

基于深度学习的中国地面气温的多模式集成预报研究

智协飞*, 王田, 季焱

南京信息工程大学 气象灾害预报预警与评估协同创新中心/气象灾害教育部重点实验室/大气科学学院, 江苏 南京 210044

* 联系人, E-mail: zhi@nuist.edu.cn

2020-02-19 收稿, 2020-04-01 接受

国家自然科学基金面上资助项目(41575104)

摘要 基于 TIGGE 资料中的欧洲中期天气预报中心、英国气象局、美国国家环境预报中心、韩国气象厅和日本气象厅 2015 年 1 月 1 日—9 月 30 日中国及周边地区地面 2 m 气温 24~168 h 集合预报资料, 利用长短期记忆神经网络(Long Short-Term Memory, LSTM)、浅层神经网络(Neural Networks, NN)、滑动训练期消除偏差集合平均(BREM)和滑动训练期多模式超级集合(SUP)方法对 2015 年 9 月 5—30 日 26 d 预报期进行集成预报试验。结果表明, BREM 对 5 个单模式进行等权集成, 预报结果易受预报效果较差模式的影响, 整体预报技巧略低于单个最优模式 ECMWF 的预报技巧。其中在新疆南部, 等权集成后的预报技巧更低。SUP 的预报结果比所有单个模式预报更为准确。在 144 h 之前, SUP 的误差明显小于 ECMWF 的预报误差, 但随预报时效增加, 误差增长幅度增大。NN 对地面气温的预报效果与 SUP 的预报效果相当。LSTM 整体预报效果最好, 特别是在预报时效较长(超过 72 h)时, 比其他方法预报准确率明显提高。LSTM 神经网络方法明显改进了我国西北、华北、东北、西南和华南大部分地区的气温预报, 但在南疆部分地区误差较大。

关键词
深度学习;
人工智能;
多模式集成
预报;
LSTM 神经网络;
地面气温

过去几十年数值天气预报的发展取得了长足的进步。受全球气候变化的影响, 近年来高温、严寒等极端天气事件频发, 人们对气温预报准确率的要求越来越高。但在天气预报业务中, 由于不同的数值模式在分辨率、初始场、资料同化技术、动力框架以及物理参数化方案等方面具有明显差异, 各模式的预报能力存在较大差异, 因此发展多模式集成预报技术十分必要(杨学胜, 2001)。多模式集成预报技术是对各模式预报的输出结果进行统计后处理, 集成多模式的预报结果来减小预报误差(智协飞等, 2013)。国内外许多专家对此技术开展了广泛深入的研究(Krishnamurti, 1999; Krishnamurti et al., 2000a, 2000b, 2007, 2009, 2016; 马清等, 2008; 智协飞等, 2009, 2010, 2013, 2015a, 2015b, 2016; 林春泽等, 2009, 2013; 陈超辉等, 2010; Zhi et al., 2011, 2012; He et al., 2015; Ji et al., 2019, 2020; 智协飞和

黄闻, 2019)。研究发现, 多模式超级集合预报的预报技巧远超过单个模式以及简单集合平均的预报技巧, 显著减小了天气预报和季节气候预测的误差。马清等(2008)运用超级集合预报方法对全球 5 个模式的温度预报进行研究, 结果表明超级集合预报相比单个模式预报不存在明显的系统误差, 而且平均绝对误差有所降低。智协飞等(2009)分别运用固定训练期和滑动训练期超级集合对地面气温进行预报, 结果发现滑动训练期超级集合的误差小于固定训练期超级集合的误差, 并且训练期的长度在一定程度上影响超级集合预报的结果, 随着训练期长度的增长, 均方根误差不断减小, 但训练期达到一定长度后, 误差减小幅度下降, 甚至出现误差增大。张玲和智协飞(2013)利用多模式集成方法对 2008 年中国南方冰冻雨雪、低温天气进行预报试验, 发现滑动训练期消除偏差集合平均方法的预报误差小于单

引用格式: 智协飞, 王田, 季焱. 2020. 基于深度学习的中国地面气温的多模式集成预报研究[J]. 大气科学学报, 43(3): 435-446.

Zhi X F, Wang T, Ji Y. 2020. Multimodel ensemble forecasts of surface air temperature over China based on deep learning approach[J]. Trans Atmos Sci, 43(3): 435-446. doi: 10.13878/j.cnki.dqkxxb.20200219003. (in Chinese).

个模式预报的误差,且能有效延长气温预报的预报时效。崔慧慧和智协飞(2013)对地面气温做延伸期预报,多模式集成的预报准确率均高于单个模式的预报准确率,超级集合的预报准确率最高。He et al.(2015)和张涵斌等(2015)运用多模式集成预报技术对西太平洋热带气旋的路径和强度预报进行研究,发现多模式集成预报效果较单模式预报效果有明显提高。Ji et al.(2020)提出了基于MODE评估的降水多模式集成预报新方法。这种基于目标评估的多模式集成预报明显优于传统的基于点对点的多模式集成预报,不但改进了降水量的预报,还使得预报的降水区更加完整,较好地保留了降水预报的空间相关性。

常见的多模式集成方法,如简单集合平均、消除偏差集合平均、基于线性回归的超级集合等都是线性统计后处理方法,虽然能够提高预报技巧,但仍有不小的改进空间。随着人工智能的热潮席卷全球,深度学习方法被广泛应用到各行各业。循环神经网络(Recurrent Neural Network, RNN)算法始于20世纪80年代,在21世纪初发展成为重要的深度学习算法。循环神经网络的优势在于它能记忆输入序列的信息,并对输入序列及其上下关系进行深度学习和理解,进而准确地做出预报。但是,传统的RNN模型梯度无法在较长的时间序列中传递,从而无法检测到较长序列的影响。LSTM(Long Short-Term Memory)是RNN算法的扩展,在RNN模型的基础上添加“门”的结构,控制信息的传递,从而将RNN的记忆功能扩展,能够运用非线性的方法从长时间的数据中学习有效信息,因而成为气温预报的首选。Shi et al.(2015)将LSTM与CNN相结合进行临近降水预报,也取得了很好的预报效果。Han et al.(2000)研究发现,LSTM神经网络对24 h和72 h的逐时次气温、湿度和风具有较好的预报效果。倪铮和梁萍(2018)利用LSTM神经网络对宝鸡市未来24 h气温做精细化预报,能较好地预报出转折性天气过程,且精细化气温预报准确率相对较高。张庭玉(2018)提出基于神经网络的空气质量预报多模式集成技术,对每个污染物所适用的神经网络集成模型进行分析和比较,找出了能有效预报空气质量的神经网络集成方法。门晓磊等(2019)利用基于岭回归、随机森林和LSTM三种多模式集成技术对4个中心模式华北地区地面气温进行多模式集成预报试验,发现3种后处理方法都能得到优于单个模式的地面气温预报。

本文利用深度学习方法对TIGGE资料中欧洲中期天气预报中心、英国气象局、美国国家环境预报中心、韩国气象厅和日本气象厅5个中心模式1~7 d预报时效的地面气温集合预报进行多模式集成预报试验,评估和比较传统的统计后处理方法及基于深度学习的多模式集成预报方法,期望得到更优的地面气温预报。

