

基于集合预报系统的日最高和最低气温预报^{* 1}

熊敏詮

XIONG Minquan

国家气象中心,北京,100081

National Meteorological Center, Beijing 100081, China

2016-08-25 收稿,2016-12-26 改回.

熊敏詮. 2017. 基于集合预报系统的日最高和最低气温预报. 气象学报, 75(2):211-222

Xiong Minquan. 2017. Calibrating daily 2 m maximum and minimum air temperature forecasts in the ensemble prediction system. *Acta Meteorologica Sinica*, 75(2):211-222

Abstract BP neural network-Self Memory method (BP-SM) is used to calibrate daily 2 m maximum (minimum) air temperature forecasts at 512 stations in China with the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) in March 2016. Seven statistical characteristics used as predictors are calculated based on 2 m air temperature model output of EPS. Daily maximum (minimum) air temperature forecasts by BP-SM, BP and direct model output (DMO) are compared. The post-processing with BP-SM is shown to improve the forecast accuracy. Compared with BP method, more advantages of BP-SM method are attained in longer predictable time. The accurate rate of daily maximum (minimum) air temperature forecasts with absolute errors less than 2°C reaches above 60% and even over 90% at some stations. Compared with DMO, the forecasting skill score of BP-SM is 30% on average, and above 60% over the eastern Tibetan Plateau. This program is obviously superior with forecast errors within 2°C (1°C). The calibrated daily 2 m maximum air temperature is slightly better than the daily 2 m minimum air temperature. By BP-SM method, the systematic deficiencies of daily 2 m maximum (minimum) air temperature forecasts are significantly reduced.

Key words Ensemble prediction system, Daily maximum air temperature, Daily minimum air temperature, BP neural network-Self Memory method

摘要 根据欧洲中心集合预报系统 2 m 气温预报的集合统计值,提出了 BP-SM 方法,针对中国 512 个台站 2016 年 3 月的日最高(低)气温做预报分析。将集合预报系统的模式直接输出、BP 和 BP-SM 方法得到的日最高(低)气温进行了比较,结果表明:预报时效越长,BP-SM 方法较之 BP 方法的预报优势也更明显;在 1 至 5 d 的预报中,BP-SM 方法显著降低了预报绝对误差,误差在 2°C 以内的准确率大部分在 60% 以上,部分站点达到 90%;正技巧评分均值大多高于 30%,在青藏高原东部和南部地区超过了 60%。预报正技巧站点次数在绝对误差 $\leq 2^\circ\text{C}$ (1°C) 范围内有所提高,对日最高气温预报准确率的提高略好于日最低气温;BP-SM 方法有效地降低了预报系统偏差,较大预报误差出现次数显著减少。

关键词 集合预报系统,日最高气温,日最低气温,BP-SM 方法

中图分类号 P456.4

1 引言

不确定性是天气预报的特征,集合预报是对不

确定性进行描述的概率预报系统。集合后处理技术关注点是如何提高集合概率预报,有直方图法(Hamill, et al, 1997, 1998)、集合映射(Roulston, et al,

* 资助课题:发展格点预报融合和订正关键技术(YBGJXM201703)。

作者简介:熊敏詮,主要从事天气学研究。E-mail: minquanxiong@sina.com

2003)、贝叶斯模型平均法(Raftery, et al, 2005)、逻辑回归(Hamill, et al, 2004)、非奇次高斯回归(Gneiting, et al, 2005)、相似法(Hamill, et al, 2006)、“预报同化”法(Stephenson, et al, 2005)等。近些年有较多的深入研究和进展,例如:Wilks等(2007)对比3种不同的后处理技术在集合气温和降水概率预报中的优劣,得到训练样本长度和预报技巧有密切关系。Schmeits等(2010)根据ECMWF的20 a总降水量集合再预报数据集,对比分析了EPS原始预报、修正的BMA和扩展逻辑方法在降水概率预报上的优劣。Roulin等(2012)使用扩展的逻辑回归提高EPS降水概率预报。Mendoza等(2015)使用不同的统计后处理技术,结合高分辨率的区域气候模式,实现降尺度的日降水概率预报。

在多模式集合、集成技术领域,中国学者有多方面进展,例如:陈超辉等(2010)开展了有限区域模式的多模式短期超级集合预报研究。王敏等(2012)基于集合2 m气温预报,对比了非均匀高斯回归方法与自适应卡尔曼滤波偏差订正方法订正效果。刘琳等(2013)基于T213集合预报极端降水天气预报指数,对中国极端强降水天气进行预报试验和分析。刘建国等(2013)根据全球多模式集合预报(TIGGE)资料,进行地面日均气温贝叶斯模型平均概率预报及其检验与评估。刘永和等(2013)对TIGGE资料在沂沭河流域6 h降水集合预报能力做了分析。杜钧等(2014)揭示了“集合异常预报法”和集合预报在极端高影响天气预报中的优越性。马旭林等(2015)利用日本气象厅区域集合预报的850 hPa气温资料,分析了基于卡尔曼滤波递减平均法的集合预报综合偏差订正。

由于次网格的不确定性影响,集合预报系统对近地面气象要素预报可靠性不如高层大气要素,如何提高近地面的气温和降水概率预报是统计后处理技术发展的主要方向。另外,确定性预报也有着巨大需求,特别是在短期预报中,人们总是根据确定性预报做出决策,集合预报提供了未来天气演变的多种可能性,通过适当的方法提炼出确定性信息也需要深入研究。针对近地面日最高(低)气温预报,根据集合预报系统的统计特征量,提出了BP-SM方法,明显提高了预报准确率。

2 资料和方法

2.1 资料

根据2015年12月1日至2016年3月31日08时(北京时,下同)的欧洲中心集合预报产品,预报时效为0到240 h都是6 h间隔输出,共51个预报成员,可得到集合平均、最低、最高、中数、众数、离散度共7个统计量。以集合预报系统6 h间隔的2 m气温的集合统计特征量,每天4个时次的高(低)值作为日最高(低)值,并为预报模型输入量。在集合统计量得到的诸多日最高(低)值中,由集合平均得到的日最高(低)气温通常有较高的精度,当对集合预报系统进行评价时,就以此作为日最高(低)气温的DMO(模式直接输出);针对中国512个台站(图1),在2016年3月1日至3月31日08时,逐日进行日最高(低)气温多个时效的预报。

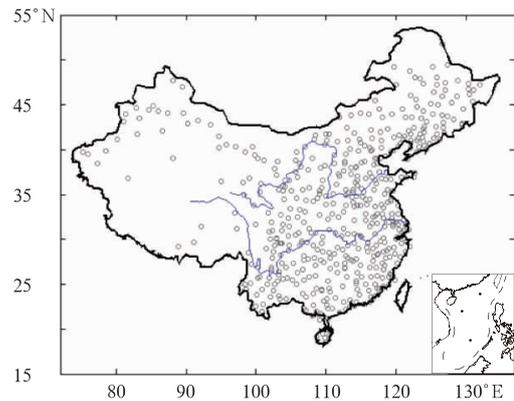


图1 512个测站的分布

Fig. 1 Geographic distribution of 512 stations in China

2.2 方法

BP(Back-Propagaion)网络是人工神经网络的重要组成部分,广泛应用于诸多领域,其由3个环节构成,一是正向传播:信号从输入层经隐层到输出层;二为反向传播:预测误差从输出层,经原来的通道到达输入层;三是迭代过程:设立误差目标,当正向传播未达到目标值,通过反向传播,调整网络权值和阈值,反复迭代直至达到目标值。本研究使用集合预报的7个统计量作为输入层的7个节点,由Kolmogorov定理,隐层节点数设为15,输出层是1

