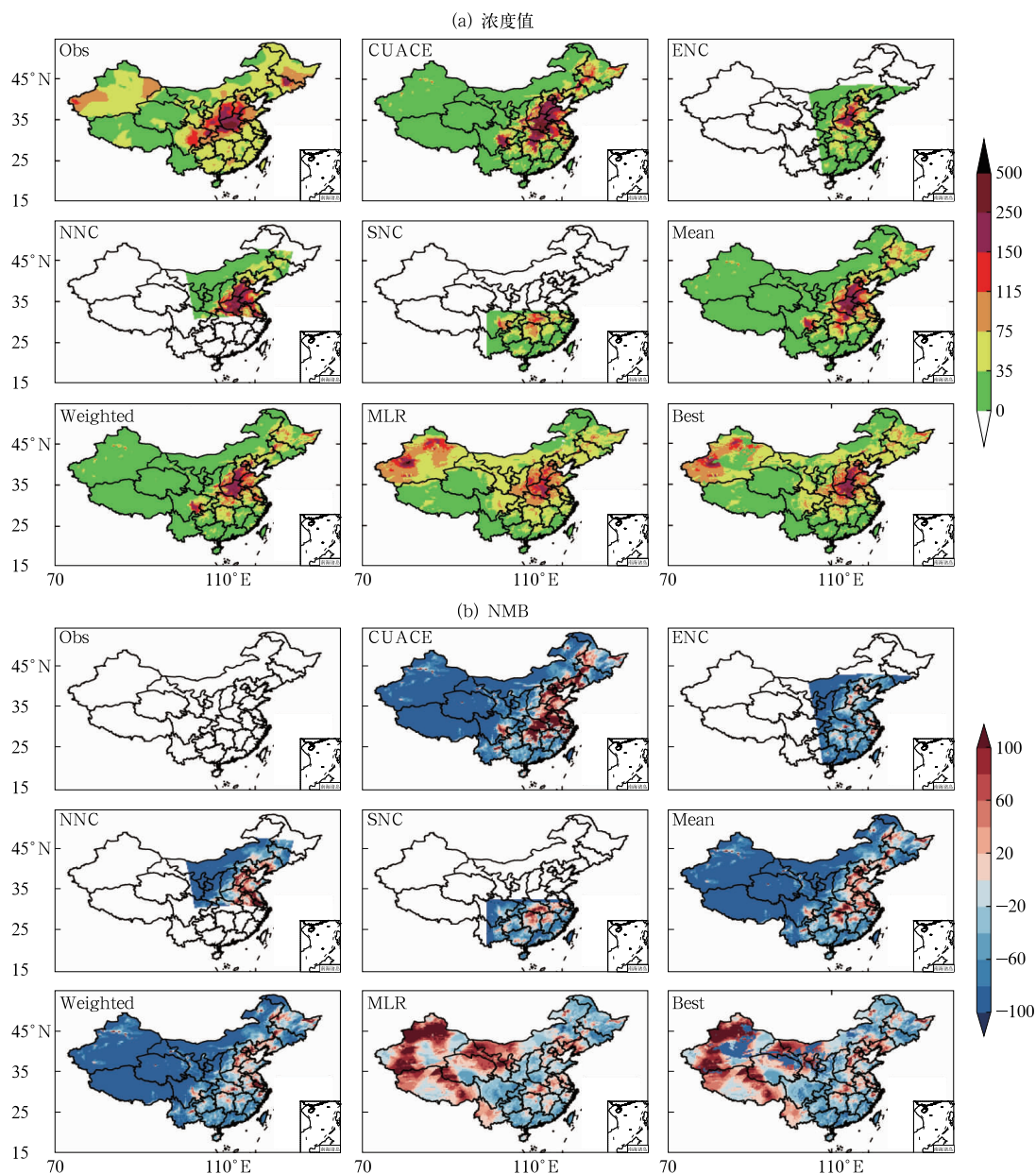


图 6 2018 年 12 月 21 日 01 时观测和各方法预报的 $\text{PM}_{2.5}$ 质量浓度 (a, 单位: $\mu\text{g} \cdot \text{m}^{-3}$) 和 NMB 值 (b, 单位: %)