1 资料与方法

1.1 资料

1.1.1 模式资料

所用模式资料取自TIGGE资料集下欧洲中期天气预报中心、英国气象局、美国国家环境预报中心、韩国气象厅、和日本气象厅5个中心全球集合预报模式每日12时(世界时,下同)起报的2 m气温各自集合成员的平均。资料长度取2015年1月1日—9月30日,研究区域为中国及其周边地区($70^{\circ}\sim 140^{\circ}\text{E}$, $15^{\circ}\sim 60^{\circ}\text{N}$),模式的水平分辨率为 $1^{\circ}\times 1^{\circ}$,预报时效为24~168 h,间隔为24 h。

1.1.2 “观测”资料

在选择“观测”资料时,应充分考虑资料覆盖的地理范围、准确度及连续性等影响因素。采用站点的观测资料应与模式再分析资料区分开来,这样可以使得多模式集成预报的检验更加客观。但是,将大范围的观测资料插值到格点上也会造成人为的插值误差,且站点资料由于受观测范围和观测要素的限制,对于模式预报检验也存在诸多不足(何诚飞,2016)。ERA-Interim再分析资料为世界公认的准确度较高的资料,许多专家将此作为“观测”资料进行科学研究。Hagedorn et al.(2012)研究表明,2 m气温的ERA-Interim再分析资料远比NCEP再分析资料准确度高。同时,多模式集成预报性能对“观测”资料的选择并不敏感,采用不同“观测”资料对模式预报进行检验,所得检验结果变化较小。所以,经过综合考量,本文选择ERA-Interim再分析资料作为“观测”资料。

地面气温的“观测”资料为逐日12时ERA-Interim再分析资料,并且将此作为参照用于预报评估和多模式集成预报方法的建模。资料长度选取2015年1月1日—10月7日,研究区域为中国及其周边地区($70^{\circ}\sim 140^{\circ}\text{E}$, $15^{\circ}\sim 60^{\circ}\text{N}$),资料的水平分辨率为 $1^{\circ}\times 1^{\circ}$ 。

1.2 方法

1.2.1 消除偏差集合平均

消除偏差集合平均 (Bias-Removed Ensemble Mean, BREM) 定义为:

$$F_{\text{BREM}} = \bar{O} + \frac{1}{N} \sum_{i=1}^N (F_i - \bar{F}_i) \quad (1)$$

其中: 对于每一个格点, F_{BREM} 为消除偏差集合平均预报结果; \bar{O} 为训练期观测值平均; N 为参与集合的模式个数; F_i 为第 i 个模式的预报值; \bar{F}_i 为第 i 个模式预报训练期的时间平均值。采用滑动训练期的方法, 将训练期设置为固定长度并逐日向后滑动, 每次只预报距离训练期临近的一天。训练期窗口不断滑动, 权重不断更新, 能够消除预报偏差, 使得预报准确率更高。最优训练期长度的选取以及滑动训练期的选取方法参见智协飞等 (2009, 2013)。

1.2.2 超级集合预报

超级集合预报运用统计方法, 将资料分为训练期和预报期, 在训练期对模式的预报结果和“观测”数据进行训练建模, 确定参与集成的模式权重系数后, 在预报期进行超级集合预报。超级集合 (SUP) 定义为:

$$F_{\text{SUP}} = \bar{O} + \sum_{i=1}^N a_i (F_i - \bar{F}_i) \quad (2)$$

其中: 对于每一个格点, F_{SUP} 为超级集合的预报结果; \bar{O} 为训练期观测值平均; N 为参与集合的模式个数; a_i 为权重系数; F_i 为第 i 个模式的预报值; \bar{F}_i 为第 i 个模式预报训练期的时间平均值。

在训练期中, 利用式 (3) 计算得到误差项的最小值, 并将得到 a_i 带入预报期的计算之中, 所有格点如此计算就可进行超级集合预报。

$$G = \sum_{i=1}^{N_{\text{train}}} (S_i - O_i)^2 \quad (3)$$

其中: G 为误差项; N_{train} 为训练期天数; O_i 为观测值。

超级集合预报仍旧使用滑动训练期方法 (智协飞等 2009, 2013; Zhi et al. 2012)。

1.2.3 浅层神经网络

神经网络模仿人脑神经系统的组织结构和部分活动机理, 通过大量具有简单功能的神经元互相连接, 形成一种复杂的自适应非线性动态系统, 对处理逻辑操作、非线性问题等具有较大优势。有关神经网络方法更详细的介绍可参考 Nielsen (2015) 的文献。

浅层神经网络由几层节点组成, 每个节点都是输入值和相应权重相乘后, 对所有乘积求和, 再加上该层一个偏置项得到, 即

$$\sum_{i=1}^n w_{ij} x_i + b_j \quad (4)$$

其中: w 为权重; x 为输入的数据; b 为偏置项; i 为上一层神经元个数; j 为当前层神经元个数。

如图 1 第一层为输入层, 例如有三个节点接收数据 x_1, x_2, x_3 , 每个节点利用式 (4) 计算, 通过激活函数 f 激活, 产生下一层的输入值, 即隐藏层的输入值, 其计算方法为

$$a_j^{(1)} = f\left(\sum_{i=1}^n w_{ij} x_i + b_j\right) \quad (5)$$

其中: a 为当前节点的输出值; x 为输入数据; w 为权重; f 为激活函数; b 为偏置项; i 为上一层神经元个数; j 为当前层神经元个数。按图 1 中结构计算得到

$$a_1^{(1)} = f(w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + w_{31}^{(1)} x_3 + b_1^{(1)}) \quad (6)$$

$$a_2^{(1)} = f(w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{32}^{(1)} x_3 + b_2^{(1)}) \quad (7)$$

$$a_3^{(1)} = f(w_{13}^{(1)} x_1 + w_{23}^{(1)} x_2 + w_{33}^{(1)} x_3 + b_3^{(1)}) \quad (8)$$

上标表示所在层数。

第二层为隐藏层, 每个神经元接收上一层输出值作为输入, 再通过激活函数激活输出, 最终得到输出层的输出 y_j 。

$$y_1 = a^{(2)} = f(w_{11}^{(2)} a_1^{(1)} + w_{21}^{(2)} a_2^{(1)} + w_{31}^{(2)} a_3^{(1)} + b_1^{(2)}) \quad (9)$$

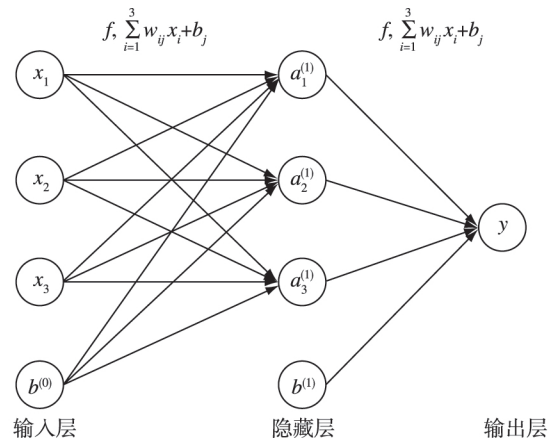


图 1 神经网络结构

Fig.1 The structure of the neural networks

使用的激活函数为 R_{eLU} :

$$R_{\text{eLU}}(z) = \max(0, z) \quad (10)$$

相对于其他激活函数 (例如: sigmoid) R_{eLU} 在计算过程中节省计算量, 不存在梯度消失问题, 使得模型的收敛速度维持在稳定状态。同时, 它也在一定程度上减小了参数的依赖关系, 缓解了过拟合问题。

