

探究基于LightGBM与地形的黄冈站点温度订正模型

李子祎

黄冈市气象局气象台, 湖北 黄冈

收稿日期: 2026年8月11日; 录用日期: 2026年9月9日; 发布日期: 2026年9月18日

摘要

针对2 m气温预报中局地效应表征不足等问题, 本文采用欧洲中心再分析资料ERA5高空数据、地面站点观测资料以及30 m分辨率地形数据, 在黄冈地区构建了基于LightGBM算法的站点精细化温度订正模型。模型对比结果表明, LightGBM模型在预报精度和训练效率方面均表现最优, 具有良好的业务化潜力。独立预报检验结果显示, 4月(7月)模型的MAE和RMSE较ERA5分别降低16.02% (11.54%)和12.63% (10.78%)、准确率提升了9.71 (5.53)个百分点。虽然模型与ERA5的误差空间分布均表现出平原低、山区高的特征, 但模型在山区的各项指标更高, 改进效果更明显, 且对温度场空间细节与局地中心的刻画能力明显优于ERA5, 可为大别山区等复杂地形区域温度精细化预报提供技术支撑。

关键词

LightGBM, 高精度地形数据, 气温订正模型

Exploring a Station-Scale Temperature Correction Model for Huanggang Based on LightGBM and Topography

Ziyi Li

Meteorological Observatory of Huanggang Meteorological Bureau, Huanggang Hubei

Received: August 11, 2026; accepted: September 9, 2026; published: September 18, 2026

Abstract

To address issues such as the inadequate representation of local effects in 2-meter air temperature forecasts, this study uses upper-air data from ERA5 reanalysis, ground station observations, and 30 m resolution topographic data to construct a regional high-resolution temperature correction model

based on the LightGBM algorithm in Huanggang. Model comparison results show that the LightGBM model performs best in both prediction accuracy and training efficiency, demonstrating strong potential for operational application. Independent forecast test results show that compared to ERA5, the model reduces MAE by 16.02% (11.54%) and RMSE by 12.63% (10.78%) in April (July), with the accuracy increased by 9.71 (5.53) percentage points. Although both the model and ERA5 exhibit a spatial error distribution characterized by lower errors in plains and higher errors in mountainous areas, the model achieves better performance metrics and more prominent improvement effects in mountainous regions. Furthermore, its ability to capture spatial details and local centers within temperature field is significantly superior to that of ERA5, providing technical support for high-resolution temperature forecasting in complex topographic regions such as the Dabie Mountains.

Keywords

LightGBM, High-Resolution Topographic Data, Temperature Correction Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

气温作为天气预报、气候分析中最基本且重要的要素之一，与人民生活密切相关。当前，数值天气预报等传统方法广泛应用于气温相关研究中，然而这类方法对气象过程复杂的非线性关系刻画能力有限[1]，在地形起伏大的区域往往存在系统性偏差[2][3]。随着人工智能技术的发展，机器学习的加入可以有效突破现有技术瓶颈[1][3]-[6]。其中，LightGBM 算法作为一种高效的梯度提升决策树算法，已在多个预报订正项目中证明了其业务化潜力[7][8]。黄冈地区地形复杂多样，数值模式直接输出的温度场往往难以准确刻画复杂地形处的局地气温差异[9]，无法满足精细化气候分析、精准气象服务的应用需求。随着遥感技术和地理信息系统的发展，已有研究证实，将地形特征引入机器学习能够有效提升复杂地形下气象要素的预报精度[10]。

鉴于此，本文使用 ERA5 再分析数据、地面观测数据以及高分辨率的 DEM、坡度和坡向数据，基于 LightGBM 算法，建立了一种适用于黄冈地区的气温订正模型。该方法旨在弥补现有温度预报方法的不足，降低复杂地形和局地气候条件对温度预报影响，提高温度预报的时空精度，以满足各行各业对精细化气温预报服务的需求。

2. 数据来源

本文使用的资料包括：(1) 黄冈市国家及区域气象自动观测站的地面常规气象观测数据、(2) 欧洲中期天气预报中心(ECMWF)第五代大气再分析数据集 ERA5 [11]高空再分析资料、(3) FathomDEM 全球 30 m 分辨率数字高程数据集[12] (此后简称为 DEM)以及派生得到的坡度坡向数据。其中，在模型训练阶段使用的是 2022~2024 年逐 3 小时的数据，且站点数据仅使用筛选出的 115 个站点；而检验阶段使用的是所有 299 个站。

3. 模型建立与评估

3.1. 特征工程

为充分挖掘多源数据中的有效信息，针对温度预报的物理机制，构建多层次特征体系：

- 1) 大气热力与动力特征：各层温度、湿度、风场及同层温度与湿度的乘积；
- 2) 垂直结构特征：各层间温度梯度；
- 3) 地形与地理特征：经纬度、高程、坡度坡向；
- 4) 时间周期特征：小时、年积日、月份的周期值。