Fig. 6 Observed and forecasted $\text{PM}_{2.5}$ mass concentrations (a, unit: $\mu\text{g} \cdot \text{m}^{-3}$) by single models and ensemble methods and the NMB (b, unit: %) between them at 01:00 BT 21 December 2018

CAUCE 模式较大高估了华北北部、四川盆地和华中地区的 $\text{PM}_{2.5}$ 质量浓度,且对轻度及以上污染范围预报偏大,与观测值间的 NMB 值在华北、华东和华中地区为 $-40\% \sim 80\%$,且没有预报出新疆地区的污染情况。NNC 模式对华北地区的预报效果与 CAUCE 模式接近,但高估程度不如 CUACE 模式严重;而 ENC 模式则与 CUACE 模式相反,即对污染区域 $\text{PM}_{2.5}$ 质量浓度的预报值偏低,NMB 值在 $-60\% \sim 0\%$,污染区域预报偏小。SNC 模式对 $\text{PM}_{2.5}$ 浓度的预报在华中部分地区偏高,NMB 值最高达到 60% ,而在华南地区则整体低估。总的来说,在该时刻污染最重的华北地区,CAUCE 模式和 NNC 模式对污染均有不同程度的高估,而 ENC 模式的预报浓度则偏低。

由于 Mean 集成和 Weighted 集成的预报效果对单个模式的表现依赖程度较高,虽然较单个模式对污染的程度和范围有所改善,但仍在局地较大高估了观测值(NMB 值高于 60%)。在华北和华中地区,MLR 集成和 Best 集成与观测之间的 NMB 值大部分在 $-20\% \sim 20\%$,少数局地在 $20\% \sim 40\%$,相较于单个模式以及 Mean 集成和 Weighted 集成有较大幅度改善。对污染程度为良($\text{PM}_{2.5}$ 质量浓度为 $35 \sim 75 \mu\text{g} \cdot \text{m}^{-3}$)区域的预报范围更大,更接近观测值。但 MLR 集成预报的华北地区 $\text{PM}_{2.5}$ 质量浓度较实况略有偏低,不及 Best 集成效果好。此外,虽然 MLR 集成和 Best 集成预报的新疆地区污染程度为良的区域范围接近观测,但却高估了天山北部和南疆盆地西部的 $\text{PM}_{2.5}$ 浓度。

3.2.2 3 d 整体

本次重污染过程主要持续时间为 12 月 19—22 日。本节评估了 12 月 19 日 20 时起报的未来 3 d (重污染主要持续时段)各方法的整体预报效果(图 7)。从 3 d 均值来看,各模式均能预报出大部分 $\text{PM}_{2.5}$ 浓度高值区,主要为京津冀中南部、汾渭平原、四川盆地和湖北西北部等地。其中,CAUCE 模式高估了华北北部、华中和华东部分地区的 $\text{PM}_{2.5}$ 浓度,低估了华南地区的 $\text{PM}_{2.5}$ 浓度,对四川盆地高 $\text{PM}_{2.5}$ 浓度的预报值偏高,但污染范围偏小。ENC 模式和 NNC 模式分别低估和高估了华北地区的 $\text{PM}_{2.5}$ 浓度,SNC 模式高估了华中地区的 $\text{PM}_{2.5}$ 浓度。总的来说,在华北地区,CAUCE 模式和 NNC 模式高估了 $\text{PM}_{2.5}$ 浓度,而在华中地区,CAUCE 模式和 SNC 模式存在类似的高估情况。相反,ENC 模式对华北和华中地区的 $\text{PM}_{2.5}$ 浓度有所低估。而