1.2.4 LSTM 神经网络

1.2.4.1 循环神经网络

循环神经网络(RNN)在浅层神经网络的基础上引入了定向循环,通过神经元记忆之前时刻时间序列变量的信息,传递到下一个时刻的计算中,能更好地建立时间序列变量间非线性的映射,对未来时刻的变化做出预报,尤其是对连续且非线性变化的气象要素,如地面气温等。有关RNN方法的描述可参看顾逸(2018)。

图2是目前循环神经网络的一种典型结构。从图中的结构可以看出,输入一组时间序列(x_1, x_2, x_3, \dots)存在两条传递信息的路径。一条是从下向上传递信息,将数据从输入层传递到隐藏层,另一条由左向右传递信息,将隐藏层的输出值传递到下一时刻作为输入。也就是说,将输入值与输入层权重的乘积分为二条,一条传递到输出层,另一条传递到下一时刻的隐藏层,从过去时间序列中学习有用信息,传递到未来时刻的数据预报中。隐藏层和输出层的输出信息将由如下公式计算得到:

$$h_t = f(u_t x_t + w_{t-1} h_{t-1} + b) \quad (11)$$

$$y_t = v_t h_t \quad (12)$$

其中: h 为隐藏层输出值; f 为激活函数; u 为输入层权重; x 为输入的时间序列数据; w 为上一时刻的数据权重; b 为偏置项; y 为输出层输出值; v 为隐藏层权重; t 为时间。

但是,传统的深度神经网络在隐藏层中很有可能出现梯度消失,导致预报模型收敛速度过慢,训练时间加长,所以RNN在时间序列后面的节点对前面节点的感知力下降,神经网络无法记忆相隔较长时刻的历史数据,极大地降低了预报的准确率。

1.2.4.2 长短期记忆神经网络

在处理与历史数据相关的序列问题时,针对循

环神经网络无法捕捉到较长时刻之前的历史数据,科学家们发展了长短期记忆(LSTM)神经网络。LSTM神经网络是一种RNN变体,通过将较长历史数据保存至后续时刻的方法,加强了神经网络的存储能力。LSTM神经网络使用累加器和“门”结构构成的细胞单元作为主体。细胞单元传递自身状态和外部输入信息,将前一时刻的细胞单元传递至当前时刻细胞单元,并由另一个单元决定何时保留或消除记忆信息。LSTM神经网络可以使用过去时刻的历史数据,在气温预报中,可以在不产生梯度衰减的情况下使用过去时刻气温值,较好地捕捉序列隐藏的关系,并为预报气温打下良好的基础。

LSTM神经网络中包含输入门、输出门、遗忘门、候选细胞、隐含层、细胞单元。

$$f(t) = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (13)$$

$$i(t) = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (14)$$

$$g(t) = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \quad (15)$$

$$o(t) = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (16)$$

$$s_t = f(t) \times s_{t-1} + i(t) \times g(t) \quad (17)$$

$$h_t = o(t) \times \tanh(s_t) \quad (18)$$

LSTM根据式(13—18)这6个核心公式进行计算。 s 为当前时刻细胞; $f(t)$ 为遗忘门方程; $i(t)$ 和 $g(t)$ 共同构成输入门方程,用来更新细胞状态 s ; $o(t)$ 是输出门方程,与 s 一起输出隐含层 h ;输入门、输出门和遗忘门中每个元素的值域都是 $[0, 1]$;候选细胞 $g(t)$ 使用值域为 $(-1, 1)$ 的双曲正切函数(\tanh)作为激活函数; W_f, W_i, W_o, W_g 为权重参数,需要经过学习获得; b_f, b_i, b_o, b_g 为可学习的偏移参数。细胞单元随着图中箭头流动方向计算(图3)。前一时刻细胞单元的信息输出将作为下一时刻细胞单元的输入。方块中为该层使用的激活函数,而圆圈中是每一步进行的运算(如向量的加减乘)。黑

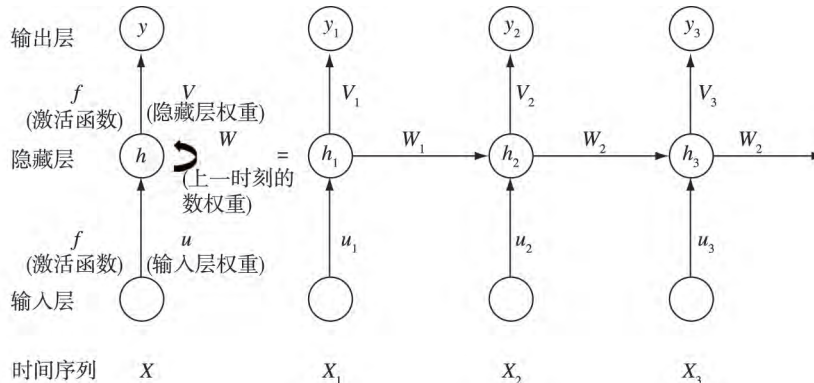


图2 典型的RNN结构示意图

Fig.2 Diagram of the typical RNN structure

线代表信息的流动。

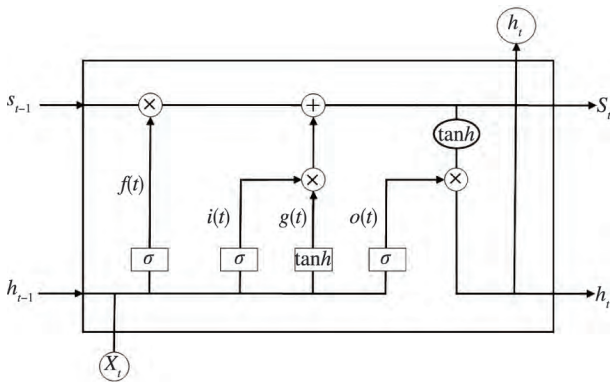


图3 LSTM 内部细胞单元

Fig.3 LSTM internal cell unit

1.2.4.3 数据说明及数据清洗

将数据分为训练集、验证集和测试集 3 个部分,训练集和验证集进行建模,测试集用于实际预报。此方法具有普适性,对于测试期的选取不太敏感。从 2015 年 1 月 1 日—9 月 4 日 ECMWF、UKMO、NCEP、KMA 和 JMA 的 2 m 气温预报值中抽取 90% 进行训练,剩下 10% 进行验证。2015 年 9 月 5—30 日进行测试。

对于 TIGGE 资料中存在的缺测情况,并没有统一的处理方法,可依据不同数据不同要素灵活处理。本文中使用的 2 m 气温要素在时间序列上有一定连续性,所以使用缺测值所在日期前一天的预报值来代替。