个节点,即预测和实际观测之差。构成3层网络进行日最高(低)气温预报,以预测误差(观测值减DMO)为目标值,出于实用考虑,迭代过程均为1000步,同时,网络学习过程采用经典的动量梯度法,步长为0.05。

泛化性是BP网络好坏的关键,一方面是BP算法本身固有缺陷造成的,网络训练过程中较易陷入局部极小而“难以自拔”,影响到网络性能;相关参数选取缺乏理论指导等。另一方面来自于问题的复杂性,首先是训练样本和独立(预报)样本差异,本研究使用的滚动建模中70个训练样本距离预报日最近的也要提前1d,当作120h预报时,相差就达到了5d,3月中国气温变化剧烈,在某些时段某个地区容易出现偏高(低)气温,那么,偏高(低)的输入量就容易出现偏高(低)的输出,即异常值;同时,训练样本变化易致网络权值变动,对于动态建模而言,每天都

有最新的样本增加,得到的训练结果会有差异,当新样本数值上出现较大差异时,网络的权值变动也大,表现为BP网络对历史样本“记忆”功能缺失及网络稳定性变化。由此,提出BP神经网络-自忆(简称BP-SM)建模方法。

自相关函数是描绘时间序列过去对未来的影响,EPS预报误差时序普遍有一阶或二阶记忆特征。如开展上海南汇(58369)2016年3月1日08时24h预报,以2015年12月21日至2016年2月29日08时的24h集合平均日最高气温预报为训练样本,得到预报偏低(共有66d)和偏高(4d)两个序列,预报偏低的误差序列自相关函数(图2)表明上一期会影响到下一期,当时滞大于3时,自相关函数位于两倍标准误差带中,为单边衰减、二阶截尾。可使用自回归动态建模,通过历史数据对未来进行预测:

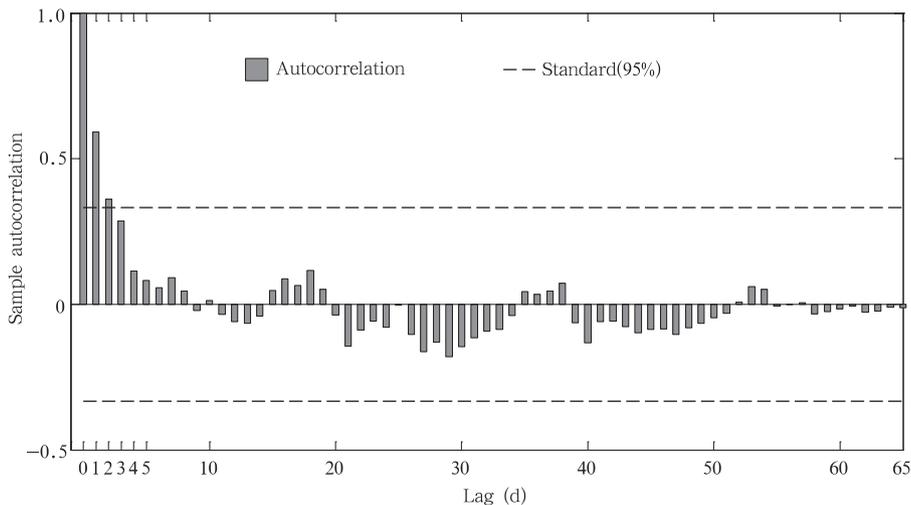


图2 2015年12月21日至2016年2月29日08时上海南汇(58369)

DMO日最高气温24h预报偏低的误差序列自相关函数

Fig. 2 Autocorrelation function of error series for DMO underestimated daily 2 m maximum air temperature forecasts over Nanhui of Shanghai during the period of 21 Dec 2015 - 29 Feb 2016 at 08:00 BT

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (1)$$

式中, y_t 是预报误差; ϕ 为自回归系数,即 y_t 记忆能力的强弱; y_{t-1} 是一阶滞后的预报误差; ε_t 为现期; $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ 满足 $E(\varepsilon_t) = 0$ 、 $E(\varepsilon_t^2) = \sigma^2$ 、 $E(\varepsilon_t \varepsilon_{t'}) = 0$ ($t \neq t'$)的白噪声过程,序列 $y_0, y_1, y_2, \dots, y_{t-1}, y_t$ 都以

式(1)相类似的表达,使用递归替代法, y_t 可写成初始值 y_{t-1} 和 ε 在第0期到第 t 期的历史值的函数:

$$y_t = \phi^{t+1} y_{-1} + \phi^t \varepsilon_0 + \phi^{t-1} \varepsilon_1 + \phi^{t-2} \varepsilon_2 + \dots + \phi \varepsilon_{t-1} + \varepsilon_t \quad (2)$$

由 $\hat{y}_t = \phi y_{t-1}$, \hat{y}_t 是 y_t 的估计值,

$$y_t - \varepsilon_t = \phi^{t+1} y_{-1} + \phi^t (y_0 - \hat{y}_0) + \phi^{t-1} (y_1 - \hat{y}_1) + \phi^{t-2} (y_2 - \hat{y}_2) + \dots + \phi (y_{t-1} - \hat{y}_{t-1}) \quad (3)$$

$$\begin{aligned} \hat{y}_t &= (\phi^{t+1} y_{-1} + \phi^t y_0 + \phi^{t-1} y_1 + \phi^{t-2} y_2 + \dots + \phi y_{t-1}) \\ &\quad - (\phi^t \hat{y}_0 + \phi^{t-1} \hat{y}_1 + \phi^{t-2} \hat{y}_2 + \dots + \phi \hat{y}_{t-1}) \quad (4) \\ &= (\phi^{t+1} y_{-1} + \phi^t y_0 + \phi^{t-1} y_1 + \phi^{t-2} y_2 + \dots + \phi y_{t-1}) \\ &\quad - (\phi^t \hat{y}_0 + \phi^{t-1} \hat{y}_1 + \phi^{t-2} \hat{y}_2 + \dots + \phi \hat{y}_{t-1}) \quad (5) \end{aligned}$$

$$\text{令 } L_t = \beta_0 y_0 + \beta_1 y_1 + \beta_2 y_2 + \dots + \beta_\tau y_\tau \quad (6)$$

式中, $\{\beta_i\}_{i=0}^\tau \approx \{\phi^{t-i}\}_{i=0}^\tau$, $\tau = t - 1$, τ 是时间常数, 即训练样本长度。式(5)左边第 2 项 $\{\hat{y}_i\}_{i=0}^t$ 并非严格的平稳时间序列, 在一些天气演变中, 变化特点往往难以估计, 通过自回归动态建模获得的 $\{\hat{y}_i\}_{i=0}^t$ 总是达不到理想的精度。若 $\{\hat{y}_i\}_{i=0}^t$ 为白噪声, 式(5)左边第 2 项较小, 可略去; 将 L_t 作为式(5)右端项形式表达, 即存在有限大的 L_t , 满足 $\hat{y}_t < L_t$ 。式(6)表明 L_t 和历史预报误差有密切关系。使用指数衰减方程描述构造函数随时间的变化: $\frac{dF(t')}{dt'} = -\frac{F(t')}{\tau}$, $t' = \tau - t$, $F(t')$ 为 t' 时刻误差系数, 方程的解析解: $F(t') = F(0) \exp(-t'/\tau)$ 。训练样本按时间顺序排列 $i = 0, 1, 2, \dots, \tau$, 愈临近当前时间的样本序列号愈大。对于有限大 L_t , 设 $\sum_{i=0}^\tau \beta_i = 1, \beta_i =$

$$\begin{aligned} &\sum_{i=0}^\tau \frac{\exp(-\frac{\tau-i}{\tau})}{\sum_{i=1}^\tau \exp(-\frac{i}{\tau})}, \text{ 当前时刻极大似然值 } L_t = \\ &\sum_{i=0}^\tau \frac{y_i \exp(-\frac{\tau-i}{\tau})}{\sum_{i=1}^\tau \exp(-\frac{i}{\tau})}; \text{ 由于要对误差区间估计, 将 } y_i \end{aligned}$$