在我国西部地区,CAUCE 模式仍低估了 1~2 个污染等级。Mean 集成和 Weighted 集成较单个模式在我国东部地区的预报结果更加接近于观测值。这两种集成预报结果均高估了华北地区的 $\text{PM}_{2.5}$ 浓度,在四川盆地的预报效果则与 CUACE 模式类似,说明这两种集成方法较依赖于单个模式,对预报效果的改善能力有限。MLR 集成预报相较于单个模式、Mean 集成和 Weighted 集成,在东部地区预报效果改善幅度较大,西部地区的低估程度有所改善,但在部分地区存在一定程度的高估。最优集成结合了各单一模式和集成方法在不同站点上的优势,在本次污染最重的河南预报的 $\text{PM}_{2.5}$ 浓度相对于 MLR 集成更高,也更加接近于观测,说明对峰值的预报能力有了进一步的提高。

从各方法预报和观测之间 NMB 值来看,在华北中南部地区,CUACE 模式高估了观测值的 $20\% \sim 80\%$,ENC 模式的 NMB 值为 $-40\% \sim 20\%$,NNC 模式的 NMB 值为 $-40\% \sim 60\%$,ENC 模式与观测值最为接近。在长三角地区,CAUCE 模式和 NNC 模式较大高估了观测值,ENC 模式的预报效果更为接近观测值。在四川盆地,SNC 模式预报和观测值之间的 NMB 值为 $-40\% \sim 20\%$,优于 CAUCE 模式。从集成结果来看,Mean 集成和 Weighted 集成预报和观测值间的 NMB 值在大部分区域较为接近,但在华中地区 Mean 集成的 NMB 值为 $-20\% \sim 60\%$,Weighted 集成的 NMB 值为 $-20\% \sim 20\%$,Weighted 集成预报效果有所改善。MLR 集成和 Best 集成的预报效果改善幅度最大,NMB 值在我国东部大部区域为 $-20\% \sim 20\%$,但在新疆北部地区存在较大高估。

就预报和观测值间的 RMSE 而言,CAUCE 模式和 ENC 模式的 RMSE 在华北南部和华中部分地区出现了超过 $115 \mu\text{g} \cdot \text{m}^{-3}$ 的高值,同样 NNC 模式在华北南部的 RMSE 也相对偏高。SNC 模式的 RMSE 大部分均低于 $75 \mu\text{g} \cdot \text{m}^{-3}$,低于其他单个模式。Mean 集成和 Weighted 集成的 RMSE 在华北南部、新疆部分地区呈高于 $75 \mu\text{g} \cdot \text{m}^{-3}$ 的高值。MLR 集成和 Best 集成的 RMSE 整体最小,全国大部分地区小于 $35 \mu\text{g} \cdot \text{m}^{-3}$,在本次过程重污染区的大部分地区为 $35 \sim 75 \mu\text{g} \cdot \text{m}^{-3}$,且相较于 Mean 集成和 Weighted 集成,新疆地区的 RMSE 有所降低。