1.2.4.4 参数调优

归一化与反归一化:为了使激活函数的梯度不至于过小,加快收敛,且有利于数据在隐藏层的传输,对输入的数据集进行归一化处理,训练后再将训练结果反归一化得到最优值。本文采用最大最小值归一化对数据进行处理。

时间步长:时间步长的选取并没有具体要求,但是随着时间步长的增加,深度学习过程中可能捕捉更多信息和细节,由于本文的资料长度有限,最终选取的时间步长为 1 d。

学习率:学习率过小,会使训练时长增加;学习率过大,会影响最终训练结果的收敛。

隐藏层节点数:隐藏层节点数过多会导致网络模型太复杂,出现过拟合现象;隐藏层节点数太少,不利于数据特征的提取和数据信息的学习。

迭代次数:根据数据量和试验过程中损失函数的下降和收敛情况最终确定。对于本研究数据的复杂程度,训练 600 次即可满足需求。

由于 LSTM 神经网络的参数设置与最终预报结果关系密切,故需要不断调参,使得预报效果最好。同时,考虑到中国及周边地区中的气候特征或地形因素,对所选区域分区调参。重点调参对象是学习率、隐藏层节点数和迭代次数。

1.3 评估方法

均方根误差(RMSE,用 R_{RMSE} 表示)能较好地反映气温预报的预报技巧,其计算公式如下:

$$R_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - Y_i)^2} \quad (19)$$

其中: n 为参与预报的格点数, F_i 为第 i 个样本预报, Y_i 为与之相对应的观测值。

2 地面气温集合预报误差评估

选取均方根误差作为评估地面气温预报技巧的指标,计算 2015 年 9 月 5—30 日 ECMWF、UKMO、NCEP、KMA 和 JMA 5 个中心地面气温预报在中国及周边地区($70^\circ \sim 140^\circ E$, $15^\circ \sim 60^\circ N$) 区域平均(在时间和空间上求平均)均方根误差。图 4 是预报期内 5 个中心模式在中国及周边地区地面气温区域平均的均方根误差。随着预报时效的延长,各中心模式的均方根误差不断增大,但均方根误差的变化幅度并不大,预报 24~168 h 均方根误差增长幅度在 $1^\circ C$ 以内。ECMWF 均方根误差最小;NCEP 均方根误差最大,JMA 次之,且与 ECMWF 相差较大;KMA 与 UKMO 的均方根误差则位居其中。

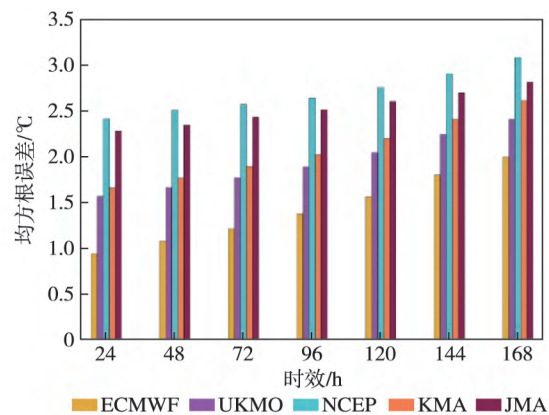


图4 2015年9月5—30日中国及周边地区ECMWF、UKMO、NCEP、KMA和JMA 24~168 h地面气温预报的区域平均均方根误差(R_{RMSE} ,单位: $^\circ C$)

Fig.4 Regional average RMSE (unit: $^\circ C$) of the 24—168 h surface air temperature forecasts of ECMWF, UKMO, NCEP, KMA and JMA in China and its adjacent area from September 5 to 30 2015

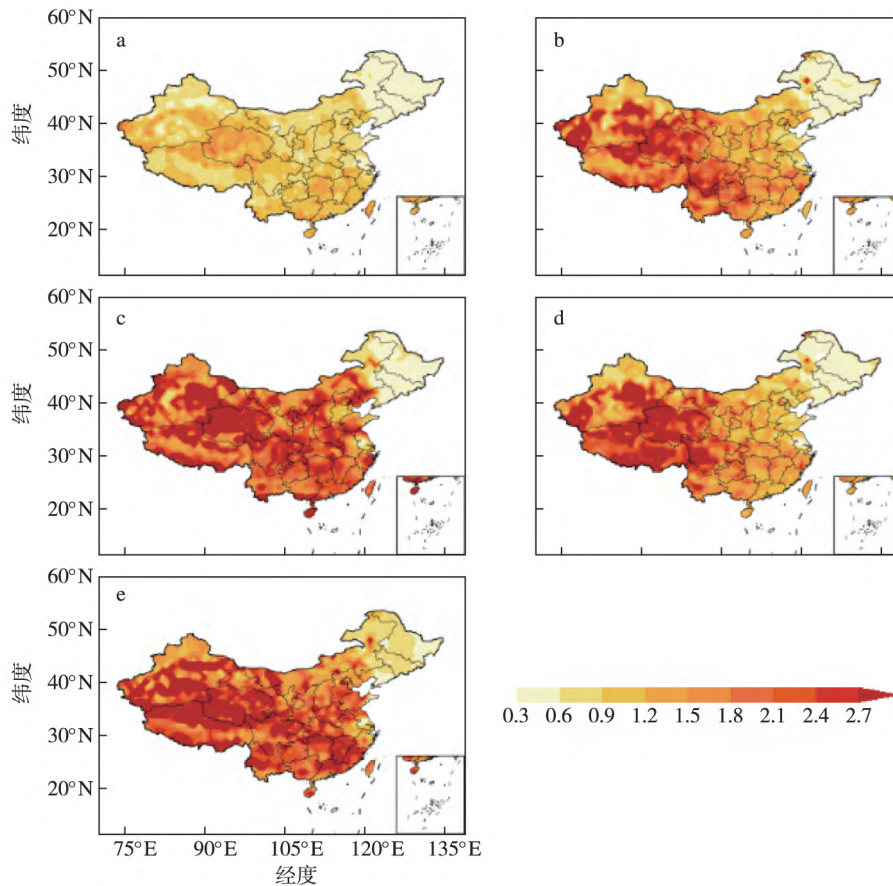


图5 2015年9月5—30日ECMWF(a)、UKMO(b)、NCEP(c)、KMA(d)和JMA(e)中国地区24h地面气温预报的时间平均均方根误差分布(单位: °C)

Fig.5 Geographical distribution of the RMSE(unit: °C) of 24 h surface air temperature forecasts of (a) ECMWF (b) UKMO (c) NCEP (d) KMA and (e) JMA models from September 5 to 30 2015 in China

