

方鸿斌,王珊珊,王晓玲,等,2024. 基于机器学习的格点气温预报订正方法[J]. 气象,50(1):103-114. Fang H B, Wang S S, Wang X L, et al, 2024. Gridded temperature forecast correction method based on machine learning[J]. Meteor Mon, 50(1):103-114(in Chinese).

## 基于机器学习的格点气温预报订正方法\*

方鸿斌<sup>1</sup> 王珊珊<sup>1</sup> 王晓玲<sup>1</sup> 谭江红<sup>2</sup> 鲁礼炳<sup>3</sup>

1 武汉中心气象台, 武汉 430074

2 湖北省襄阳市气象局, 襄阳 441022

3 湖北省荆州市气象局, 荆州 434022

**提 要:** 使用 2017 年 9 月至 2021 年 3 月国家级业务化运行的智能网格实况分析产品和欧洲中期天气预报中心全球模式(EC)产品,根据湖北省的地理分布特征构建 6 个分区,采用基于 LightGBM 机器学习算法建立的气温预报方法,生成湖北省  $0.05^{\circ} \times 0.05^{\circ}$  格点气温预报产品。利用 2021 年 4—9 月的预报产品和格点实况资料进行检验,结果表明:基于机器学习的气温预报方法(MLT)取得了较好的预报效果,其在 0~72 h 时效内优于中央气象台下发的气温精细化指导预报(SCMOC)和 EC 产品;MLT 在山区的误差较平原大,但山区的订正幅度大于平原,日最高气温的订正幅度大于日最低气温的订正幅度;4—9 月 MLT、SCMOC、EC 产品的平均绝对误差(MAE)日变化都呈现了白天偏高、夜间偏低、午后凸起的单峰特征,MLT 的 MAE 值较 SCMOC 和 EC 产品的更低,并且在转折性天气中仍具有优势;站点检验与格点检验结论一致,基于格点建模的气温预报产品对站点预报同样得到了订正。机器学习在格点气温的模式订正方面可以作为一个行之有效的手段。

**关键词:** 格点气温,机器学习,特征选择,分区建模

**中图分类号:** P456

**文献标志码:** A

**DOI:** 10.7519/j.issn.1000-0526.2023.111701

## Gridded Temperature Forecast Correction Method Based on Machine Learning

FANG Hongbin<sup>1</sup> WANG Shanshan<sup>1</sup> WANG Xiaoling<sup>1</sup> TAN Jianghong<sup>2</sup> LU Libing<sup>3</sup>

1 Wuhan Central Meteorological Observatory, Wuhan 430074

2 Xiangyang Meteorological Office of Hubei Province, Xiangyang 441022

3 Jingzhou Meteorological Office of Hubei Province, Jingzhou 434022

**Abstract:** Using the smart grid reality analysis product (CLDAS-V2.0) and the European Centre for Medium-Range Weather Forecast Global Model (EC) product, which were operated at the national level from September 2017 to March 2021, six regions are constructed according to the geographical distribution characteristics of Hubei Province. The temperature forecast model established by LightGBM machine learning algorithm is used to generate  $0.05^{\circ} \times 0.05^{\circ}$  gridded temperature forecast products in Hubei Province, and the forecast products are verified by the forecast products and grid data from April to September 2021. The results show that the temperature prediction method (MLT) based on machine learning has achieved a good forecast effect, and MLT is superior to SCMOC and EC model products in 0—72 h lead time. The error of MLT in mountain area is larger than that in plain area, but the correction amplitude of MLT in mountain area is larger than that in plain area, and the correction amplitude of daily maximum temperature is larger

\* 湖北省气象局科技支撑项目(2022Y01)、中国气象局重点创新团队(CMA2022ZD04)共同资助

2022 年 11 月 27 日收稿; 2023 年 11 月 17 日收修定稿

第一作者:方鸿斌,主要从事天气预报客观算法研究. E-mail: fhbky2010@163.com

通讯作者:王珊珊,主要从事短时短期天气预报分析研究. E-mail: 52862776@qq.com

than that of daily minimum temperature. The diurnal variations of MAE in MLT, SCMOC and EC model products from April to September present the characteristics of higher in daytime, lower over night, and convex single-peak in afternoon. The MAE value of MLT is lower than that of SCMOC and EC model products, and still has an advantage in changeable weather. The results of site test and grid test are consistent, and the temperature forecast product based on grid modeling is also corrected. Machine learning can be used as an effective means to correct the pattern of gridded temperature.

**Key words:** gridded temperature, machine learning, feature selection, partition modeling

## 引 言

气温的高分辨率格点预报是精细化预报业务中的一项重要工作,提高格点气温预报的准确率是气象预报业务智能化、客观化的发展要求,是气象工作“观测精密、预报精准、服务精细”目标的内在要求。近年来,在国内外大力发展无缝隙精细化预报的背景下,我国的格点实况分析场产品也得到了长足的发展(师春香等,2019;俞剑蔚等,2019)。2017年7月国家级格点实况分析产品(CLDAS-V2.0),正式投入业务化运行。与此同时,机器学习在气象预报领域的应用也正在蓬勃发展,国际上已经有越来越多的机器学习方法应用在天气预报领域中(Haupt et al,2021;Kashinath et al,2021)。欧洲中期天气预报中心在2021年初发布的未来十年发展规划(Machine learning at ECMWF: A roadmap for the next 10-years, <https://www.ecmwf.int/en/library/81207-machine-learning-ecmwf-roadmap-next-10-years>)中阐明了未来机器学习技术在其数值天气预报流程中的重要性。我国学者在气象领域应用机器学习技术上也进行了诸多探讨和实践(许小峰,2018;李扬等,2021;周康辉等,2021a,杨绚等,2022)。通过预报员主观手动订正千米级分辨率的气温网格预报难以做到准确高效,而利用高分辨率格点实况产品,通过机器学习算法订正精细化的格点预报产品是未来的一大趋势。

国内常见的气温模式后处理订正方法有模式输出统计(MOS)、卡尔曼滤波、人工神经网络、站点订正值向格点传递法、最优集合预报订正法、滑动平均和空间误差综合订正技术、机器学习订正方法等(潘留杰等,2017;雷彦森等,2018;郝翠等,2019;薛湛彬等,2019;门晓磊等,2019;任萍等,2020;陈昱文等,2020)。以上方法基本是针对站点的预报订正,订正后的2 m气温有不同程度的改进,但优势不够突

出。基于本地业务基础和天气特点,也有多种后处理方法。如,吴启树等(2017)发展了最优TS评分订正算法,王建鹏等(2018)研发了动态交叉最优要素预报方法,盛春岩等(2020)选择最优的数值预报产品,分析不同客观释用方法预报效果,研究适合的最优集成预报方法,生成气温智能网格最优集成预报产品。这些方法考虑了本地地理、气候等特征以及模式偏差,但仍然没能很好地解决实况分辨率粗的问题,并且方法复杂不易推广,没能充分利用数值模式环流预报较为准确的优势,中高层物理量对气温的影响考虑不足。而针对气温预报,除了需要考虑天气因子的影响外,地理特征、日变化的特征以及模式固有的偏差都是非常重要的。因此本文提出,结合地理特征、日变化特征、融合多种影响气温预报的物理因子及模式偏差,基于LightGBM(light gradient boosting machine)算法建立湖北省格点气温预报方法,将该方法生成的温度预报产品命名为MLT(matching learning temperature)。