从预报和观测值间的 R 可以看出,MLR 集成和 Best 集成与观测值间的 R 在华北大部分地区均高于 0.6,表现出强相关性,在华中和华南大部分地

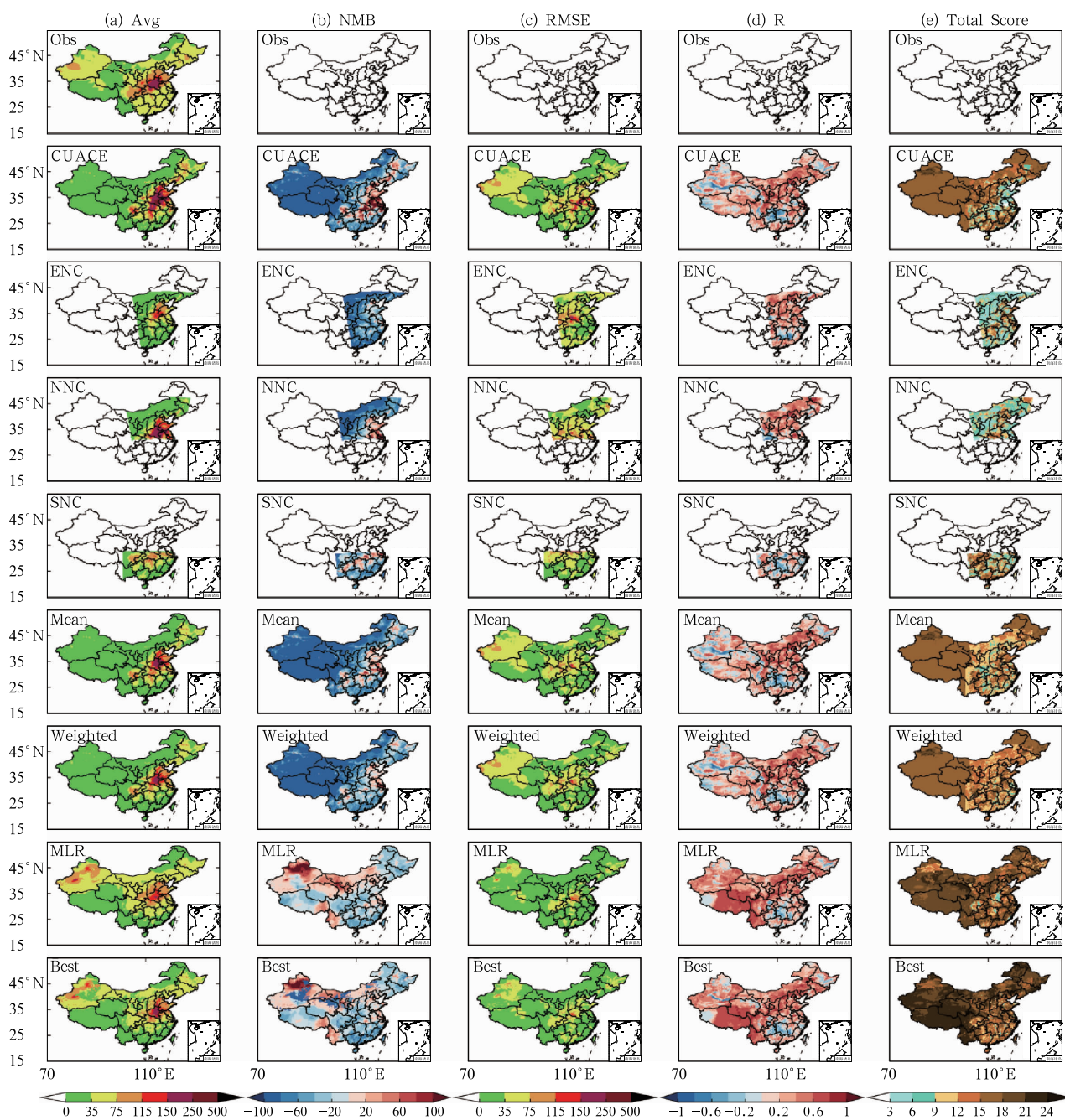


图 7 2018 年 12 月 19 日 20 时起报未来 3 d 各模式和各集成预报的 $\text{PM}_{2.5}$ 平均质量浓度(a, 单位: $\mu\text{g} \cdot \text{m}^{-3}$) 以及观测值间的 NMB(b, 单位: %)、RMSE(c, 单位: $\mu\text{g} \cdot \text{m}^{-3}$)、R(d) 和总体评分(e)

Fig. 7 Observed and forecasted average $\text{PM}_{2.5}$ mass concentrations (a, unit: $\mu\text{g} \cdot \text{m}^{-3}$) by single models and ensemble methods, NMB (b, unit: %), RMSE (c, unit: $\mu\text{g} \cdot \text{m}^{-3}$), R (d) and total score between them (e) for the next 3 days beginning at 20:00 BT 19 December 2018

区为 0.4~0.6。相对于各单一模式和其他集成方法,在全国更多地区表现出强相关性,体现了这两种集成方法对污染物浓度变化趋势较高的预报能力。

总体评分上来说,我国东部大部分地区单个模式的评分要低于各集成模式,而 MLR 集成和 Best 集成评分又显著高于 Mean 集成和 Weighted 集成。

但 MLR 集成在污染最重的河南地区低至 9 分以下,预报效果不佳。总体而言, Best 集成在全国最多的区域有着更高的总体得分,说明最优集成能够有效提升综合预报效果。