图5是中国区域预报期内ECMWF、UKMO、NCEP、KMA和JMA 24h地面气温预报的时间平均的均方根误差地理分布。ECMWF在5个模式中误差整体最小,西部的均方根误差略大于东部的均方根误差;UKMO和KMA均方根误差处于中间水平,西部地区的均方根误差明显大于东部地区的均方根误差;NCEP和JMA预报效果相比其他模式整体较差,均方根误差整体偏大,均方根误差在2.5 °C以上。综上,在预报期内,对中国及周边地区地面气温的预报,ECMWF的预报效果最好。当然,ECMWF的预报效果好可能和用于预报评估的“观测”资料来源于ECMWF ERA-Interim再分析资料也有一定关系,特别是在缺少实测资料的戈壁和沙漠地区,ERA-Interim再分析资料对ECMWF预报评估的影响更大。同时,各个中心模式的预报效果区别较大,并且在地理分布上存在明显的差异。因此,我们尝试应用多模式集成技术,并将模式预报技巧最高的ECMWF与传统的多模式集成、浅层神经

网络和LSTM神经网络进行比较。

3 多种方案的地面气温多模式集成预报试验

将单模式中预报效果最好的ECMWF与传统多模式集成方法、浅层神经网络和LSTM集成预报进行比较。图6为2015年9月5—30日中国及周边地区地面气温ECMWF、SUP、BREM、NN和LSTM不同预报时效的区域平均的均方根误差。BREM均方根误差相比于ECMWF略高,但是比其他单模式预报的均方根误差小,主要原因是BREM对5个单模式进行等权计算,因此易受预报效果较差模式的影响;SUP的均方根误差明显低于BREM和最好的单模式ECMWF的均方根误差,其中24h预报的均方根误差比ECMWF的低0.15 °C左右。但随着预报时效延长,二者差别越来越小,到168h, SUP和ECMWF均方根误差相差无几。24~144h预报, NN的均方根误差略大于SUP的,168h预报,

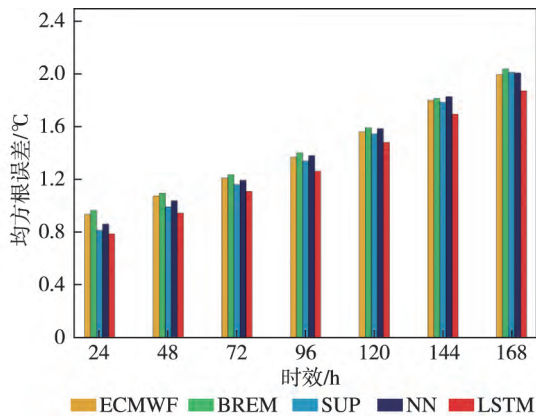


图6 2015年9月5—30日中国及周边地区地面气温ECMWF、SUP、BREM、NN和LSTM 24~168 h预报的区域平均均方根误差(单位: °C)

Fig.6 Regional average RMSE (unit: °C) of 24—168 h surface air temperature forecasts of ECMWF, SUP, BREM, NN and LSTM in China and its adjacent area from September 5 to 30, 2015

NN和SUP的均方根误差非常接近。相比ECMWF、NN和SUP的预报,LSTM预报具有明显优势,24~168 h的预报效果明显优于ECMWF单模式预报及BREM、NN和SUP多模式集成预报。从中还可以看出,预报时效超过96 h,LSTM预报的优势更加明显。

为此,图7选取预报期内ECMWF、SUP、BREM、NN和LSTM在预报时效为96 h地面气温预报的平均均方根误差进行比较分析。从图7中可以看出,青海、四川、重庆及湖南部分地区地面气温BREM预报的均方根误差较小,但在新疆南部均方根误差很大。BREM在南疆误差较大的主要原因是,ECMWF在南疆的均方根误差较小,约在2 °C以下,而其他4个中心的模式在南疆的均方根误差均远大于ECMWF在此地的均方根误差。特别是NCEP的预报,在此地的均方根误差达到4.5 °C以上。而BREM方法对单模式的等权计算,使多模式集成预报结果受NCEP等误差较大的预报模式的影响较大。SUP和NN预报的均方根误差较ECMWF的均方根误差在我国大部分地区更小,误差大值区明显缩小。LSTM预报比ECMWF预报均方根误差明显更小,在西北、西南、华南等地误差显著减小。

为更加突出地反映LSTM预报的优越性,计算了LSTM预报比ECMWF、BREM、SUP和NN预报在每个格点的均方根误差减小百分比(图8)及总体均方根误差减小的格点占比(LSTM相对于其他方法的预报技巧有所提高的格点数目占所选区域内总

格点数的百分比,即图8中暖色区域占中国区域的百分比)。从图8中可以看出,LSTM预报比ECMWF预报误差减小的格点占总格点数的71%,西北、西南、华南、华北北部和东北大部分地区均方根误差减小了10%以上,部分地区均方根误差减小了50%以上。但是在塔里木盆地及周边地区LSTM预报均方根误差较大,甚至超过了ECMWF预报的均方根误差。LSTM预报比BREM预报均方根误差减小的格点占总格点数的75%,在西北、西南及华南地区误差明显减小。LSTM预报相比SUP预报,其误差减小不如前两种预报明显,误差减小的格点占总格点数的69%左右,误差减小了14%~50%,其中在西北、西南、华南的误差减小明显。LSTM预报较NN预报误差减小的格点占总格点数的72%,但均方根误差减小的幅度不大,西北地区预报技巧改善较明显。

图9为2015年9月5—30日ECMWF、BREM、SUP、NN和LSTM 24~168 h地面气温预报的区域平均均方根误差。随着预报时效的延长,不同多模式集成方案及ECMWF单模式预报的均方根误差递增。总体上,SUP预报的均方根误差比ECMWF预报的均方根误差小。BREM比SUP预报均方根误差略大,和ECMWF预报相比,均方根误差没有明显改善。总体上,LSTM预报技巧最高,均方根误差最小。LSTM比SUP 24~72 h预报的均方根误差改进不明显,但对96~168 h预报有明显改进,最大改进达0.5 °C左右。同时,BREM、SUP、NN和LSTM预报的均方根误差趋势和单个模式ECMWF大致相同,在预报后期,96~168 h预报的均方根误差都有增大的趋势。因此,在本文建立的LSTM模型中,将5个单个模式的预报值作为特征,“观测”值作为目标,不断对地面气温的预报值进行订正和拟合,达到减小系统误差的目的,最终得到的预报值的变化趋势也在一定程度上依赖于单个模式。

4 结论和讨论

运用滑动训练期消除偏差集合平均、多模式超级集合、浅层神经网络和长短期记忆神经网络(LSTM)4种方法对ECMWF、NCEP、UKMO、KMA和JMA 5个中心的1~7 d地面气温集合预报进行多模式集成,比较了4种方法的预报技巧,得到以下几点结论。