按正负值分成两个序列, 分别计算出上界 E_{up} 和下界 E_{down} 。将 BP 方法得到的预报值限定在 $[E_{\text{down}}, E_{\text{up}}]$ 区域内, 称为 BP-SM 方法。

采用滚动建模方法进行实时训练和预报, 例如: 当要做 2016 年 3 月 1 日 08 时的 24 h 预报, 就以 2015 年 12 月 26 日至 2016 年 2 月 27 日集合产品为训练样本建模, 而 2016 年 3 月 1 日 08 时的 120 h 预报, 其训练样本则是 2015 年 12 月 21 日至 2016 年 2 月 22 日的资料, 依此类推。模型就可以反映出

数值模式最近的误差特点, 有助于提高预报精度。

预报评价方面, 日常业务中使用预报绝对误差、准确率和技巧评分, 具体如下: 最高(低)气温预报平均绝对误差: $T_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |F_i - O_i|$; 预报准确率: $TT_K = \frac{N_{r_K}}{N_{f_K}} \times 100\%$; 其中, N 是参加检验预报总次数(例如计算 3 月 1 日全国 24 h 的气温预报平均绝对误差时, N 即为 512), F_i 为第 i 次的预报, O_i 第 i 次观测; $K = 1$ 时, $|F_i - O_i| \leq 1^\circ\text{C}$; $K = 2$ 表示 $|F_i - O_i| \leq 2^\circ\text{C}$, N_{r_K} 是预报正确的次数, N_{f_K} 为总预报次数(例如计算 3 月 1 日全国 24 h 的气温预报绝对误差 $\leq 1^\circ\text{C}$ 的准确率, N_{f_1} 为 512)。技巧评分 $SST = \frac{T_{\text{BP}} - T_{\text{DMO}}}{T_{\text{DMO}}} \times 100\%$, T_{BP} 是使用 BP-SM 方法得到的日最高(低)气温预报平均绝对误差, T_{DMO} 是通过集合平均得到的最高(低)气温预报平均绝对误差。

3 结果分析

3.1 BP-SM 和 BP 方法的比较

通过 BP-SM 方法得到的预报绝对误差小于相应的 BP 方法, 称为“正效应”; 反之, 为“负效应”; 若 BP 预测值位于期望区内, 则两者一致(重合); 将“正效应”中的各单站预报绝对误差降低值求和、再取平均, 即得“正效应”产生的预报误差减少值; 同理, 有“负效应”站点的预报绝对误差增加值的均值为“负效应”增加的误差值, 由此, 可以直观地分析两种方法预报的优劣。在 512 个台站 1 至 10 d 的日最高气温预测中, 有如下特点(图 3a): 首先, 对于中长时效(3 至 10 d), BP-SM 法有一定优势, 例如第 10 天的预报, “正效应”绝对误差降低值 2.41°C , 而“负效应”引起的增加值是 1.71°C , 两者数值相差较大, 对应是预报准确率的明显提高; 其次, 随着时效的延长, “负效应”预报次数逐渐下降, 而“正效应”次数总体均衡或略有增长, 如第 1 天“正(负)效应”次数是 2688(3091), 第 10 天是 3814(1064); 最后, BP 算法并不是十分完善的神经网络, “正效应”和“重合”次数大致是 3000 : 10000, 占有相当比例, 既便是在较高的“重合”次数中还有异常值, 表明 BP-SM 方法的积极意义和深入分析之价值。

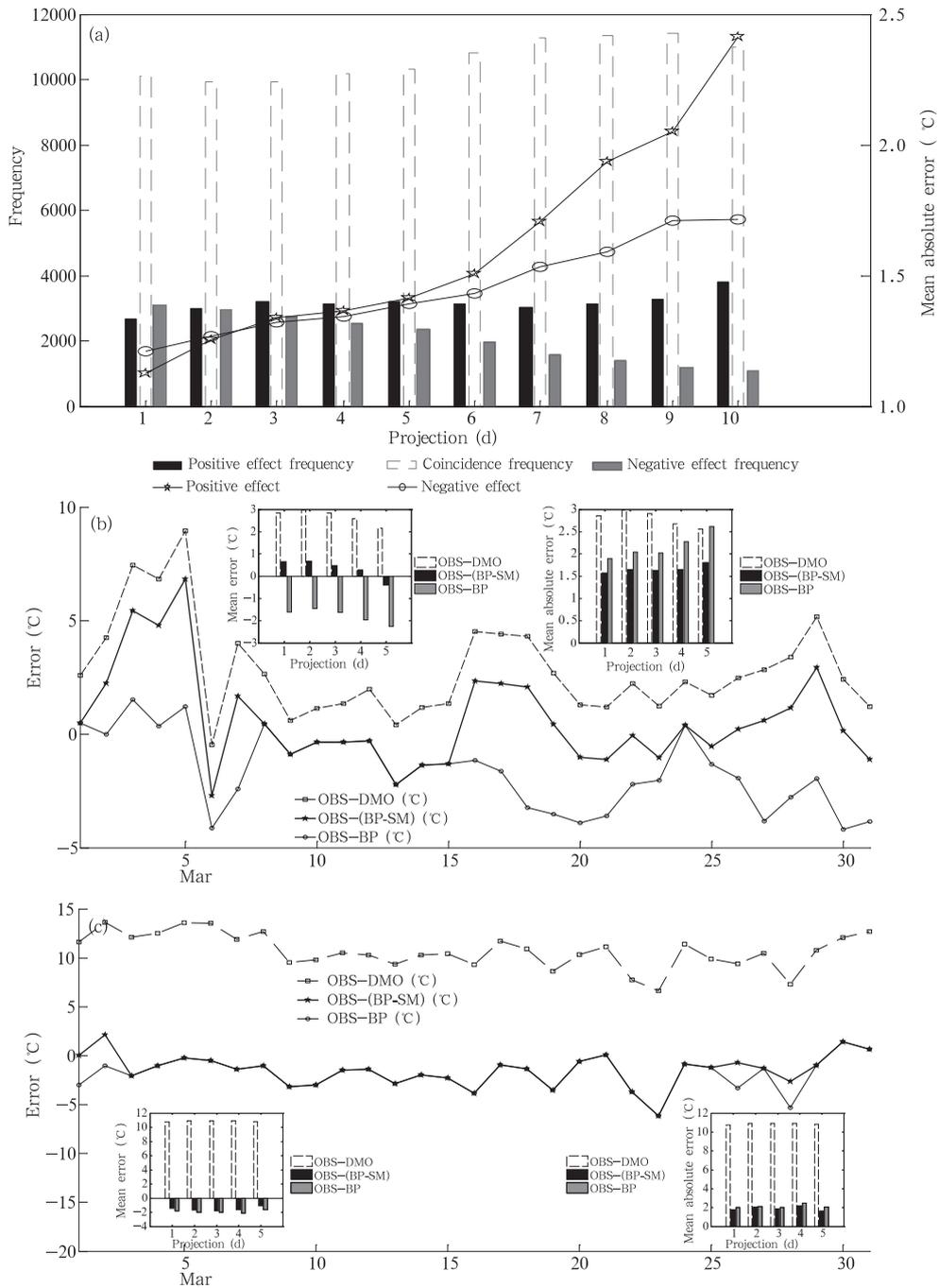


图3 2016年3月DMO、BP和BP-SM方法的日最高气温预报比较(a. 1至10 d的BP-SM方法“正(负)效应”预报次数和误差增减值; b. 上海南汇24 h预报误差的逐日比较(主图), 1—5 d的预报误差和绝对误差月均值比较(子图); c. 拉萨市24 h预报误差的逐日比较(主图), 1—5 d的预报误差和绝对误差月均值比较(子图))