## 1 数据资料

### 1.1 数 据

本文资料采用的是2017年9月至2021年9月 $28.9^{\circ}\sim 33.4^{\circ}\text{N}$ 、 $108.25^{\circ}\sim 116.25^{\circ}\text{E}$ 范围的EC模式数据、CLDAS-V2.0逐小时格点气温实况数据及中央气象台下发的气温格点预报产品(SCMOC)。EC模式数据时间分辨率为3 h,空间分辨率为 $0.125^{\circ}\times 0.125^{\circ}$ ;格点气温实况的时间分辨率为1 h,空间分辨率为 $0.05^{\circ}\times 0.05^{\circ}$ (约5 km);SCMOC时间分辨率为1 h,空间分辨率为 $0.05^{\circ}\times 0.05^{\circ}$ 。

### 1.2 数据处理

本文采用线性插值方法分别在时间、空间上做了插值。提取逐小时格点实况数据与数值预报要

素,并进行时空对齐,去掉缺失、存在异常的样本,形成机器学习样本数据库。为了使建模的特征因子更接近真实值,选取 08 时和 20 时起报的 0~11 h 时效 EC 模式预报产品作为样本。本文采用的 EC 模式没有 24 h 高温、低温产品,所以在检验时,采用的是由逐 3 h 高温、低温统计出来的 24 h 高温、低温。基于气候差异和地形特征,对湖北省进行地理分区,6 个分区示意图如图 1。28.9°~33.4°N、108.25°~116.25°E 范围内 6 个矩形块分别为鄂西南、鄂中南(江汉平原南部)、鄂东南、鄂西北、鄂中北(江汉平原北部)、鄂东北。本文对数据进行裁剪,应用机器学习算法,针对 6 个分区分别建立预报模型。

## 2 方 法

### 2.1 机器学习算法

机器学习中,常见的算法是决策树类(树模型)。该算法具有可解释性较强,在天气预报中的连续性变量(气温、湿度、气压)预报效果出色,不亚于深度学习,具有训练更高效的特点(杨璐等,2021)。近年在树模型算法中梯度提升算法 (gradient boosting decision tree, GBDT) 较为流行,例如 XGBoost (extreme gradient boosting)算法、LightGBM 算法。

LightGBM 是一种基于梯度提升树的机器学习方法,是 2016 年由微软公司开发的一种比较快速的梯度提升框架,较 XGBoost 有明显优势。LightGBM 是对 GBDT 的高效实现,原理上它和 GBDT 及 XGBoost 类似,都采用损失函数的负梯度作为当前决策树的残差近似值,去拟合新的决策树。XGBoost 的不足在于,计算信息增益需要扫描所有样本,从而找到最优划分点;在面对大量数据或者特征维度很高时,其效率和扩展性很难使人满意。考虑到本文采用的数据体量较大,训练样本较多,而 LightGBM 利用了单边梯度采样算法和互斥特征绑定算法,能够很好地解决上述问题(Ke et al,2017),因此本文选择 LightGBM 机器学习算法。

### 2.2 特征工程

机器学习模型的好坏需要建立在对天气物理规律清晰、明确的认识基础之上(周康辉等,2021b)。正确的特征选择可以降低机器学习任务的难度,减少计算和存储开销,同时优化学习模型(孙全德等,2019;Pan et al,2019,Wang et al,2021)。因此,本文根据预报员经验初选出能够反映气团冷暖、日照辐射、日变化等 56 种特征因子,通过机器学习(统计分析)特征重要性排序,综合考虑硬件环境、业务运行耗时及检验评估结果,进一步筛选 38 个特征因子

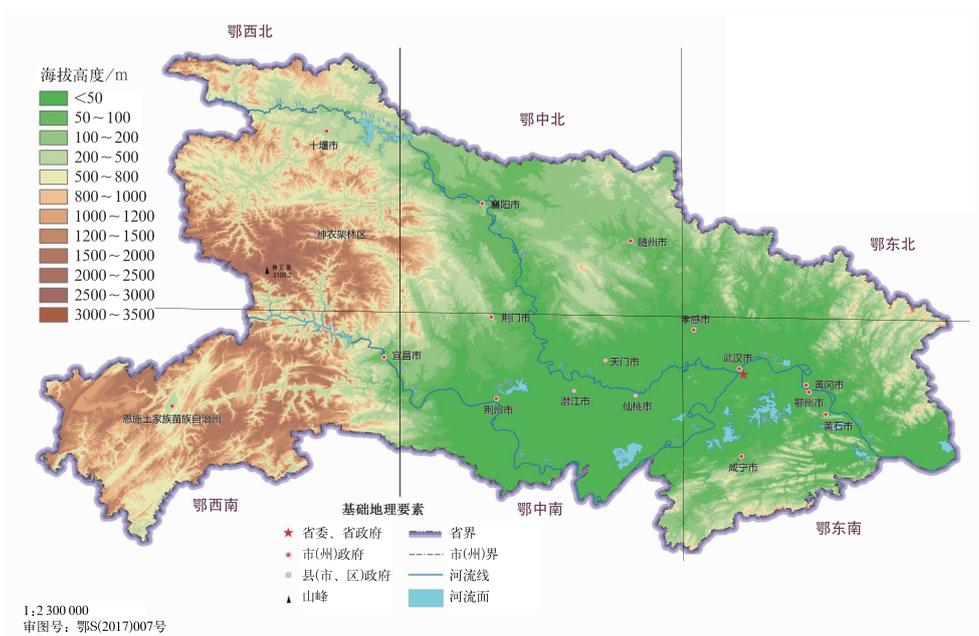


图 1 湖北省 6 个分区示意图

Fig.1 Schematic diagram of research regional division of Hubei Province

形成最优因子集进行建模。

构建模型使用的要素主要有预报场的各层风场、湿度场、温度场、海平面气压场以及平均位势高度(用 500 hPa 场代替)。本文选取的风场数据层次包括 10 m、925 hPa、850 hPa、700 hPa,一定程度上体现了影响某地的天气系统。相对湿度,用于体现当地上空的天空状况以衡量辐射因子。气温方面,选取了 1000~850 hPa 气温,用以表征大气的基本冷暖状态。

相似的天气特征,由于格点所处位置的不同,格点实况值往往有较大差异。因此构建特征时,分别把 6 个分区的格点进行编码。每个分区有约 2840 个格点,每个分区用从 0 开始的正整数,逐一将格点编码,并在训练时把格点编码作为类别特征。经机器学习特征重要性排序,格点编码排名第一,说明格点的位置是气温客观预报应该考虑的一个重要特征。此外,气温存在明显的日变化特征,因此在逐小时的样本中将小时作为特征之一,并且把小时作为类别变量。预报员做主观预报时往往会考虑到模式近期预报的偏差,所以在 24 h 内的预报建模当中,使用过去 5 d 对应时刻的气温预报偏差作为特征之一进行建模。本文建模及预报流程如图 2。

### 2.3 训练方法

LightGBM 将连续特征离散化成直方图特征,减少了数据的存储空间和计算复杂度;训练过程中通过高效的垂直并行化计算,提高了训练速度;通过设置类别变量参数(categorical\_feature),选择类别变量,基于类别特征值的最优分割,对经纬度编码和小时特征设置为类别变量,较 one-hot 编码更方便、