1) 滑动训练期消除偏差集合平均对5个单模式进行等权计算,易受预报效果较差模式的影响,虽

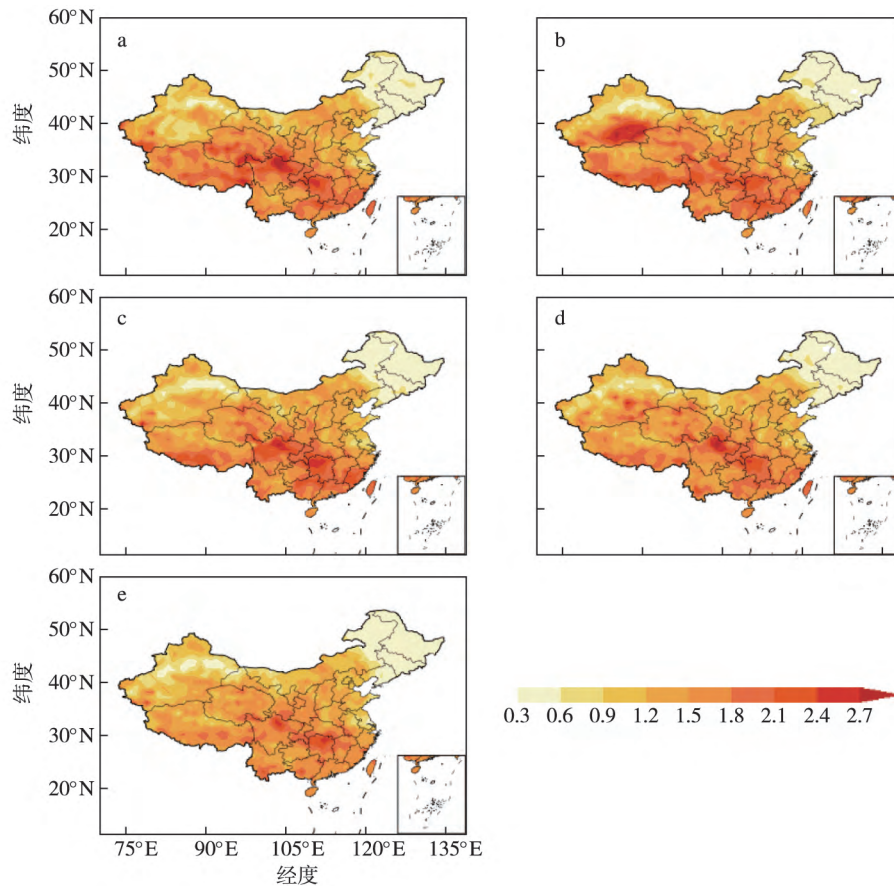


图7 2015年9月5—30日中国地区ECMWF(a)、BREM(b)、SUP(c)、NN(d)和LSTM(e)预报时效为96h地面气温预报的时间平均均方根误差分布(单位: °C)

Fig.7 Geographical distribution of the RMSE (unit: °C) of 96 h surface air temperature forecasts of (a) ECMWF, (b) BREM, (c) SUP, (d) NN and (e) LSTM in China from September 5 to 30, 2015

然预报效果总体上优于大部分单个模式的预报效果,但不如单个模式中预报效果最好的ECMWF预报,特别是在新疆南部,在等权计算之后的预报值仍存在较大误差。

2) 滑动训练超级集合的预报技巧优于单个模式的预报技巧,在144h之前,超级集合预报误差明显小于ECMWF的预报误差。

3) 基于LSTM神经网络的多模式集成预报效果最优,特别是在预报时效较长(超过72h)时,相

比其他方法其误差明显降低。LSTM神经网络方法明显改进我国西北、华北、东北、西南和华南大部分地区的气温预报技巧,但在南疆部分地区误差较大。

鉴于LSTM神经网络方法在预报时效较长(超过96h)时预报效果较其他多模式集成预报方法具有明显优势,因此对中期预报、延伸期预报可以选用LSTM神经网络方法。另外,在极端高、低温天气的预报上,传统的多模式集成方法尚有一些不足,也可尝试使用LSTM神经网络方法进行改进。

参考文献(References)

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[J].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 [J]. Acta Meteorol Sin, 68(4): 464-476. (in Chinese).

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(2):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]. Trans Atmos Sci, 36(2): 165-173. (in Chinese).

顾逸.2018.基于长短期记忆循环神经网络及其结构约减变量的中长期径流预报研究[D].武汉:华中科技大学. Gu Y. 2018. Research on mid-

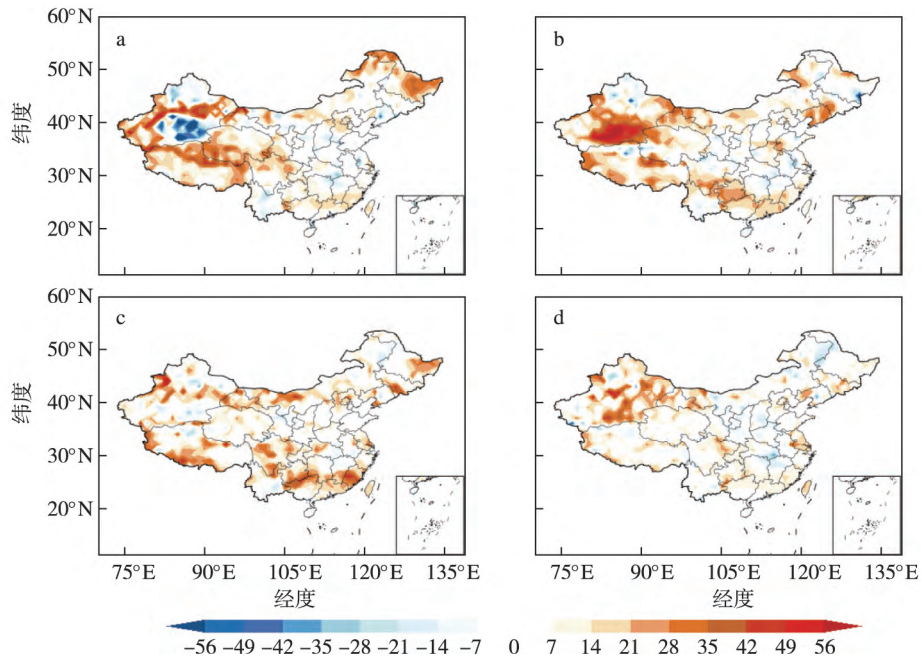


图 8 2015 年 9 月 5—30 日与 LSTM 比 ECMWF(a)、BREM(b)、SUP(c)、NN(d) 分别在预报时效为 96 h 的中国地面气温时间平均的均方根误差减小的百分比分布(单位: %)

Fig.8 Geographical distribution of forecast RMSE decrease rate of LSTM 96 h forecast compared with (a) ECMWF (b) BREM (c) SUP (d) NN forecasts from September 5 to 30 2015 in China(units: %)

long-term runoff forecasting based on long-short-term memory recurrent networks and its structural reduction variant [D]. Wuhan: Huazhong University of Science and Technology. (in Chinese).

Hagedorn R, Buizza R, Hamill T M, et al. 2012. Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts [J]. *Quart J Roy Meteor Soc*, 138(668): 1814-1827.

Han J W, Pei J, Yin Y W. 2000. Mining frequent patterns without candidate generation [C] // Proceedings of the 2000 ACM SIGMOD international conference on management of data. New York: ACM Press.

何诚飞. 2016. 基于卡尔曼滤波的地面与 500 hPa 多模式集成预报对比研究 [D]. 南京: 南京信息工程大学. He C F. 2016. Comparison of multimodel ensemble forecasts at surface and 500 hPa based on the Kalman filter method [D]. Nanjing: Nanjing University of Information Science & Technology. (in Chinese).