综上所述, MLR 集成和 Best 集成能较大幅度改善本次污染过程中 $\text{PM}_{2.5}$ 浓度的预报效果,但

Best 集成在全国更多的区域有着较高的总体评分,且对于污染最重的河南地区预报效果更佳。

由于 $\text{PM}_{2.5}$ 轻度污染的阈值为 $75 \mu\text{g} \cdot \text{m}^{-3}$ 。所以针对某一城市, $\text{PM}_{2.5}$ 浓度大于(小于) $75 \mu\text{g} \cdot \text{m}^{-3}$ 被定义为一个过程的开始(结束)。表 4 展示了本次过程中污染最重区域河北南部和河南北部各城市 Best 集成预报的污染过程开始和结束时间与观测的偏差。可以看出受静稳形势和本地排放影响,该区域大部分城市的过程开始时间在 18 日 19—21 时, Best 集成预报的开始时间比观测早了 0

~4 h。仅在新乡,预报的过程开始时间比观测晚了 1 h。在所有城市,集成预报的平均开始时间比观测仅早了 1.8 h,误差较小。说明集成预报对静稳形势下污染物的累积过程预报较好。受北方冷空气影响,23 日 03—13 时,本次污染过程自北向南逐渐减弱消散。集成预报的过程结束时间比观测早了 3~10 h,平均值为 6.9 h。说明集成预报的本次冷空气到达时间偏早,对环流形势改变引起的污染物浓度变化的预报能力需要进一步提升。

表 2 最优集成预报的各城市污染过程开始和结束时间与观测的偏差
Table 2 Biases between best ensemble forecast and observed onset and end time of the pollution process in each city

城市	开始时间/BT	预报偏差/h	结束时间/BT	预报偏差/h
保定	18 日 00 时	+4	23 日 03 时	+3
石家庄	18 日 19 时	+3	23 日 04 时	+9
邢台	18 日 21 时	+4	23 日 08 时	+10
邯郸	18 日 20 时	+3	23 日 09 时	+10
安阳	18 日 19 时	0	23 日 10 时	+8
鹤壁	18 日 20 时	+1	23 日 10 时	+5
新乡	18 日 19 时	-1	23 日 11 时	+6
郑州	18 日 19 时	0	23 日 13 时	+4
平均	/	+1.8	/	+6.9

注:预报偏差的正值和负值分别表示预报的时间点比观测偏早和偏晚。
Note: The positive and negative values of forecast deviation indicate that the time point of forecast is earlier and later than observation, respectively.

4 结合未来环境智能网格预报的讨论

中国气象局智能网格预报发展规划中对环境气象的要求是建立全国 5 km 分辨率精细化网格雾、霾、沙尘预报。网格化多模式集成空气质量预报是其中的重要环节,在此就其现阶段存在的问题和未来发展计划进行讨论。

(1)目前,通过直接插值得到的污染物实况精细化程度还不能满足要求。未来可通过卫星反演的气溶胶和气体特征估算地表污染物浓度,再融合地面站点污染物浓度观测,最后基于地形和下垫面特征进一步调整,得到更为精准的网格化污染物浓度实况。其中利用卫星资料反演夜间污染物浓度是一大难点。

(2)由于计算机时和次网格技术发展的限制,现在所用各空气质量数值预报模式的空间分辨率均达不到 5 km 的要求,从一定程度上影响了集成预报的准确率。未来可考虑用降尺度手段间接提高现有空气质量数值预报的空间分辨率(曹勇等,2016)。

(3)多元线性回归集成虽然能在一定程度上提

升单一模式对我国西部受沙尘区域污染物浓度的预报能力,但和观测相比还存在较大误差。国家气象中心同时也发展了一套多模式集成沙尘网格预报,未来可考虑将其融合进网格化多模式集成空气质量预报中,以提升其对沙尘过程的预报能力。