高效,可以有效反映时空特征。

在硬件环境及时间允许范围内进行个别关键参数的网格搜索,寻找最优参数,6 个分区建模的关键参数保持一致,LightGBM 建模关键参数的最优值详见表 1。值得说明的是,迭代次数(num\_ iterations)设置为 10 000,同时设置早停(earlystopping\_ rounds)参数为 30,即如果一个验证集的评估指标在最近 30 次循环中没有继续改善(误差降低或者准确率提高),训练将提前停止训练,而不会进行到第 10 000 次。6 个分区虽然面积相差不大,但是由于迭代循环中,平原地区的误差改进小进而迭代次数相对少,所以耗时短,山区则相反。机器学习训练调优之后,保存一份含有最优参数的模型,用来进行预报。基于机器学习的气温预报 MLT 从获取 EC 模式数据到未来 24 h 的产品生成,可以在 3 min 内完成。

表 1 LightGBM 建模关键参数的最优值

Table 1 Optimal values of key parameters of LightGBM modeling

参数名称	参数简介	参数值
boosting	弱学习器的类型	gbdt
objective	学习任务及相应的学习目标	huber
metric	用于指定评估指标(l2 表示均方根误差)	l2_root
device_type	建模使用的设备类型,可选 CPU 或 GPU	CPU
num_ iterations	迭代的次数	10000
earlystopping_ rounds	迭代提前终止	30
learning_rate	学习率	0.05
num_leaves	叶子个数	800
max_depth	最大深度	10
min_data_in_leaf	叶节点样本的最少数量用于防止过拟合	30
feature_fraction	对特征随机采样的比例	0.675
lambda_l2	正则化权重项	20.04
categorical_ feature	指定哪些是类别特征(小时,格点编码所在列)	31,32

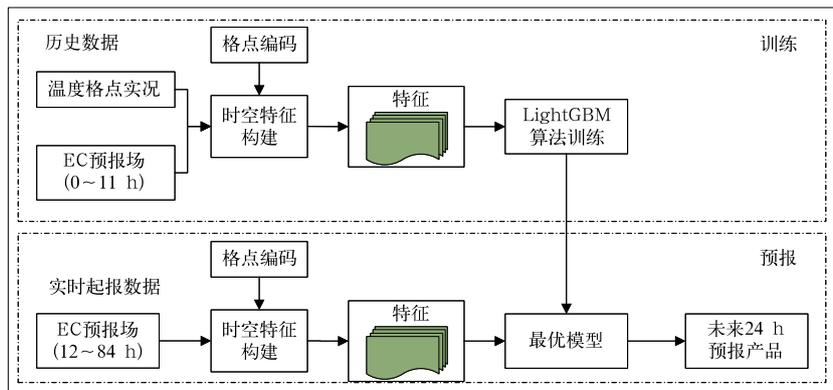


图 2 机器学习气温预测流程图

Fig. 2 Temperature prediction flow chart of machine learning

### 2.4 检验方法

智能网格检验业务相关要求使用平均绝对误差(MAE)进行气温预报性能对比。机器学习一般采用均方根误差(RMSE)作为损失函数进行回归建模。分析气温预报偏高、偏低问题,通常使用平均误差(ME);在对比多种产品的站点预报性能时,本文使用气温的预报准确率。即本文共使用了4种检验方法,分别是ME、MAE、RMSE、气温预报准确率。以上检验方法均是参照全国智能网格预报业务产品检验评估办法(2021版)。

## 3 结果分析

### 3.1 高温、低温预报产品误差的空间分布对比

利用2021年4—9月时间段的格点实况数据进行

行检验,对比24、48、72 h时效MLT、SCMOC、EC预报产品的高温、低温预报在6个分区的MAE和RMSE(图3)。由图3a、3b可见,MLT低温MAE在6个分区24~72 h时效预报误差均在1℃左右,与EC相比,西部山区低温订正幅度较大,平原地区订正幅度略小。MLT低温在6个分区的RMSE均在1℃左右,与MAE相差不大,由于RMSE对误差的反映更敏感,也证明了MLT低温较接近于实况。由图3c、3d可见,MLT高温与EC预报产品相比,西部山区高温订正幅度较大,平原地区订正幅度略小。SCMOC高温24 h预报较EC预报产品有正订正,随着时效的延长,订正幅度降低。MLT高温在6个分区的RMSE,明显低于EC和SCMOC预报产品的相应数值,且与MAE相差不大。综合分析表明,MLT高温、低温的预报效果优于SCMOC与EC预报产品,预报误差较小且较稳定。

图4、图5分别展示了湖北省区域内MLT、

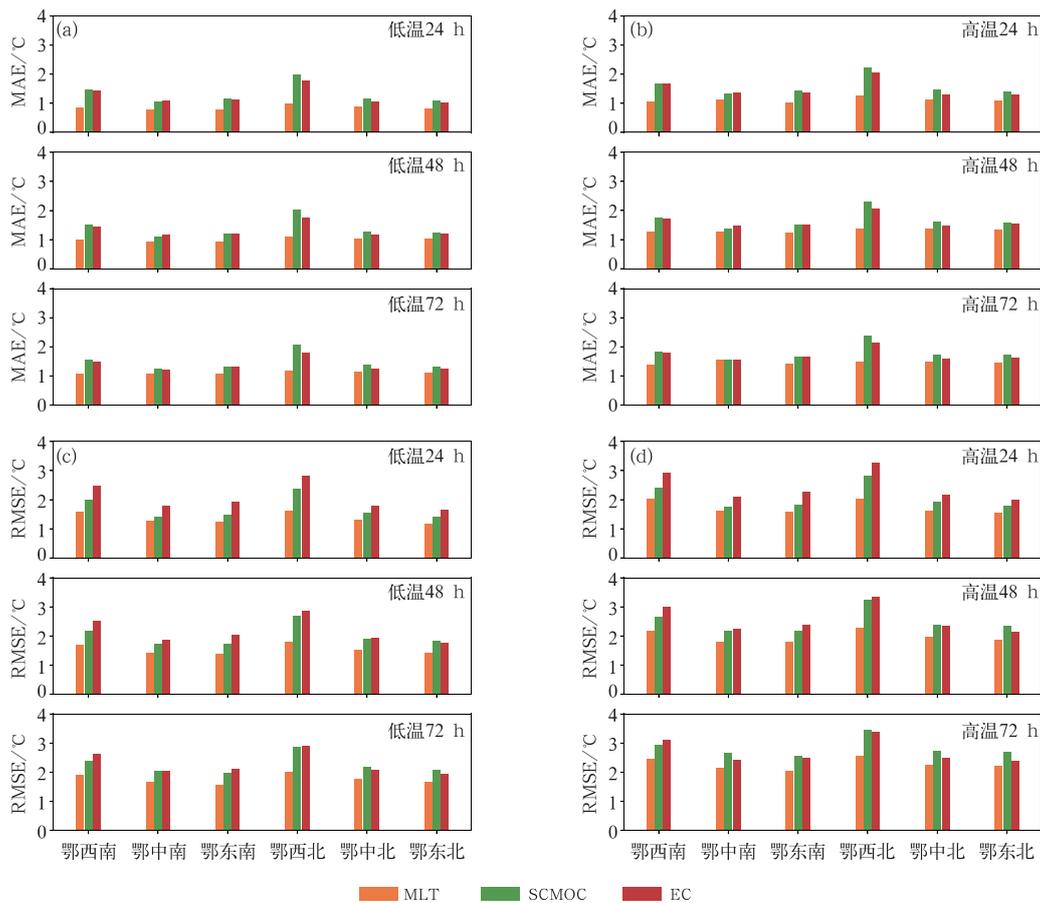


图3 2021年4—9月08时起报的24 h、48 h、72 h时效MLT、SCMOC、EC(a,c)日最低气温和(b,d)日最高气温预报产品在湖北省6个分区的(a,b)平均绝对误差和(c,d)均方根误差  
Fig.3 (a, b) MAE and (c, d) RMSE of (a, c) daily minimum temperature and (b, d) daily maximum temperature by MLT, SCMOC, EC with 24 h, 48 h, 72 h lead time initiated at 08:00 BT from April to September 2021 in six regions of Hubei Province