He C F, Zhi X F, You Q L, et al. 2015. Multi-model ensemble forecasts of tropical cyclones in 2010 and 2011 based on the Kalman Filter method [J]. *Meteorol Atmos Phys*, 127(4): 467-479.

Ji L Y, Zhi X F, Zhu S P, et al. 2019. Probabilistic precipitation forecasting over east Asia using Bayesian model averaging [J]. *Wea Forecasting*, 34(2): 377-392.

Ji L Y, Zhi X F, Simmer C, et al. 2020. Multi-model ensemble forecast of precipitation based on an object-based diagnostic evaluation [J]. *Mon Wea Rev*, 148(6): 2591-2606.

Krishnamurti T N. 1999. Improved weather and seasonal climate forecasts from multimodel superensemble [J]. *Science*, 285(5433): 1548-1550.

Krishnamurti T N, Kishtawal C M, Shin D W, et al. 2000a. Improving tropical precipitation forecasts from a multianalysis superensemble [J]. *J Climate*, 13(23): 4217-4227.

Krishnamurti T N, Kishtawal C M, Zhang Z, et al. 2000b. Multimodel ensemble forecasts for weather and seasonal climate [J]. *J Climate*, 13(23): 4196-4216.

Krishnamurti T N, Gnanaseelan C, Chakraborty A. 2007. Prediction of the diurnal change using a multimodel superensemble. Part I: precipitation [J]. *Mon Wea Rev*, 135(10): 3613-3632.

Krishnamurti T N, Sagadevan A D, Chakraborty A, et al. 2009. Improving multimodel weather forecast of monsoon rain over China using FSU superensemble [J]. *Adv Atmos Sci*, 26(5): 813-839.

Krishnamurti T N, Kumar V, Simon A, et al. 2016. A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes [J]. *Rev Geophys*, 54(2): 336-377.

林春泽, 智协飞, 韩艳, 等. 2009. 基于 TIGGE 资料的地面气温多模式超级集合预报 [J]. *应用气象学报*, 20(6): 706-712. Lin C Z, Zhi X F, Han

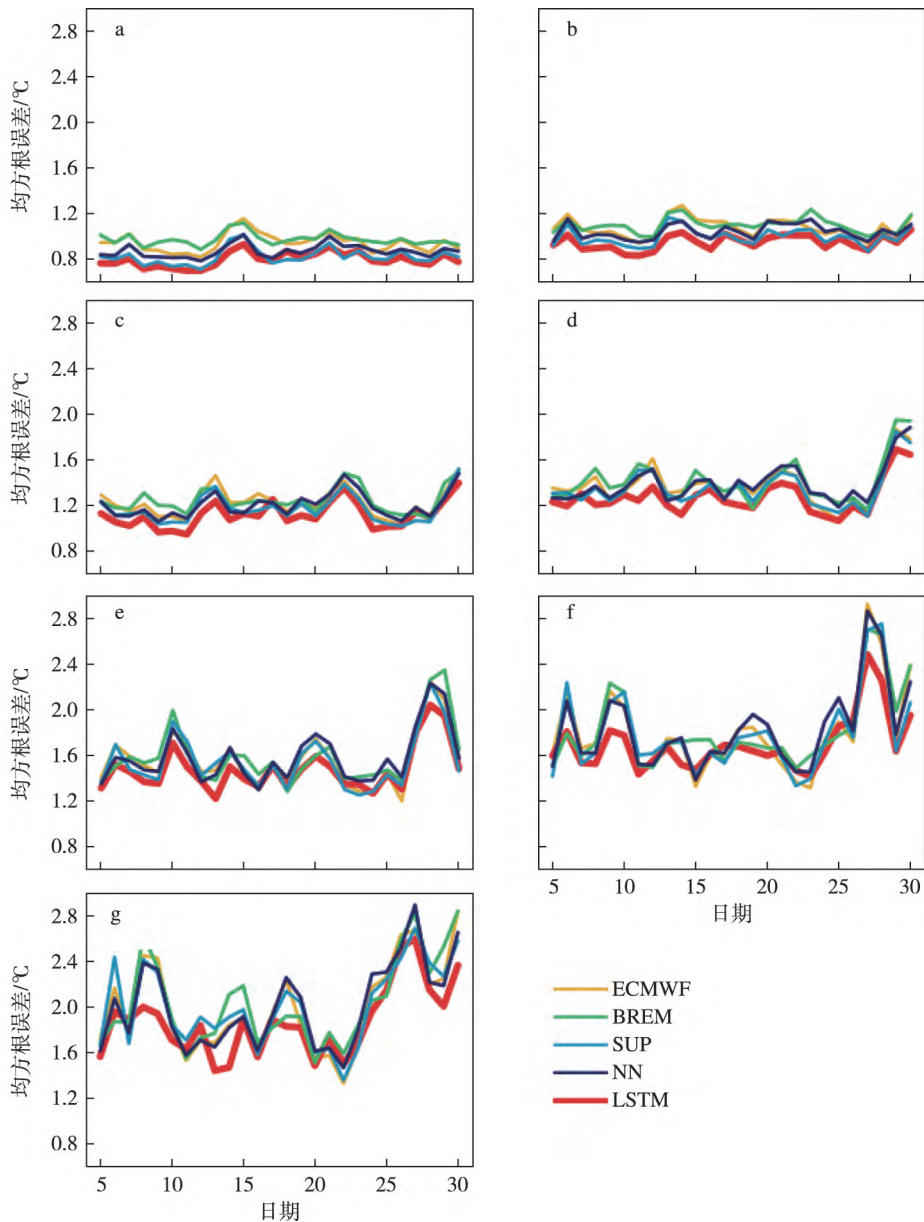


图9 2015年9月5—30日中国及周边地区的ECMWF、BREM、SUP、NN、LSTM 24 h (a)、48 h (b)、72 h (c)、96 h (d)、120 h (e)、144 h (f)和168 h (g)地面气温预报的区域平均均方根误差(单位: °C)

Fig.9 The RMSE (unit: °C) of the surface air temperature forecasts of ECMWF, SUP, BREM, NN and LSTM with (a) 24 h, (b) 48 h, (c) 72 h, (d) 96 h, (e) 120 h, (f) 144 h and (g) 168 h forecast lead time in China and its adjacent area from September 5 to 30, 2015

Y et al. 2009. Multimodel superensemble forecasts of the surface temperature using the TIGGE data [J]. J Appl Meteor Sci, 20(6): 706-712. (in Chinese).

林春泽, 祁海霞, 智协飞, 等. 2013. 中国夏季降水多模式集成概率预报研究[J]. 暴雨灾害, 32(4): 354-359. Lin C Z, Qi H X, Zhi X F, et al., 2013. Study on multi-model ensemble probability forecast for summer precipitation in China [J]. Torrential Rain and Disasters, 32(4): 354-359. (in Chinese).

马清, 龚建东, 李莉, 等. 2008. 超级集合预报的误差订正与集成研究[J]. 气象, 34(3): 42-48. Ma Q, Gong J D, Li L, et al., 2008. Study of bias-correction and consensus in regional multi-model super-ensemble forecast [J]. Meteor Mon, 34(3): 42-48. (in Chinese).