(4)目前,国家气象中心每天发布未来 3 d 空气污染气象条件主观预报,其对 2018 年全国轻度及以上污染的 TS 评分为 0.23~0.34(张天航等,2019a)。未来可考虑将其融合进网格化多模式集成空气质量预报中,进一步提高后者的预报效果。

(5)未来可考虑基于网格化污染物实况对多模式集成空气质量预报进行逐小时滚动订正,以提高短临期的预报效果。

5 结 论

对 2018 年 12 月 19—22 日一次重污染过程中网格化多模式集成的表现评估有如下结果。

(1)通过对 19 日 20 时预报前 50 d 各单一模式、Mean 集成、Weighted 集成和 MLR 集成预报效果评估,得到当天最优集成在我国东部和东北地区

格点应以 Mean 集成、Weighted 集成和 MLR 集成为主,在河北中南部格点以 NNC 模式为主,全国其他区域格点为 MLR 集成。

(2)对污染较重时刻(2018年12月21日01时)的评估结果显示,Best 集成相较于单个模式对重污染的预报效果有较大幅度改善。在重污染地区预报与观测之间的 NMB 值大部分在 $-20\% \sim 20\%$,少数局地 $20\% \sim 40\%$,且对污染程度为良及以上区域的预报范围更接近观测值。

(3)对污染过程整体的评估结果显示,Best 集成在所有单一模式和集成方法中的综合预报效果最好,在最多的格点有着较高的总体评分。在大部分污染区域,其与观测之间的 NMB 值为 $-20\% \sim 20\%$,RMSE 为 $35 \sim 75 \mu\text{g} \cdot \text{m}^{-3}$, R 大于 0.4。此外,在污染最重区域的 8 个城市,Best 集成预报的污染过程平均开始和结束时间分别比观测时间早 1.8 和 6.9 h。说明集成预报对静稳形势下污染物的累积过程预报较好,但对环流形势改变引起的污染物浓度变化的预报能力需要进一步提升。