SCMOC 和 EC 高温、低温预报产品的 MAE 空间分布对比。EC 预报产品高温较大的 MAE 除位于鄂西山区外,鄂东北的大别山区和鄂东南的幕府山区也有两个小中心。从 SCMOC 和 MLT 高温预报产品的订正效果看,SCMOC 在平原地区 24 h 订正较好,MAE 基本在  $1.5^{\circ}\text{C}$  以下,但是鄂西山区和大别山、幕府山一带订正效果不明显。SCMOC 在 48 h 和 72 h 对山区和平原的订正效果均不明显。MLT 在山区有明显的订正效果,24~48 h 西部山区的 MAE 在  $2^{\circ}\text{C}$  以下,72 h 的 MAE 有所增大,但是比 EC 预报产品低  $0.4\sim 0.5^{\circ}\text{C}$ ,比 SCMOC 低  $0.1\sim 0.3^{\circ}\text{C}$ 。综上所述,MLT 对 EC 高温、低温预报产品有较好的订正效果,与 SCMOC 相比,MLT 高低温的偏差也有一定的降低,特别是西部山区降低的幅度更大。

对比 SCMOC 和 MLT 的方法,可以发现 SCMOC 主要基于 CMA-GFS 模式为基础,采用格点化 MOS 方法(韦青等,2020;赵声蓉等,2012),其主要思路和 MLT 的方法是一致的,都是通过选取最优的要素,利用回归的方法建立气温预报方法,但是在回归方法、要素和建模方案上存在一定的差别。SCMOC 产品输出的是全国的气温,相较 MLT 不够精细,虽

然鄂西北和鄂西南都是山区,但是气温变化规律不同,鄂西北升温和降温幅度都很大,而鄂西南一般云系多,夜间气温降幅小,因此精细的分区有利于找到更准确的地域规律。另外 MLT 是直接利用格点气温实况建模,而 SCMOC 是先利用 MOS 建模得到城镇站点预报结果,再通过插值得到格点预报(金荣花等,2019),即 MLT 减少了中间环节,更接近于格点实况。

### 3.2 格点预报产品误差的日变化特征

对比 2021 年 4—9 月 08 时起报的 SCMOC、EC 预报产品 24 h 逐小时整点气温预报的 MAE 日变化特征(图 6),可以发现 24 h 时效内三种预报产品的逐小时平均绝对误差在白天较大,夜间较小;误差最大出现在 7 h、8 h 时效,误差最小出现在 23 h 时效左右。图 6a、6d 分别为鄂西南与鄂西北的气温误差日变化特征,可以看出湖北西部高海拔山地区域逐小时平均绝对误差较其他地区大,15 时和 16 时的气温误差最大;平原地区占比较多的地区(图 6b、6c、6e、6f),逐小时 MAE 较小;三种预报产品在 6 个分区表现出了一致的误差日变化特征,即同时段山区误差较大,同地区每个时段 MLT 的预报效果都

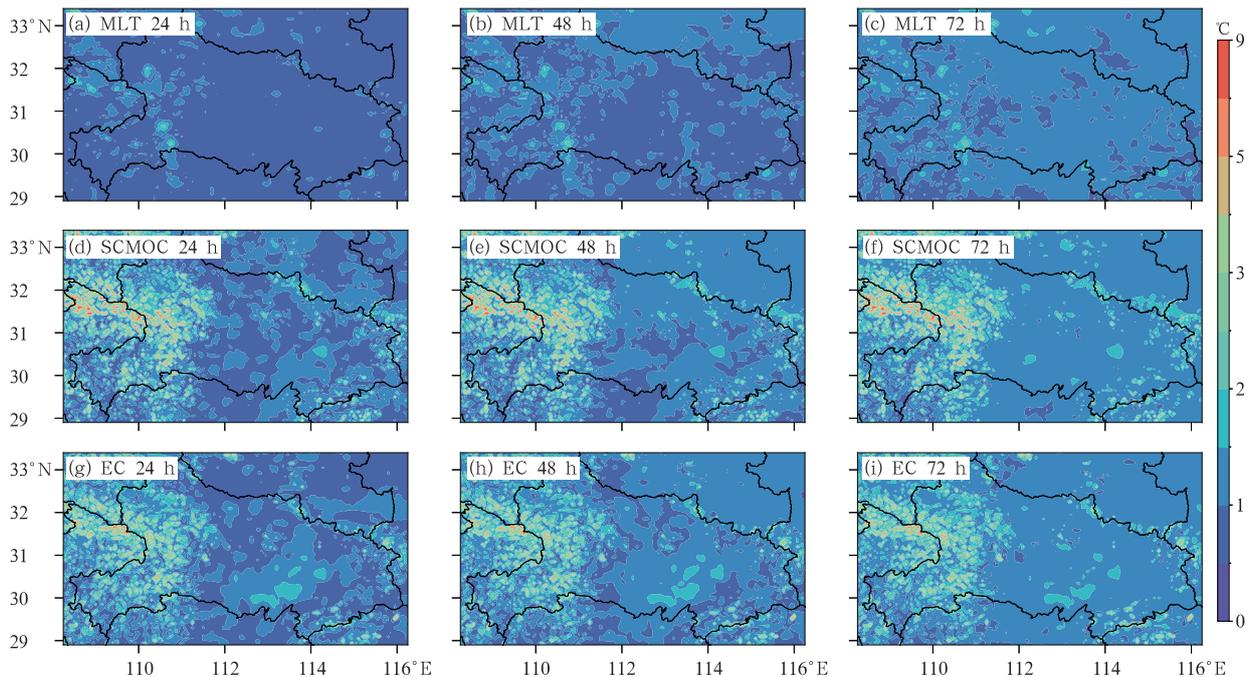


图 4 2021 年 4—9 月 08 时起报的 24 h、48 h 和 72 h 时效 MLT、SCMOC、EC 日最低气温预报产品的平均绝对误差

Fig. 4 MAE of daily minimum temperature forecast products by MLT, SCMOC and EC with 24 h, 48 h and 72 h lead time initiated at 08:00 BT from April to September 2021