Fig. 3 Comparisons of DMO, BP and BP-SM for monthly mean daily 2 m maximum air temperature forecasts during the period of 1–31 Mar 2016, (a) frequency, increment and decrement to forecast errors of positive and negative effects for BP-SM method over 10 d lead time; (b) comparisons (main diagram) of daily 24 h forecast errors over Nanhui of Shanghai, and monthly mean of daily 2 m minimum air temperature forecast errors and absolute errors (subgraph) over 5 d lead time; (c) comparisons (main diagram) of daily 24 h forecast errors over Lhasa and monthly mean of daily 2 m minimum air temperature forecast errors and absolute errors (subgraph) over 5 d lead time)

上海南汇日最高气温 24 h 预报误差逐日变化曲线(图 3b 主图): DMO 误差起伏大, 最大值 8.97°C (5 日), 最小值 -0.470°C (6 日), 预报值大多低于实况; BP 方法只在头几天有较高的预报精度, 但中下旬误差曲线走势完全不同于 DMO, 而且预报值明显比实况高; BP-SM 的误差曲线和 DMO 的线型走势相近, 特别是在中下旬, 有力地修正了原始 BP 预报误差曲线, 较大幅度地降低了误差, 主要原因是 BP-SM 考虑了 EPS 预报误差的历史信息。5 个不同时效的误差月均值: DMO(BP-SM、BP) 分别是: 2.833°C (0.637°C 、 -1.648°C)、 2.965°C (0.666°C 、 -1.464°C)、 2.838°C (0.483°C 、 -1.656°C)、 2.580°C (0.293°C 、 -2.005°C)、 2.155°C (-0.360°C 、 -2.198°C)。平均绝对误差(简称“MAE”, 下同)月均值: 2.863°C (1.589°C 、 1.935°C)、 2.987°C (1.656°C 、 2.06°C)、 2.909°C (1.632°C 、 2.057°C)、 2.670°C (1.627°C 、 2.287°C)、 2.554°C (1.796°C 、 2.624°C)。BP-SM 的误差在所有 5 个时效中是最小的, 尤其是显著降低了误差(图 3b 子图); 在较长时效(第 4 天和第 5 天)预报中, BP-SM 方法的有效性增强。

在日常天气预报业务中, BP-SM 方法的使用分两步进行, 以上海南汇站为例, 在 2016 年 3 月 30 日 08 时, 进行日最高气温的 24 h 预报, 下文较详细地介绍了预报过程。第 1 步(历史资料的 BP 建模和 SM 方法计算): 选择 2016 年 1 月 19 日至 2016 年 3 月 29 日 08 时的日最高气温 24 h 预报, 共计 70 d 资料作为训练样本集, 每天的日最高气温 7 个预报统计量作为 BP 方法的输入值, 而 BP 网络的目标值是 DMO 误差, 完成 BP 网络的构建; 根据 DMO 误差正、负值划分成预报“偏低”和“偏高”两个序列, “偏低”序列有 67 d, 只有 3 d 属于预报“偏高”, 使用式(6)可分别得到两个边界值 -0.32 和 2.27°C 。第 2 步(实时预报): 将当天 7 个日最高气温 24 h 预报的统计量输入到已建好的模型中, 2016 年 3 月 30 日 08 时 EPS 在上海南汇的 7 个预报统计量依次是: 16.27 、 16.41 、 18.15 、 15.1 、 16.12 、 15.8 、 0.67°C , 经 BP 网络预报值是 23.11°C ; 根据上一步 SM 方法求得的边界值加上 DMO 的 24 h 预报(16.27°C), 即有 $[15.95^{\circ}\text{C}, 18.54^{\circ}\text{C}]$; BP 法预报已经超出 SM 方法确定的边界, 可判定 BP 网络输出属于异常值, 就以邻近边界的 18.54°C 作为 BP-SM 方法最终预报。3 月 30 日 08 时—31 日 08 时上海南汇日最高气温

观测值是 18.7°C , 那么, DMO(BP-SM、BP) 预报误差是 2.43°C (0.16°C 、 -4.41°C), BP-SM 法预报精度高。

拉萨(55591)日最高气温 24 h 预报误差逐日变化曲线(图 3c 主图): DMO 误差曲线总体平稳, BP-SM 和 BP 方法预报基本重合, 只有 3 d 的预报有较大的差别, 在 1 日、26 日、28 日的 BP-SM(BP) 方法预报误差值对应是: 0.053°C (-2.974°C)、 -0.714°C (-3.299°C)、 -2.655°C (-5.292°C), BP-SM 法的预报值更接近于实况。5 个不同时效的误差月均值: DMO(BP-SM、BP) 分别是: 10.725°C (-1.476°C 、 -1.844°C)、 10.928°C (-1.672°C 、 -2.003°C)、 10.893°C (-1.790°C 、 -2.008°C)、 10.904°C (-1.661°C 、 -2.096°C)、 10.812°C (-1.047°C 、 -1.605°C)。MAE 月均值: 10.725°C (1.759°C 、 1.987°C)、 10.928°C (2.034°C 、 2.083°C)、 10.893°C (1.881°C 、 2.024°C)、 10.904°C (2.204°C 、 2.455°C)、 10.812°C (1.647°C 、 2.0312°C)。在 1 至 5 d 的预报中(图 3c 子图), BP-SM 方法都有优势。

BP 网络预测精度和数值模式预报能力密切相关。有较大误差的训练样本反馈在网络权值上, 网络不稳定性增大, 导致较大的输出偏差, 表现为网络学习和记忆是不稳定的, 有别于人类大脑记忆的稳定性, 也是泛化性问题的反映; 中长期(2 d 以上)预报中, 数值模式预报误差逐渐增大, BP-SM 方法有较好的效果; 对于短时(1 d)预报, BP 方法预报有较高精度, 需要更细致的分析, BP-SM 方法才能达到预期。

3.2 绝对误差、准确率和技巧评分空间分布

日最高气温未来 5 d 的预报, DMO 的 $\text{MAE} \leq 2^{\circ}\text{C}$ 的站点位于: 东北地区和华北平原大部分地区、长江中下游平原北部, 但是, 在青藏高原东部和南部, DMO 的 MAE 超过了 4°C , 局部地区超过 16°C 。BP-SM 方法得到的误差分布上, $\text{MAE} \leq 2^{\circ}\text{C}$ 的站点数在 DMO 相对应的区域都有所扩大, 并且新增了青藏高原中部、内蒙古高原、黄土高原、四川盆地和云南的部分地区, 中国的大部分地区的误差都小于 4°C , 较之 DMO, 有明显的提高。从技巧评分来看(图 4a—e), 对青藏高原上站点气温预报提高显著, 一般超过 60% 的正技巧, 局部超过 80%; 其他大部分地区也有较高的正技巧; 只有在东北北部、内蒙古中部、华北平原西南部和华南西部的部分地区, 呈现