参考文献

- 曹勇,刘凑华,宗志平,等,2016. 国家级格点化定量降水预报系统[J]. 气象,42(12):1476-1482. Cao Y, Liu C H, Zong Z P, et al, 2016. State-level gridded quantitative precipitation forecasting system[J]. Meteor Mon, 42(12):1476-1482(in Chinese).
- 刘琳,白永清,林春泽,等,2018. 华中区域空气质量数值预报系统评估及气溶胶辐射效应的模拟研究[J]. 气象,44(9):1179-1190. Liu L, Bai Y Q, Lin C Z, et al, 2018. Evaluation of regional air quality numerical forecasting system in central China and its application for aerosol radiative effect[J]. Meteor Mon, 44(9):1179-1190(in Chinese).
- 瞿元昊,许建明,Brasseur G,等,2018. 利用多模式最优集成方法预报上海 $\text{PM}_{2.5}$ [J]. 环境科学学报,38(9):3449-3456. Qu Y H, Xu J M, Brasseur G, et al, 2018. Application of OCF on $\text{PM}_{2.5}$ forecast in Shanghai[J]. Acta Sci Circumst, 38(9):3449-3456(in Chinese).
- 吴兑,汤仕文,邓雪娇,等,2010. 霾的观测和预报等级:QX/T 113-2010[S]. 北京:气象出版社. Wu D, Tang S W, Deng X J, et al, 2010. Observation and forecasting levels of haze: QX/T 113-2010[S]. Beijing: China Meteorological Press(in Chinese).
- 吴剑斌,肖林鸿,晏平仲,等,2017. 最优集成方法在城市臭氧数值预报中的应用研究[J]. 中国环境监测,33(4):213-230. Wu J B, Xiao L H, Yan P Z, et al, 2017. Application of optimal consensus forecast in urban ozone prediction[J]. Environ Monit China, 33(4):213-230(in Chinese).
- 杨关盈,邓学良,吴必文,等,2017. 基于 CUACE 模式的合肥地区空气质量预报效果检验[J]. 气象与环境学报,33(1):51-57. Yang G Y, Deng X L, Wu B W, et al, 2017. Verification of CUACE model in Hefei, Anhui Province[J]. J Meteor Environ, 33(1):51-57(in Chinese).
- 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 X F, Ge B Z, Wang Z F, et al, 2018. Application of improved super ensemble forecast method for O_3 and its performance evaluation over the Yangtze River Delta region[J]. Chin J Atmos Sci, 42(6):1273-1285(in Chinese).
- 张恒德,张庭玉,李涛,等,2018. 基于 BP 神经网络的污染物浓度多模式集成预报[J]. 中国环境科学,38(4):1243-1256. Zhang H D, Zhang T Y, Li T, et al, 2018. Forecast of air quality pollutants' concentrations based on BP neural network multi-model ensemble method[J]. China Environ Sci, 38(4):1243-1256(in Chinese).
- 张天航,迟茜元,饶晓琴,等,2019a. 2018 年国家级空气质量主客观预报 TS 评分对比检验[J]. 环境工程技术学报,9(3):213-222. Zhang T H, Chi Q Y, Rao X Q, et al, 2019a. Verification of national subjective and objective air quality forecast in 2018 by TS score[J]. J Environ Eng Technol, 9(3):213-222(in Chinese).
- 张天航,王继康,张恒德,等,2019b. 一种最优多模式集成方法在我国重污染区域 $\text{PM}_{2.5}$ 浓度预报中的应用[J]. 环境工程技术学报,9(5):520-530. Zhang T H, Wang J K, Zhang H D, et al, 2019b. Application of a best multi-model ensemble method in $\text{PM}_{2.5}$ forecast in heavily polluted regions of China[J]. J Environ Eng Technol, 9(5):520-530(in Chinese).
- 张小曳,汤洁,王亚强,等,2006. 沙尘暴天气监测规范:GB/T 20476—2006[S]. 北京:中国标准出版社. Zhang X Y, Tang J, Wang Y Q, et al, 2006. Technical regulations of sand and dust storm monitoring:GB/T 20476—2006[S]. Beijing: China Standards Press(in Chinese).
- 赵秀娟,徐敬,张自银,等,2016. 北京区域环境气象数值预报系统及 $\text{PM}_{2.5}$ 预报检验[J]. 应用气象学报,27(2):160-172. Zhao X J, Xu J, Zhang Z Y, et al, 2016. Beijing regional environmental meteorology prediction system and its performance test of $\text{PM}_{2.5}$ concentration[J]. J Appl Meteor Sci, 27(2):160-172(in Chinese).
- 周骥,孙庆华,许建明,等,2018. 上海地区不同 $\text{PM}_{2.5}$ 污染过程对炎症应激影响的差异性[J]. 气象,44(12):1612-1617. Zhou J, Sun Q H, Xu J M, et al, 2018. Effects of different $\text{PM}_{2.5}$ pollution processes on inflammatory stress in Shanghai Area[J]. Meteor Mon, 44(12):1612-1617(in Chinese).
- Cheng Z, Jiang J K, Fajardo O, et al, 2013. Characteristics and health impacts of particulate matter pollution in China(2001—2011)[J]. Atmos Environ, 65:186-194.
- Fang X, Zou B, Liu X P, et al, 2016. Satellite-based ground $\text{PM}_{2.5}$ estimation using timely structure adaptive modeling[J]. Remote Sens Environ, 186:152-163.
- Ma Z W, Hu X F, Sayer A M, et al, 2016. Satellite-based spatiotemporal trends in $\text{PM}_{2.5}$ concentrations: China, 2004—2013[J]. Environ Health Perspect, 124(2):184-192.
- Zhang Y, Li Z Q, 2015. Remote sensing of atmospheric fine particulate matter ($\text{PM}_{2.5}$) mass concentration near the ground from satellite observation[J]. Remote Sens Environ, 160:252-262.
- Zhou G Q, Xu J M, Xie Y, et al, 2017. Numerical air quality forecasting over eastern China: an operational application of WRF-Chem[J]. Atmos Environ, 153:94-108.