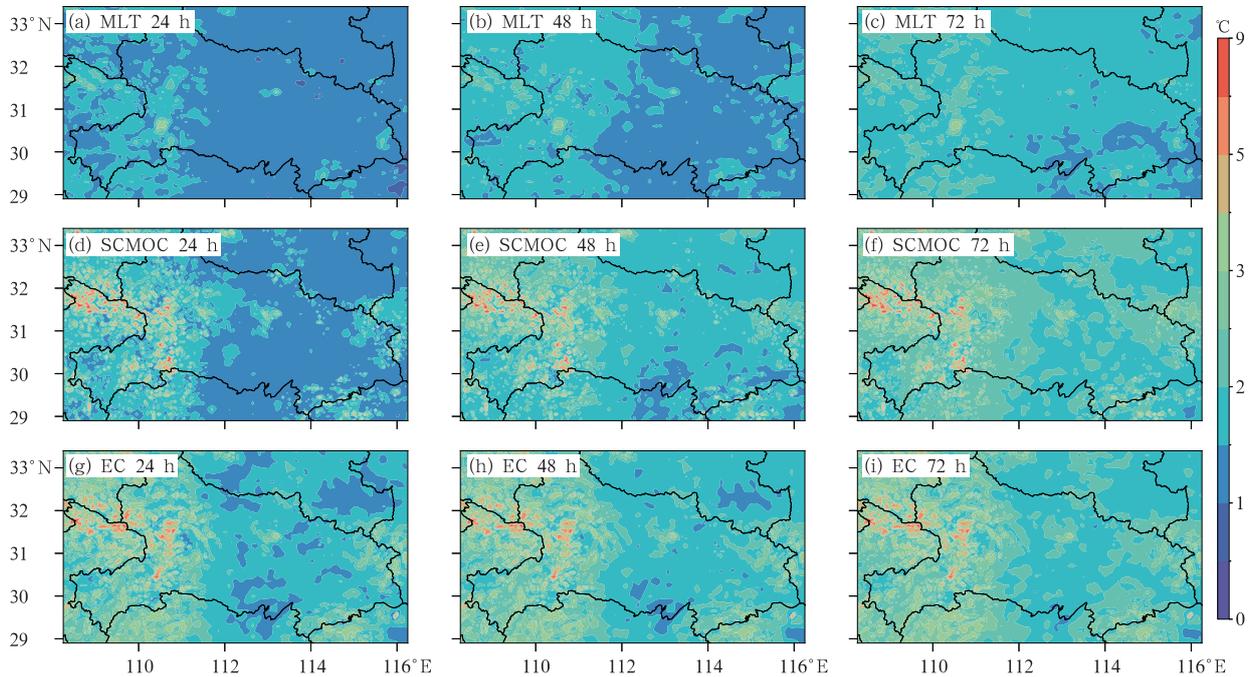


图 5 2021 年 4—9 月 08 时起报的 24 h、48 h 和 72 h 时效 MLT、SCMOC、EC 日最高气温预报产品的平均绝对误差  
 Fig. 5 MAE of daily maximum temperature forecast products by MLT, SCMOC and EC with 24 h, 48 h and 72 h lead time initiated at 08:00 BT from April to September 2021

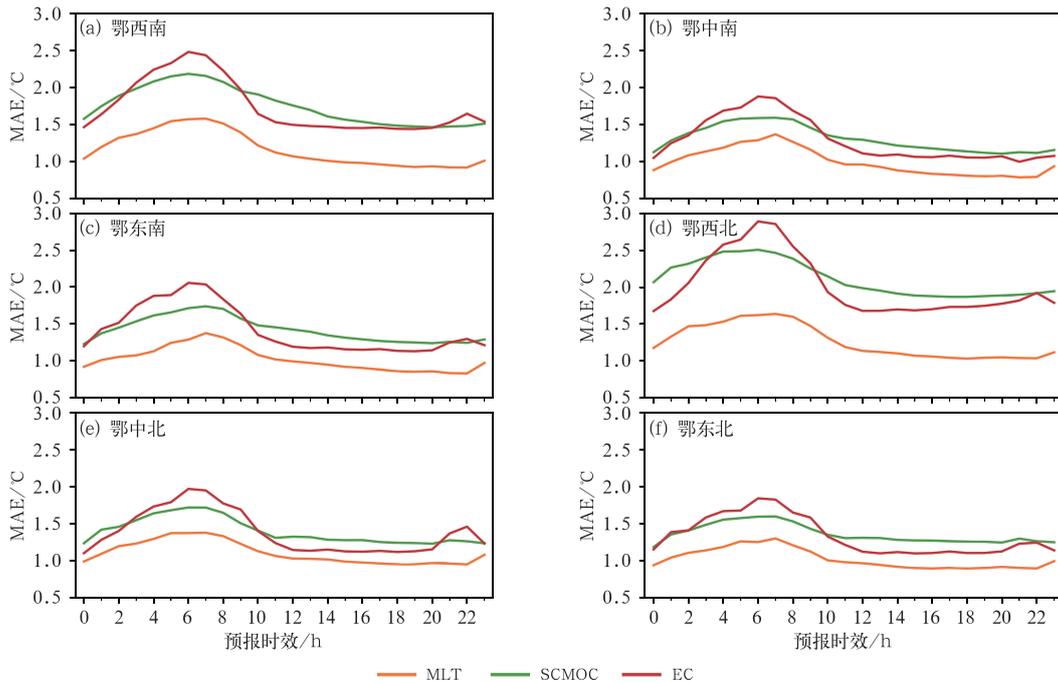


图 6 2021 年 4—9 月 08 时起报的 24 h 时效 MLT、SCMOC、EC 逐小时整点气温预报产品在湖北省 6 个分区的平均绝对误差

Fig. 6 MAE of hourly temperature forecast products by MLT, SCMOC, EC with 24 h lead time initiated at 08:00 BT from April to September 2021 in six regions of Hubei Province

是最好的。20 时起报的产品(图 7)也有类似的这种日变化特征,高温时段误差较大。

结合图 6、图 7 对不同分区整点气温的 MAE 日变化分析可以看出,SCMOC 在平原地区低温预报

误差较小,但山区误差仍较大,而 MLT 低温在平原和山区都较小。在 24 h 时效内,EC 对白天的气温预报误差明显偏大(1~3℃),尤其是中午至下午时段,鄂西山区更为突出(1.5~3.0℃);SCMOC 对白天的气温有一定的正订正效果,但对夜间气温却是负效果;MLT 对白天、夜间的气温都有明显的订正效果,尤其是对山区白天的气温订正幅度更大。

### 3.3 站点气温预报误差的空间分布及日变化特征

以湖北经纬度范围内的 171 个站实况作为真值,检验 MLT、SCMOC 在 2021 年 4 月 1 日至 9 月 30 日的逐小时气温。由图 8 可见,SCMOC 小时气温的 MAE 和 ME 明显高于 MLT(图 8a、8b);SCMOC 在湖北地区的气温预报东部偏高,西部偏低;SCMOC、MLT 的准确率分别是 77%、84%,MLT 较 SCMOC 提高了 7%(图 8e)。站点的空间误差检验表明:SCMOC、MLT 在东部平原地区的预报较西部山区更好,同时 MLT 在西部山区的预报结果比 SCMOC 更有优势,该结论与格点检验的结论一致。

对比分析逐小时整点气温预报的日变化特征(图 9)。EC、SCMOC 和 MLT 的准确率都存在单低谷的日变化特征,即在 14—17 时出现准确率低谷

值。上述时段大致为汛期最高温的出现时间,气温起伏较大,导致预报准确率较低。MLT 在站点逐小时整点预报中每一个观测时间预报准确率都高于 SCMOC,这与格点的平均绝对误差日变化结论保持一致。形成上述检验结果的原因可能是:汛期白天气温变化幅度较大,高温出现时段三种产品的预报误差较大,所以在 14—17 时出现准确率低谷值;而 MLT 建模时是逐 1 h 作为一个样本,以小时为单位将时间作为类别变量建模,可以有效学习到气温日变化。

### 3.4 转折天气个例检验

该算法目前已经在业务中实时运行,本文虽然没有将常规天气和转折天气分开建模,但是由于模型在训练的过程中充分考虑了各层的湿度,可以理解为考虑了云的情况,所以该产品在转折天气中仍有较好的表现。2021 年 4—9 月的检验结果表明,MLT 的 24 h 高温预报准确率优于 EC 和 SCMOC;同时,其在晴雨转换天气中也有较好表现。

以 2021 年 4 月 7 日为例,湖北省自西向东出现了降雨,SCMOC 的高温预报比实况偏高,准确率仅为 56.7%,MLT 的准确率达到 96.1%。图 10a、10b