负技巧,数值大致在 0 至 -20%,随着预报时效的延长,负技巧面积向东略有扩大。负技巧区出现在 DMO 预报准确率比较高的地区:负技巧站点的 DMO 绝对误差 $\leq 2^{\circ}\text{C}$ 准确率基本在 70% 以上,而在

华北平原及其以北地区的相应站点 DMO 准确率大多超过 80%,部分达到 90%,反映了当模式预报有较高准确率、预报偏差微弱时,通过 BP-SM 方法达到更高的准确率还需要深入的研究。

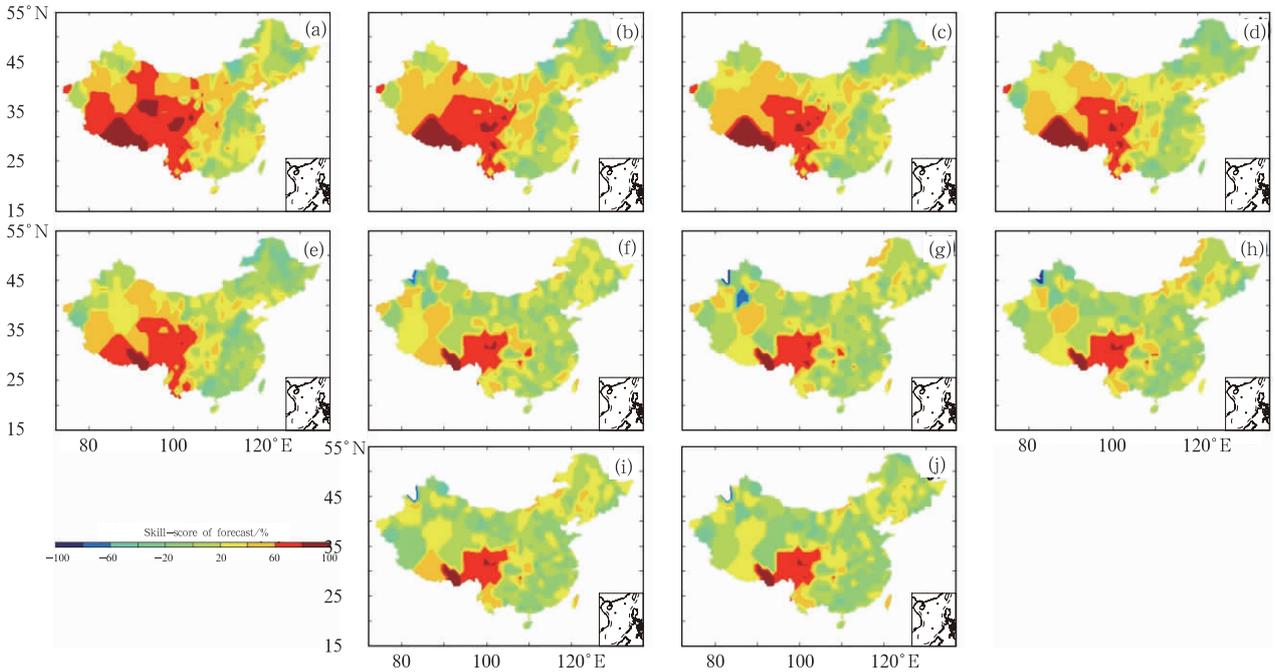


图 4 2016 年 3 月 BP-SM 方法在 1 至 5 d 气温预报的月平均技巧评分
(a—e. 最高气温, f—j. 最低气温)

Fig. 4 Monthly mean skill-scores of temperature forecasts over China for the period of 1 Mar 2016 – 31 Mar 2016
(a – e. daily 2 m maximum air temperature, f – j. daily 2 m minimum air temperature)

最低气温预报 MAE 低于最高气温,在江南、华南地区 DMO 的 5 d 预报中 MAE 大都在 2°C 以内,准确率也在 70% 以上,在中国南方的部分地区超过了 80% (日最高气温准确率高值区出现在长江以北的平原地区),青藏高原东部和南部的误差值也有所下降。BP-SM 方法在中国大部分地区都小于 2°C ,即使在青藏高原也低于 4°C ,明显降低了 DMO 预报误差;不同时效都有较高的准确率,在中国中部和东部沿海的部分地区的预报准确率超过了 80%。技巧评分方面(图 4f—g),青藏高原东部超过了 60%,负技巧评分也出现在原始 DMO 准确率较高的站点,主要位于中国南方、华北和西北的部分地区,但是负技巧分都较小(0 至 -10%)。

从日最高(低)气温预报技巧空间分布来看,在青藏高原东部和南部都有较高的正技巧,BP-SM 方法对日最高气温预报准确率提高能力好于日最低气

温;负技巧评分站点有相对固定的“区域性”特点,随着预报时效的延长(从 1 到 5 d 的预报),该区域逐渐向附近扩散、面积也增大;而且负技巧区通常出现在 DMO 准确率较高的站点。

3.3 不同误差范围准确率的变化特点

预报技巧评分是根据所有预报绝对误差的总和计算的,没有考虑到预报误差的离散度,结果是某个较大的预报误差可能会导致整个预报评分大幅的变化;基于神经网络方法预报的过拟合现象是普遍存在的,个别时次的预报异常容易对评分产生负面影响,技巧评分就难以全面反映该方法预报的优劣;同时,日常实践中,预报异常值往往可以通过主、客观方式得到有效抑制,人们更为关心的是某个具体误差段的准确率;因此只是笼统地给出所有预报误差段的技巧评分难以全面反映方法的优劣,那么,以日常使用的准确率为基准,绘出了绝对误差 $\leq 2^{\circ}\text{C}$