门晓磊, 焦瑞莉, 王鼎, 等. 2019. 基于机器学习的华北气温多模式集合预报的订正方法[J]. 气候与环境研究, 24(1): 116-124. Men X L, Jiao R L, Wang D, et al., 2019. A temperature correction method for multi-model ensemble forecast in North China based on machine learning [J]. Clim Environ Res, 24(1): 116-124. (in Chinese).

- Nielsen M A 2015. Neural networks and deep learning [EB/OL]. [2020-03-26]. <http://neuralnetworksanddeeplearning.com>.
- 倪铮, 梁萍 2018. 基于 LSTM 深度神经网络的精细化气温预报初探 [J]. 计算机应用与软件, 35(11): 233-236, 271. Ni Z, Liang P 2018. Fine temperature forecast based on lstm deep neural network [J]. Comput Appl Softw, 35(11): 233-236, 271. (in Chinese).
- Shi X J, Chen Z R, Wang H et al. 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [EB/OL]. <https://arxiv.org/abs/1506.04214>.
- 杨学胜 2001. 业务集合预报系统的现状及展望 [J]. 气象, 27(6): 39. Yang X S 2001. The new development and the outlook of the operational ensemble prediction system [J]. Meteor Mon, 27(6): 39. (in Chinese).
- 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(9): 1058-1067. Zhang H B, Zhi X F, Wang Y N et al. 2015. Comparison of multimodel ensemble methods for western Pacific tropical cyclone forecast based on TIGGE data [J]. Meteor Mon, 41(9): 1058-1067. (in Chinese).
- 张玲, 智协飞 2013. 2008 年初中国南方冰冻雨雪天气的多模式集成预报 [J]. 热带气象学报, 29(3): 393-402. Zhang L, Zhi X F 2013. Multimodel ensemble forecasting of low temperature and icy weather over central and Southern China during early 2008 [J]. J Trop Meteor, 29(3): 393-402. (in Chinese).
- 张庭玉 2018. 基于神经网络的空气质量预报多模式集成技术研发与应用 [D]. 南京: 南京信息工程大学. Zhang T Y 2018. Research and application on multi-model ensemble methods for forecast of air quality based on neural network [D]. Nanjing: Nanjing University of Information Science & Technology. (in Chinese).
- 智协飞, 黄闻 2019. 基于卡尔曼滤波的中国区域气温和降水的多模式集成预报 [J]. 大气科学学报, 42(2): 197-206. Zhi X F, Huang W 2019. Multimodel ensemble forecasts of surface air temperature and precipitation over China by using Kalman filter [J]. Trans Atmos Sci, 42(2): 197-206. (in Chinese).
- 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(5): 569-574. Zhi X F, Lin C Z, Bai Y Q et al. 2009. Superensemble forecasts of the surface temperature in Northern Hemisphere middle latitudes [J]. Sci Meteor Sin, 29(5): 569-574. (in Chinese).
- 智协飞, 伍清, 白永清 等 2010. 基于 IPCC-AR4 模式资料的地面气温超级集合预测 [J]. 气象科学, 30(5): 708-714. Zhi X F, Wu Q, Bai Y Q et al. 2010. The multimodel superensemble prediction of the surface temperature using the IPCC AR4 scenario runs [J]. Sci Meteor Sin, 30(5): 708-714. (in Chinese).
- Zhi X F, Zhang L, Bai Y Q 2011. Application of the multi-model ensemble forecast in the QPF [C] // International Conference on Information Science and Technology, March 26-28, 2011. Nanjing: 657-660.
- Zhi X F, Qi H X, Bai Y Q et al. 2012. A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data [J]. Acta Meteorol Sin, 26(1): 41-51.
- 智协飞, 季晓东, 张璟 等 2013. 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, 36(3): 257-266. Zhi X F, Ji X D, Zhang J et al. 2013. Multimodel ensemble forecasts of surface air temperature and precipitation using TIGGE datasets [J]. Trans Atmos Sci, 36(3): 257-266. (in Chinese).
- 智协飞, 张璟, 段晚锁 2015a. 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(4): 767-776. Zhi X F, Zhang J, Duan W S 2015a. Application of sensitive area for target observation associated with El Niño-Southern Oscillation predictions to multimodel ensemble forecast of the tropical Pacific sea surface temperature [J]. Chin J Atmos Sci, 39(4): 767-776. (in Chinese).
- 智协飞, 孙晶, 周文友 2015b. 2009 年夏季西太平洋台风的集合预报和多模式集成预报试验 [J]. 大气科学学报, 38(5): 633-640. Zhi X F, Sun J, Zhou W Y 2015b. Ensemble and multimodel ensemble forecasts of western Pacific typhoons during summer 2009 [J]. Trans Atmos Sci, 38(5): 633-640. (in Chinese).
- 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(1): 64-71. Zhi X F, Zhao H, Zhu S P et al. 2016. Superensemble hindcast of surface air temperature using CMIP5 multimodel data [J]. Trans Atmos Sci, 39(1): 64-71. (in Chinese).

Multimodel ensemble forecasts of surface air temperature over China based on deep learning approach

ZHI Xiefei ,WANG Tian ,JI Yan

Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters(CIC-FEMD) /Key Laboratory of Meteorological Disasters , Ministry of Education(KLME) /College of Atmospheric Sciences ,Nanjing University of Information Science and Technology ,Nanjing 210044 ,China

Based on the 1—7 days ensemble forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF) ,US National Centers for Environmental Prediction (NCEP) ,and the Japan Meteorological Agency (JMA) ,the UK Met Office (UKMO) as well as the Korea Meteorological Administration (KMA) in the TIGGE datasets ,the multimodel ensemble forecasts of the surface air temperature in China and its adjacent area during the period from 1 January to 30 September 2015 were conducted by using long-term memory (LSTM) neural networks ,neural networks (NN) ,bias-removed ensemble mean (BREM) and the superensemble (SUP) with sliding training period for the forecast period from September 5 to 30 ,2015. The results showed that the BREM forecast was no better than the ECMWF forecast due to the impact of low skill model forecasts among the five models. The forecast skill of SUP was better than that of all the single models. For 24–144h forecasts ,the root mean square error (RMSE) of SUP was significantly smaller than that of ECMWF forecast. As the forecast lead-time increased ,the RMSE increased as well. The forecast skill of NN was roughly equivalent to that of SUP. Overall ,the LSTM approach showed the best forecast performance ,especially when the forecast lead-time was longer than 72 h ,the RMSE of the LSTM forecast was considerably smaller than that of ECMWF ,BREM ,NN and SUP forecasts. The LSTM neural networks approach significantly reduced the forecast RMSE of the surface air temperature in the northwestern ,northern ,northeastern ,southwestern ,and southern China. However ,the RMSE of the LSTM forecast was relatively larger in southern Xinjiang area compared with ECMWF forecast.

deep learning; artificial intelligence; multimodel ensemble; LSTM; surface air temperature

doi: 10. 13878/j.cnki.dqkxxb.20200219003

(责任编辑: 袁东敏)