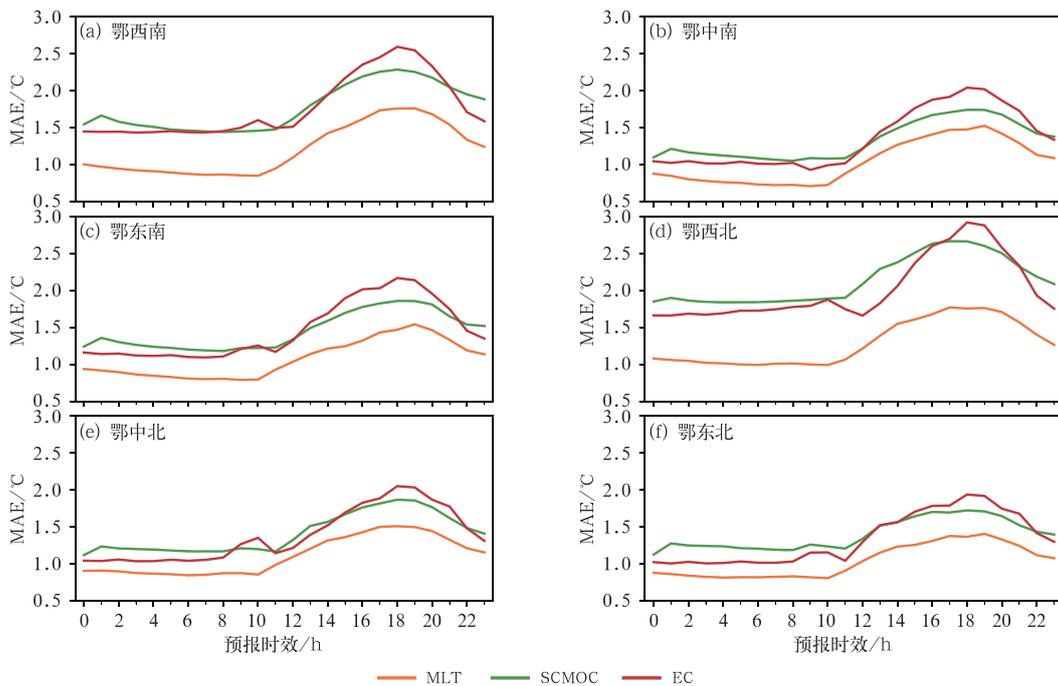
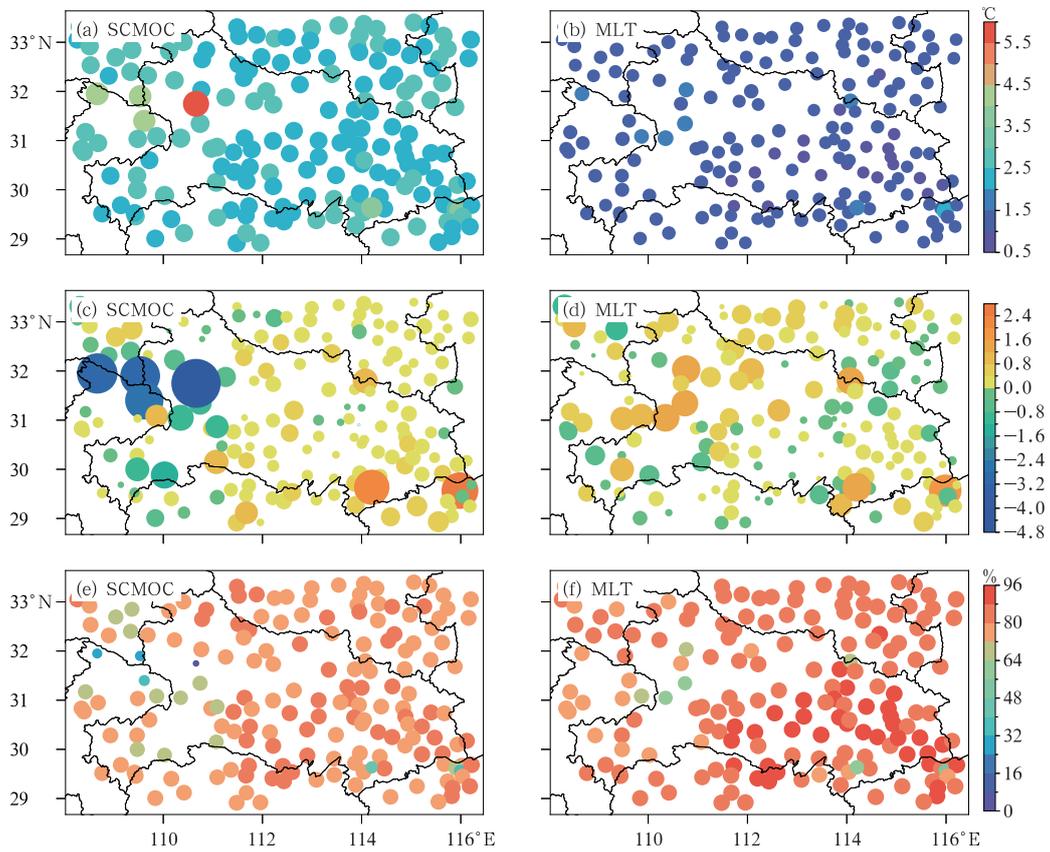


图 7 2021 年 4—9 月 20 时起报的 24 h 时效 MLT、SCMOC、EC 逐小时整点气温预报产品在湖北省 6 个分区的平均绝对误差

Fig. 7 MAE of hourly temperature forecast products by MLT, SCMOC, EC with 24 h lead time initiated at 20:00 BT from April to September 2021 in six regions of Hubei Province



注:各圆点以相应站点数值为半径;设置时以批量站点数据的平均值为基准,当某站点数值超出或小于基准时,则以其绝对值为半径,以凸显偏差情况。

图 8 2021 年 4—9 月 SCMOC、MLT 湖北省各站点整点气温预报  
(a, b)平均绝对误差, (c, d)平均误差, (e, f)预报准确率

Fig. 8 (a, b) MAE, (c, d) ME, (e, f) prediction accuracy of hourly temperature forecast by SCMOC and MLT at Hubei Province from April to September 2021

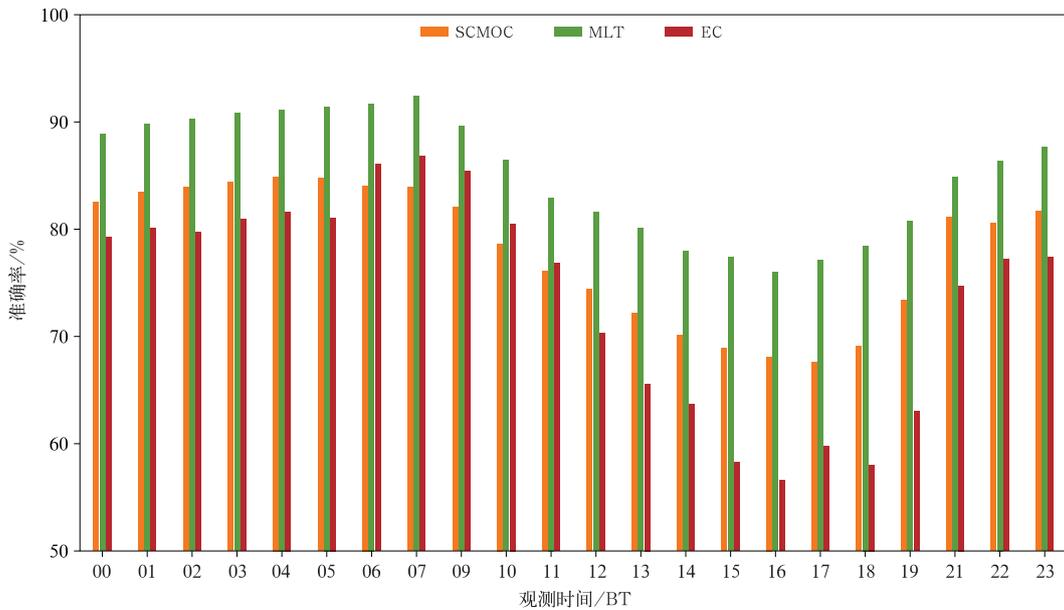


图 9 2021 年 4—9 月湖北省各站点整点气温预报准确率

Fig. 9 Prediction accuracy of station hourly temperature in Hubei Province from April to September 2021