(1℃)准确率不同变化方向的站点数比率。

图 5a 显示了各个预报时效中,512 个台站预报评分正负技巧站点数和分值。日最高气温中,BP-SM 在 1 至 5 d 的预报正技巧比例(站点数)是 86%

(440)、82% (420)、77% (394)、74% (379)、68% (348),而相应的预报技巧是 38%、35%、33%、32%、30%。

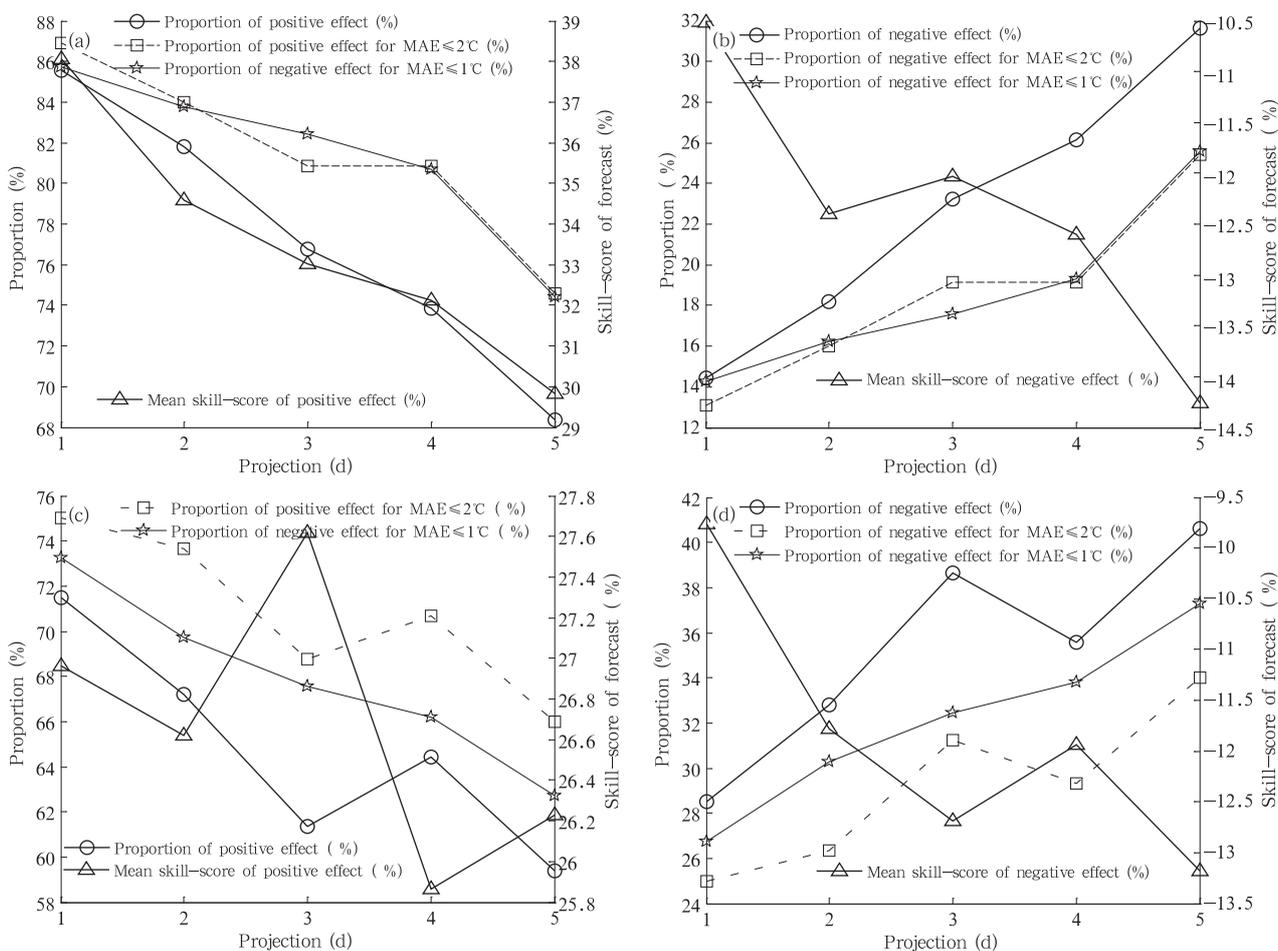


图 5 2016 年 3 月 DMO 和 BP-SM 方法在不同误差范围的技巧评分比较

(a. 日最高气温的正技巧比较, b. 日最高气温的负技巧比较, c. 日最低气温的正技巧比较, d. 日最低气温的负技巧比较)

Fig. 5 The comparisons of skill-scores in different error ranges for DMO and BP-SM method during 1-31 Mar 2016

- positive skill-score of daily 2 m maximum air temperature forecasts,
- negative skill-score of daily 2 m maximum air temperature forecasts,
- positive skill-score of daily 2 m minimum air temperature forecasts,
- negative skill-score of daily 2 m minimum air temperature forecasts)

通过 BP-SM 方法得到的预报绝对误差 ≤ 1℃ 准确率呈现上升趋势的比例(站点数)有 87% (445)、84% (430)、82% (420)、81% (415)、75% (384);对于绝对误差 ≤ 2℃ 准确率,也有相似的变化特点,图形上处于“绞合”状态。即在较小预报误差范围内,BP-SM 方法优势凸出。负技巧评分而言(图 5b),1

至 5 d 站点比例依次是 14%、18%、23%、26%、32%,平均分值是 -11%、-12%、-12%、-13%、-14%;而绝对误差 ≤ 1℃ 准确率呈下降趋势的站点比例为 14%、16%、18%、19%、26%。所以,在较小预报误差范围内,BP-SM 方法优势凸出。

日最低气温的预报误差比较也有上述类似特

点。BP-SM 方法预报正技巧(图 5c)1 至 5 d 的站点比例为 71%、67%、61%、64%、59%;绝对误差 $\leq 1^{\circ}\text{C}$ 准确率有提高的站点比例依次是 73%、70%、68%、66%、63%;平均正技巧分值是 27%、27%、28%、26%、26%。负技巧站点数则在 29%—41%;准确率下降(绝对误差 $\leq 1^{\circ}\text{C}$)的站点数介于 26%和 37%;而负技巧得分大约为 -11% 。

综上所述,BP-SM 方法在 5 d 的预报中正技巧站点数有较高的比例,分值也大多位于 30%附近,日最高气温预报的提高幅度要略好于最低气温。对于绝对误差 $\leq 2^{\circ}\text{C}$ (1°C),正技巧站点数占比和分值都要高于总体评价,同时,负技巧站点数也比较低,说明在人们较关注的预报误差范围,BP-SM 方法表现更佳。结合实际气象工作者通行的预报准确率分析,BP-SM 方法表现出了更优越的预报特性,说明了 BP 神经网络的泛化问题,BP-SM 方法虽然能大幅减少由此引发的异常值,但是还不能完全排除,也由于 BP 网络的高度非线性,难以得到精确的数值解,所以就表现在 DMO 预报准确率比较高的部分

站点出现负技巧现象。

3.4 系统偏差比较

根据日最高气温的 24 h 预报资料。对预报系统偏差进行分析,加深对 BP-SM 方法的认识,有助于方法完善和逐步提高预报精度。

散点图(图 6)直观地展现了预报的系统偏差,在对称轴($y=x$)附近(图 6a),离散点呈发散、非对称性,图中使用不同颜色表示误差范围,由于 DMO 预报误差普遍较大,远距离的数值点显得较多;另外,数值点大多位于对称轴线的下方,即 DMO 预报“偏低”。BP-SM 方法(图 6b)减小了系统性偏差,数值点在对称轴附近呈现收敛、对称性,远离对称轴的离散点较少,大多数集中于 $0-3^{\circ}\text{C}$,点的密度变得更紧凑、密集;数据点以对称轴为轴心呈“准对称”分布,DMO 预报“偏低”现象得到有效校正。DMO 在 6 个误差范围站次数(图 6c)依次是 4070、3630、2590、1830、1290、2440,而 BP-SM 为 6810、4660、2410、1080、488、397,误差在 $0-1^{\circ}\text{C}$ 、 $1-2^{\circ}\text{C}$ 范围的预报站次数增加明显,同时又大幅降低了 $>5^{\circ}\text{C}$ 误差站次。

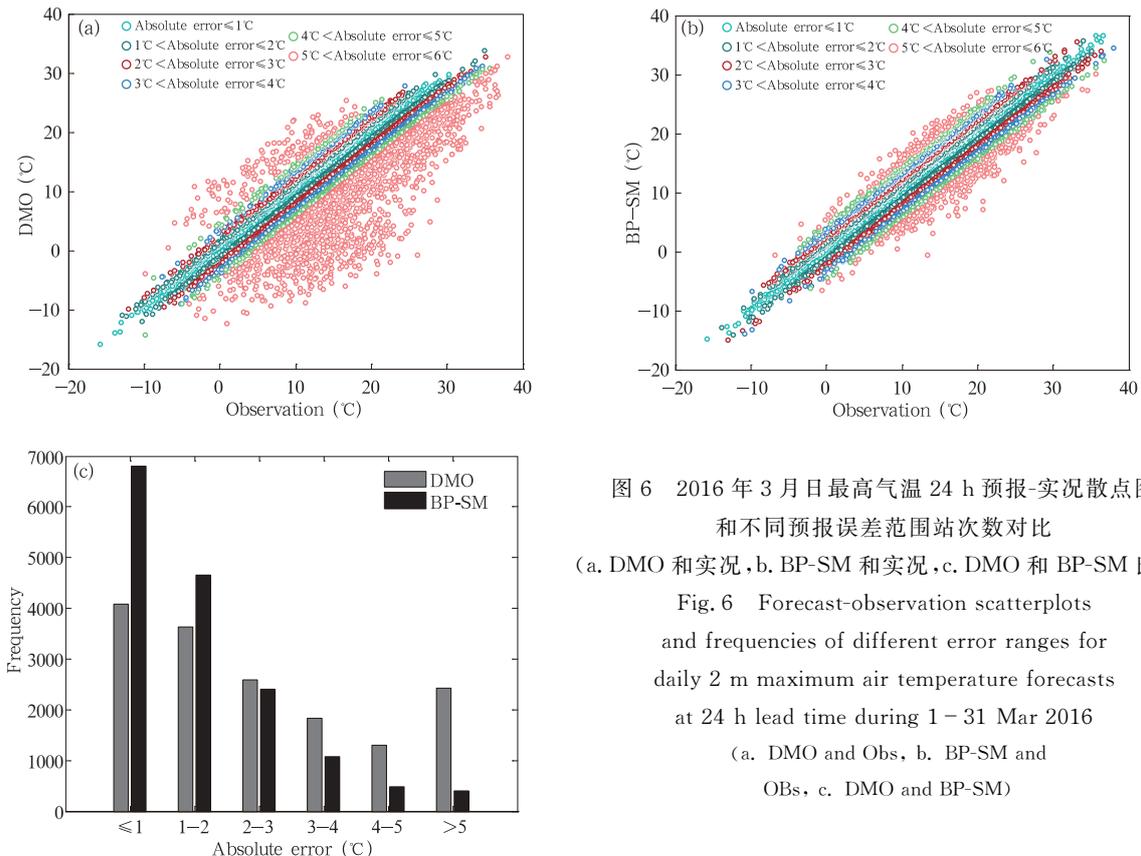


图 6 2016 年 3 月日最高气温 24 h 预报-实况散点图
和不同预报误差范围站次数对比
(a. DMO 和实况, b. BP-SM 和实况, c. DMO 和 BP-SM 比较)

Fig. 6 Forecast-observation scatterplots
and frequencies of different error ranges for
daily 2 m maximum air temperature forecasts
at 24 h lead time during 1-31 Mar 2016

(a. DMO and Obs, b. BP-SM and
Obs, c. DMO and BP-SM)

在中国的预报误差空间分布上,中西部地区 DMO 预报能力较弱,以 100°E 为界,分析中国西部、东部地区预报误差变化特点。西部地区,尤其是青藏高原东部地区,大致是 $0\text{--}20^{\circ}\text{C}$ (图 7a),实况和 DMO 差别悬殊,DMO 预报偏低,个别站次的误差超过了 20°C ,而经过 BP-SM 方法再预报后,得到极大的改进。图 7a 和 b 前后对比,表明在东部和西部,散点分布都变得更“收敛和对称”,BP-SM 有效

降低了数值模式预报的系统偏差。西部地区的 DMO 柱状图(图 7c)在 6 个误差范围中站次数分别是 139、175、200、236、190、544,而改进后为 562、392、255、145、76、54;DMO 预报误差迥异于其他地区,其中误差大于 5°C 站次较多,改进后,预报精度提高十分明显,特别是在较大误差预报。东部地区(图 7d)有相似的特点。

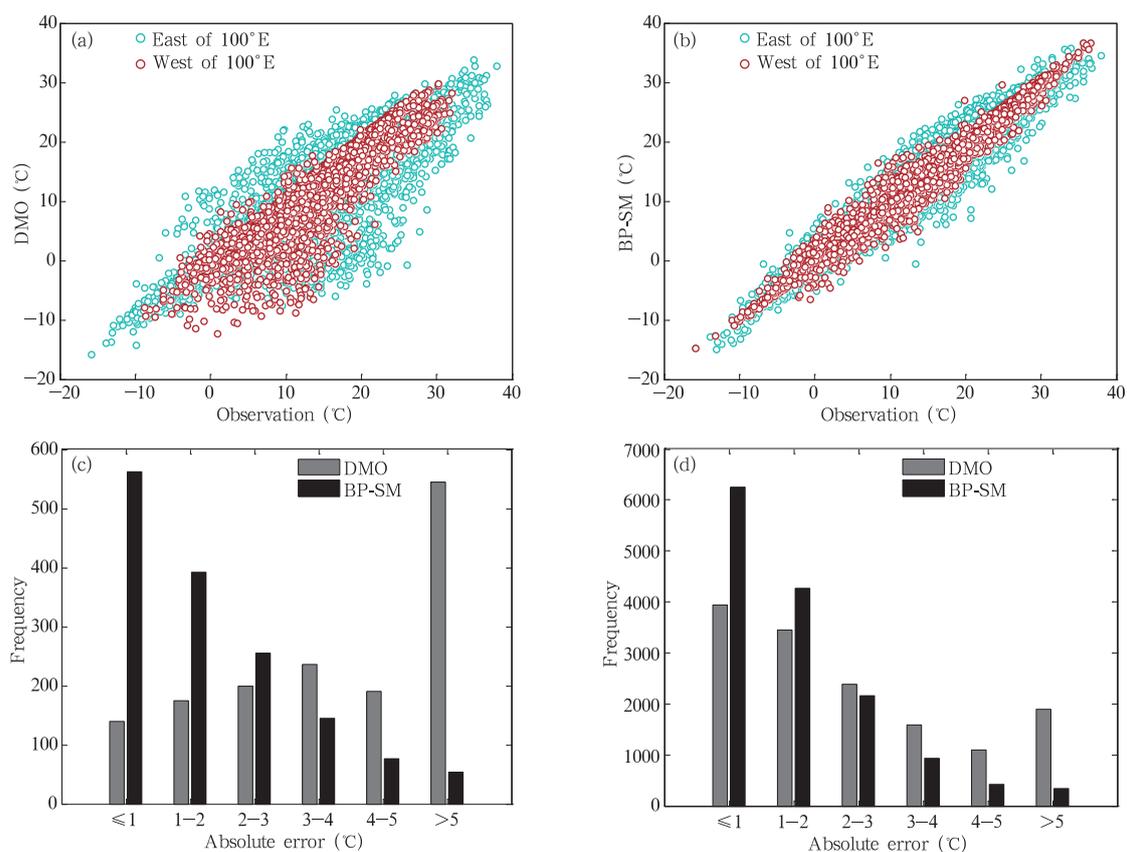


图 7 2016 年 3 月 100°E 以西和以东地区日最高气温 24 h 预报-实况散点图
(a. DMO 和实况, b. BP-SM 和实况) DMO 和 BP-SM 方法的日最高气温 24 h 预报
在不同误差范围站次数对比 (c. 100°E 以西地区, d. 100°E 以东地区)

Fig. 7 Forecast-observation scatterplots for daily 2 m maximum temperature forecasts at 24 h lead time over the west and east of 100°E in the during 1–31 Mar 2016:

(a) DMO, (b) BP-SM. Frequency comparisons between DMO and BP-SM methods in various error ranges over different areas: (c) to the west of 100°E , (d) to the east of 100°E