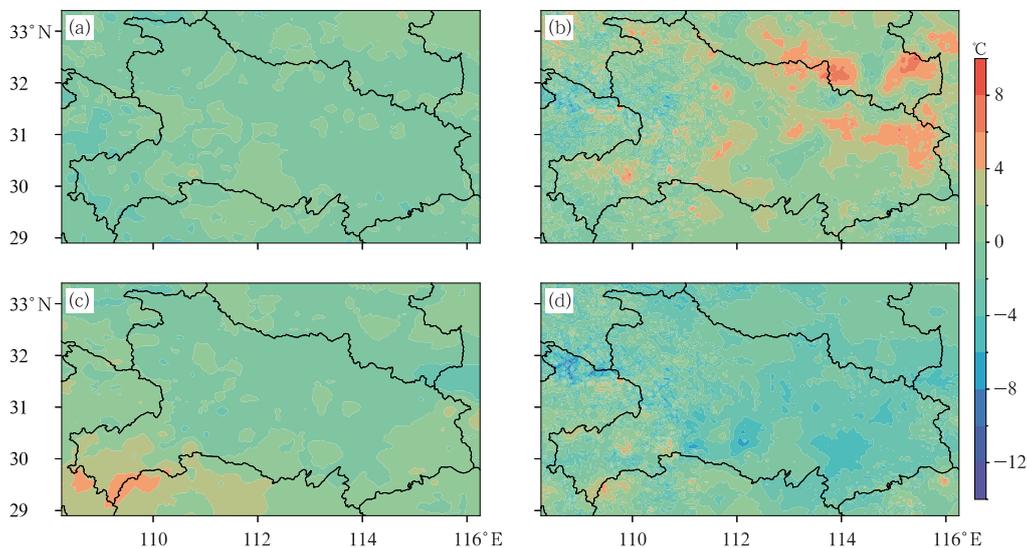


图 10 2021 年 4 月 (a,b)6 日 20 时,(c,d)11 日 20 时起报的 24 h 时效 (a,c)MLT 和 (b,d)SCMOC 格点日最高气温平均绝对误差

Fig. 10 MAE of daily maximum temperature forecast with 24 h lead time by (a, c) MLT and (b, d) SCMOC initiated at (a, b) 20:00 BT 6 and (c, d) 20:00 BT 11 April 2021

分别为 MLT 和 SCMOC 的高温预报 MAE, SCMOC 为  $2.09^{\circ}\text{C}$ , MLT 则为  $0.69^{\circ}\text{C}$ , 即 MLT 比 SCMOC 的 MAE 降低了  $1.40^{\circ}\text{C}$ , 预报结果更接近实况。2021 年 4 月 12 日过程, 是一次雨转晴的过程, SCMOC 的高温预报比实况偏低, 准确率为 52.3%, MAE 为  $2.19^{\circ}\text{C}$  (图 10d), MLT 的高温预报也偏低, 但是偏低的幅度要小于 SCMOC, MAE 为  $0.95^{\circ}\text{C}$  (图 10c), 即 MLT 比 SCMOC 的 MAE 降低了  $1.24^{\circ}\text{C}$ , 准确率也达到 89.2%。

## 4 结论与讨论

本文建立了一种基于机器学习的气温预报方法 (MLT), 应用于湖北地区, 并与 SCMOC 和 EC 模式的预报产品进行了对比检验。主要有以下结论:

(1) 通过格点气温实况产品和主客观筛选的 30 多个特征量, 利用 LightGBM 方法建立的格点气温预报方法, 可以提供时间分辨率为 1 h、空间分辨率为  $0.05^{\circ}\times 0.05^{\circ}$  的气温预报产品。在 0~72 h 时效内 MLT 的 MAE、RMSE 较 SCMOC 与 EC 的低。

(2) MLT 在山区的误差较平原大, 但山区的订正幅度大于平原, 日最高温的订正幅度大于日最低温。RMSE 比 MAE 略大, 但误差分布是一致的。在模型中设置格点编码和小时特征为类别变量, 较

好地分别反映了格点气温的地理特征和日变化特征。

(3) MLT 以站点为真值检验, 仍然有很好的预报效果。这说明应用机器学习基于格点实况建模的格点气温预报产品对站点预报同样有订正作用。

(4) 2021 年 4—9 月 MLT、SCMOC、EC 的 MAE 日变化都呈现了白天偏高、夜间偏低, 午后凸起的单峰特征, MLT 不仅在湖北全省累计平均误差较低, 在单时次、单区域的平均误差较 SCMOC、EC 仍较低, 并且在转折性天气中仍具有优势。检验结果表明, MLT 在湖北的气温预报准确率优于 EC 和 SCMOC。

由此可见, 机器学习在格点气温的模式订正方面可以作为一个行之有效的手段。本文运用的机器学习方法对 EC 气温预报模式订正表现出了较好的效果, 对其他模式或其他要素预报订正仍具有很大的潜力。未来可以在天气学的物理意义指导下进行更精细的数据处理、特征工程, 应可挖掘出更多的可订正的时空误差信息, 进而达到更好的预报效果。

## 参考文献

- 陈昱文, 黄小猛, 李熠, 等, 2020. 基于 ECMWF 产品的站点气温预报集成学习误差订正[J]. 应用气象学报, 31(4): 494-503. Chen Y W, Huang X M, Li Y, et al, 2020. Ensemble learning for bias correction of station temperature forecast based on ECMWF