不同区域的 DMO 系统性偏差表现各异,可能是次网格不确定性、中国天气及地理复杂性密切相关,散点图呈发散、非对称性分布,将预报误差由小到大划分成 6 个区段,相应的预报站次数并非均匀地落在每个区段中,DMO 预报次数随着误差增大

而逐渐衰减,只是到了 $>5^{\circ}\text{C}$ 段有“翘尾”现象,因为文中将误差 $>5^{\circ}\text{C}$ 都统一划归到此段的缘故;而在西部地区 DMO 预报误差次数在 6 个区段的分布形态就比较奇特,可能是误差分段和集合预报在此区域的低准确率两方面原因引起的。BP-SM 方法则将

DMO 误差各异分布都转换成一致的衰减形态,而且衰减率也明显增大。

4 结论和讨论

(1) BP-SM 和 BP 方法预报精度对比分析中:随着预报时效的延长,BP-SM 方法优势也逐渐增大,“负效应”预报次数明显下降,表明 BP-SM 法能有效地减少 BP 网络预报异常值。上海南汇多个预报时效比较中:BP 网络的不稳定性导致较大的预报误差;而 BP-SM 法避免了上述现象,在较长时效预报中,BP-SM 法有更大的正向作用。通过上海南汇和拉萨两单站的 DMO 预报误差变化特点,可初步获悉 EPS 误差时空特点和 BP 网络预报准确度有密切关系。以上海南汇为例,较细致地介绍了 BP-SM 方法日常预报过程。

(2) BP-SM 方法十分显著地提高了日最高(低)气温预报准确率,尤其是在青藏高原东部和南部地区。和 DMO 相比,BP-SM 方法在大部分地区都有较高的正技巧,部分地区超过了 60%,负技巧站点通常出现在 DMO 预报准确率较高的部分地区,技巧分值较小(0—20%)。

(3)具体比较了绝对误差 $\leq 2^{\circ}\text{C}$ (1°C)准确率变化特点,在未来 5 d 预报中,BP-SM 方法在此误差范围内的正技巧站点比例表现出更大的优势,同时,对日最高气温预报准确率的提高能力要略好于日最低气温。探讨了 BP-SM 方法出现负技巧预报的原因,说明 BP-SM 方法有较强的非线性映射能力,当面对函数逼近精度和模型容错性等问题时,还要与天气变化特点相联系,可能会达到更好的预报效果。

(4)实况-预报散点图及在东、西部站点比较中,DMO 呈现发散、非对称性分布特点,而 BP-SM 方法减小了系统性偏差,数值点也收敛、对称地分布。BP-SM 方法也改变了 DMO 在 6 个误差区段分布形态的不一致性,使得误差越大区段站点数越少,分布结构呈快速衰减趋势。

传统的动力统计释用方法通常根据模式输出的动力因子进行建模预报,实践中如何获得最优预报子集往往是难点,也是预报质量优劣的关键;当面临集合预报海量的因子输出时,预报因子集选择将显得尤为困难;BP-SM 方法直接使用集合气温预报统计特征值,避免了上述问题。日最高(低)气温是日常气象研究和业务服务的重要预报量,受到广泛的

关注,基于集合预报系统而提出快速、准确建模和优化的 BP-SM 方法,可操作性强、实用性好、能在实践中发挥作用并将得到进一步完善。

参考文献

- 陈超辉,李崇银,谭言科等. 2010. 基于交叉验证的多模式超级集合预报方法研究. 气象学报, 68(4): 464-476. Chen C H, Li C Y, Tan Y K, et al. 2010. Research of the multi-model super-ensemble prediction based on cross-validation. Acta Meteor Sinica, 68(4): 464-476 (in Chinese)
- 杜钧, Grumm R H, 邓国. 2014. 预报异常极端高影响天气的“集合异常预报法”:以北京 2012 年 7 月 21 日特大暴雨为例. 大气科学, 38(4): 685-699. Du J, Grumm R H, Deng G. 2014. Ensemble anomaly forecasting approach to predicting extreme weather demonstrated by extremely heavy rain event in Beijing. Chinese J Atmos Sci, 38(4): 685-699 (in Chinese)
- 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. 大气科学, 37(1): 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. Chinese J Atmos Sci, 37(1): 43-53 (in Chinese)
- 刘琳,陈静,程龙等. 2013. 基于集合预报的中国极端强降水预报方法研究. 气象学报, 71(5): 853-866. Liu L, Chen J, Cheng L, et al. Study of the ensemble-based forecast of extremely heavy rainfalls in China: Experiments for July 2011 cases. Acta Meteor Sinica, 71(5): 853-866 (in Chinese)
- 刘永和,严中伟,冯锦明等. 2013. 基于 TIGGE 资料的沂沭河流域 6 小时降水集合预报能力分析. 大气科学, 37(3): 539-551. Liu Y H, Yan Z W, Feng J M, et al. 2013. Predictability of 6-hour precipitation in the Yishu River basin based on TIGGE data. Chinese J Atmos Sci, 37(3): 539-551 (in Chinese)
- 马旭林,时洋,和杰等. 2015. 基于卡尔曼滤波递减平均算法的集合预报综合偏差订正. 气象学报, 73(5): 952-964. Ma X L, Shi Y, He J, et al. 2015. The combined descending averaging bias correction based on the kalman filter for ensemble forecast. Acta Meteor Sinica, 73(5): 952-964 (in Chinese)
- 王敏,李晓莉,范广洲等. 2012. 区域集合预报系统 2 m 温度预报的校准技术. 应用气象学报, 23(4): 395-401. Wang M, Li X L, Fan G Z, et al. 2012. Calibrating 2 m temperature forecast for the regional ensemble prediction system at NMC. J Appl Meteor Sci, 23(4): 395-401 (in Chinese)
- Gneiting T, Raftery A E, Westveld A H, et al. 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon Wea Rev, 133(5): 1098-1118
- Hamill T M, Colucci S J. 1997. Verification of Eta-RSM short-range ensemble forecasts. Mon Wea Rev, 125(6): 1312-1327
- Hamill T M, Colucci S J. 1998. Evaluation of Eta-RSM ensemble

- probabilistic precipitation forecasts. *Mon Wea Rev*, 126(3): 711-724
- Hamill T M, Whitaker J S, Wei X. 2004. Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon Wea Rev*, 132(6): 1434-1447
- Hamill T M, Whitaker J S, Mullen S L. 2006. Reforecasts: An important dataset for improving weather predictions. *Bull Amer Meteor Soc*, 87(1): 33-46
- Mendoza P A, Rajagopalan B, Clark M P, et al. 2015. Statistical postprocessing of high-resolution regional climate model output. *Mon Wea Rev*, 143(5): 1533-1553
- Raftery A E, Gneiting T, Balabdaoui F, et al. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Mon Wea Rev*, 133(5): 1155-1174
- Roulin E, Vannitsem S. 2012. Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon Wea Rev*, 140(3): 874-888
- Roulston M S, Smith L A. 2003. Combining dynamical and statistical ensembles. *Tellus A*, 55(1): 16-30
- Schmeits M J, Kok K J. 2010. A comparison between raw ensemble output, (modified) Bayesian model averaging, and extended logistic regression using ECMWF ensemble precipitation reforecasts. *Mon Wea Rev*, 138(11): 4199-4211
- Stephenson D B, Coelho C A S, Doblas-Reyes F J, et al. 2005. Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus A*, 57(3): 253-264
- Wilks D S, Hamill T M. 2007. Comparison of ensemble-MOS methods using GFS reforecasts. *Mon Wea Rev*, 135(6): 2379-2390