- products[J]. *J Appl Meteor Sci*, 31(4):494-503(in Chinese).
- 郝翠,张迎新,王在文,等,2019. 最优集合预报订正方法在客观温度预报中的应用[J]. *气象*, 45(8):1085-1092. Hao C, Zhang Y X, Wang Z W, et al, 2019. Application of analog ensemble rectifying method in objective temperature prediction[J]. *Meteor Mon*, 45(8):1085-1092(in Chinese).
- 金荣花,代刊,赵瑞霞,等,2019. 我国无缝隙精细化网格天气预报技术进展与挑战[J]. *气象*, 45(4):445-457. Jin R H, Dai K, Zhao R X, et al, 2019. Progress and challenge of seamless fine gridded weather forecasting technology in China[J]. *Meteor Mon*, 45(4):445-457(in Chinese).
- 李扬,刘玉宝,许小峰,2021. 基于深度学习改进数值天气预报模式和预报的研究及挑战[J]. *气象科技进展*, 11(3):103-112. Li Y, Liu Y B, Xu X F, 2021. Advances and challenges for improving numerical weather prediction models and forecasting using deep learning[J]. *Adv Meteor Sci Technol*, 11(3):103-112(in Chinese).
- 雷彦森,蔡晓军,王文,等,2018. 遗传算法优化的 BP 神经网络在地面温度多模式集成预报的应用研究[J]. *气象科学*, 38(6):806-814. Lei Y S, Cai X J, Wang W, et al, 2018. Application research of BP neural network optimized by genetic algorithm in multi-model ensemble forecast about ground temperature[J]. *J Meteor Sci*, 38(6):806-814(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]. *Climatic Environ Res*, 24(1):116-124(in Chinese).
- 潘留杰,薛春芳,王建鹏,等,2017. 一个简单的格点温度预报订正方法[J]. *气象*, 43(12):1584-1593. Pan L J, Xue C F, Wang J P, et al, 2017. A simple grid temperature forecast correction method[J]. *Meteor Mon*, 43(12):1584-1593(in Chinese).
- 任萍,陈明轩,曹伟华,等,2020. 基于机器学习的复杂地形下短期数值天气预报误差分析与订正[J]. *气象学报*, 78(6):1002-1020. Ren P, Chen M X, Cao W H, et al, 2020. Error analysis and correction of short-term numerical weather prediction under complex terrain based on machine learning[J]. *Acta Meteor Sin*, 78(6):1002-1020(in Chinese).
- 盛春岩,范苏丹,荣艳敏,等,2020. 几种气温客观预报方法对比及最优集成预报研究[J]. *气象*, 46(10):1351-1361. Sheng C Y, Fan S D, Rong Y M, et al, 2020. Comparison of several objective methods and optimal consensus forecast study of temperature[J]. *Meteor Mon*, 46(10):1351-1361(in Chinese).
- 师春香,潘旻,谷军霞,等,2019. 多源气象数据融合格点实况产品研制进展[J]. *气象学报*, 77(4):774-783. Shi C X, Pan Y, Gu J X, et al, 2019. A review of multi-source meteorological data fusion products[J]. *Acta Meteor Sin*, 77(4):774-783(in Chinese).
- 孙全德,焦瑞莉,夏江江,等,2019. 基于机器学习的数值天气预报风速订正研究[J]. *气象*, 45(3):426-436. Sun Q D, Jiao R L, Xia J J, et al, 2019. Adjusting wind speed prediction of numerical weather forecast model based on machine learning methods[J]. *Meteor Mon*, 45(3):426-436(in Chinese).
- 王建鹏,薛春芳,潘留杰,等,2018. 陕西省精细化网格预报业务系统技术方法[J]. *气象科技*, 46(5):910-918. Wang J P, Xue C F, Pan L J, et al, 2018. Operation system of fine grid forecast in Shaanxi Province: technical methods[J]. *Meteor Sci Technol*, 46(5):910-918(in Chinese).
- 韦青,代刊,林建,等,2020. 2016—2018 年全国智能网格降水及温度预报检验评估[J]. *气象*, 46(10):1272-1285. Wei Q, Dai K, Lin J, et al, 2020. Evaluation on the 2016—2018 fine gridded precipitation and temperature forecasting[J]. *Meteor Mon*, 46(10):1272-1285(in Chinese).
- 吴启树,韩美,刘铭,等,2017. 基于评分最优化的模式降水预报订正算法对比[J]. *应用气象学报*, 28(3):306-317. Wu Q S, Han M, Liu M, et al, 2017. A comparison of optimal-score-based correction algorithms of model precipitation prediction[J]. *J Appl Meteor Sci*, 28(3):306-317(in Chinese).
- 许小峰,2018. 从物理模型到智能分析——降低天气预报不确定性的新探索[J]. *气象*, 44(3):341-350. Xu X F, 2018. From physical model to intelligent analysis: a new exploration to reduce the uncertainty of weather forecast[J]. *Meteor Mon*, 44(3):341-350(in Chinese).
- 薛湛彬,陈娴,张瑛,等,2019. ECMWF 高分辨率模式 2 m 温度预报误差订正方法研究[J]. *气象*, 45(6):831-842. Xue C B, Chen X, Zhang Y, et al, 2019. Bias correction method for the 2 m temperature forecast of ECMWF high resolution model[J]. *Meteor Mon*, 45(6):831-842(in Chinese).
- 杨璐,南刚强,陈明轩,等,2021. 基于三种机器学习方法的降水相态高分辨率格点预报模型的构建及对比分析[J]. *气象学报*, 79(6):1022-1034. Yang L, Nan G Q, Chen M X, et al, 2021. The construction and comparison of high resolution precipitation type prediction models based on three machine learning methods[J]. *Acta Meteor Sin*, 79(6):1022-1034(in Chinese).
- 杨绚,代刊,朱跃建,2022. 深度学习技术在智能网格天气预报中的应用进展与挑战[J]. *气象学报*, 80(5):649-667. Yang X, Dai K, Zhu Y J, 2022. Progress and challenges of deep learning techniques in intelligent grid weather forecasting[J]. *Acta Meteor Sin*, 80(5):649-667(in Chinese).
- 俞剑蔚,李聪,蔡凝昊,等,2019. 国家级格点实况分析产品在江苏地区的适用性评估分析[J]. *气象*, 45(9):1288-1298. Yu J W, Li C, Cai N H, et al, 2019. Applicability evaluation of the national gridded real-time observation datasets in Jiangsu Province[J]. *Meteor Mon*, 45(9):1288-1298(in Chinese).
- 赵声蓉,赵翠光,赵瑞霞,等,2012. 我国精细化客观气象要素预报进展[J]. *气象科技进展*, 2(5):12-21. Zhao S R, Zhao C G, Zhao R X, et al, 2012. The development of objective meteorological element forecast in China[J]. *Adv Meteor Sci Technol*, 2(5):12-21

- (in Chinese).
- 周康辉,郑永光,韩雷,等,2021a. 机器学习在强对流监测预报中的应用进展[J]. 气象,47(3):274-289. Zhou K H, Zheng Y G, Han L, et al, 2021a. Advances in application of machine learning to severe convective weather monitoring and forecasting[J]. Meteor Mon, 47(3):274-289(in Chinese).
- 周康辉,郑永光,王婷波,2021b. 利用深度学习融合 NWP 和多源观测数据的闪电落区短时预报方法[J]. 气象学报,79(1):1-14. Zhou K H, Zheng Y G, Wang T B, 2021b. Very short-range lightning forecasting with NWP and observation data: a deep learning approach[J]. Acta Meteor Sin, 79(1):1-14 (in Chinese).
- Kashinath K, Mustafa M, Albert A, et al, 2021. Physics-informed machine learning: case studies for weather and climate modelling [J]. Philos Trans Roy Soc A Math Phys Eng Sci, 379(2194): 20200093
- Ke G L, Meng Q, Finley T, et al, 2017. LightGBM: a highly efficient gradient boosting decision tree [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc.
- Haupt S E, Chapman W, Adams S V, et al, 2021. Towards implementing artificial intelligence post-processing in weather and climate: proposed actions from the Oxford 2019 workshop[J]. Philos Trans Roy Soc A Math Phys Eng Sci, 379(2194): 20200091.
- Pan B X, Hsu K, Aghakouchak A, et al, 2019. Improving precipitation estimation using convolutional neural network[J]. Water Resour Res, 55(3):2301-2321.
- Wang F, Tian D, Lowe L, et al, 2021. Deep learning for daily precipitation and temperature downscaling[J]. Water Resour Res, 57(4):e2020WR029308.

(本文责编:戴洋)