分别计算 LightGBM 和随机森林特征重要性、互信息、皮尔逊相关系数 4 项指标，通过算数平均得到综合重要性得分，得分越高代表特征对气温的综合解释能力越强。按综合得分降序选取前 30 个核心特征作为模型输入，以降低模型复杂度，其中排名前 5 的分别是 DEM、1000 hPa 温度、小时正弦值、极大风速、1000 hPa 比湿。

3.2. 模型参数优化

采用随机搜索与 2 折 K-fold 交叉验证完成超参数寻优，优化参数包括 `n_estimators`、`learning_rate`、`max_depth`、`num_leaves`、`min_child_samples`、`min_child_weight`、`subsample`、`colsample_bytree`、`reg_alpha`、`reg_lambda`、`min_gain_to_split`、`min_sum_hessian_in_leaf`、`max_bin`，得到 LightGBM 模型的最优参数组合：`num_leaves` 为 191，`learning_rate` 为 0.05，`max_depth` 为 15，`reg_alpha` 为 0.005，`reg_lambda` 为 0.05。数据集划分流程如图 1 所示。

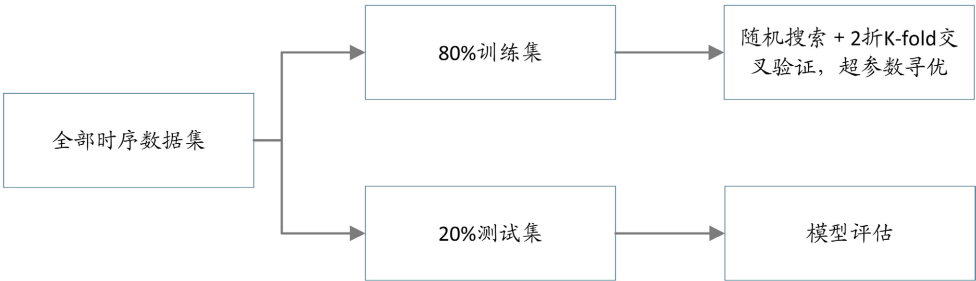


Figure 1. Flowchart of the dataset partitioning process
图 1. 数据集划分流程图

4. 实验结果与分析

4.1. 模型训练性能对比

Table 1. Comparison of forecasting performance among different algorithm models
表 1. 不同算法模型预报性能对比

算法模型	MAE/℃	RMSE/℃	R ²	CV R ²	训练时间/s
线性回归	1.582	2.052	0.958	0.952 ± 0.011	12.56
岭回归	1.581	2.051	0.958	0.952 ± 0.011	13.12
LightGBM	0.657	0.876	0.992	0.975 ± 0.005	26.32
XGBoost	0.692	0.922	0.991	0.977 ± 0.004	31.22
随机森林	1.101	1.447	0.979	0.963 ± 0.007	79.45
集成模型	0.757	1.017	0.990	0.987 ± 0.006	146.22

为验证 LightGBM 模型的可靠性，选取线性回归、岭回归、XGBoost、随机森林和集成算法模型进行对比(表 1)。LightGBM 模型综合表现最好，其 MAE (0.657℃)与 RMSE (0.876℃)最低，较线性模型 MAE

降低了约 0.92°C ，且其 R^2 高达 0.992，表明其解释能力与拟合效果最好。此外，LightGBM 模型的 CVR^2 为 0.975，交叉验证波动仅 ± 0.005 ，模型的稳定性较强，仅略逊于 XGBoost (± 0.004)；集成算法虽然 CVR^2 (0.987) 最高，泛化能力最优，但其运算时间大幅上升。LightGBM 完成训练仅耗时 26.32s，远低于集成模型 (146.22 s)。因此，LightGBM 模型在误差、拟合度与运算效率上综合表现最佳，更适配业务应用的场景。

4.2. 预报检验结果

为验证该 LightGBM 模型在全新气象条件下的订正能力，选用 2025 年 4 月和 7 月的 ERA5 高空和地面数据作为模型输入，以同期 299 个地面气象站逐 3 小时实测 2 m 温度为真值，从整体精度、空间分布及温度场分布四个方面与 ERA5 的 2 m 温度产品进行对比分析。

LightGBM 模型与 ERA5 的整体定量评估结果显示，模型的各项评估指标均优于 ERA5。从误差指标来看，4 月(7 月)模型的 MAE 和 RMSE 分别为 1.520°C (1.441°C) 和 2.020°C (1.921°C)，较 ERA5 分别降低了 0.290°C (0.188°C) 和 0.292°C (0.232°C)，说明误差得到了较为明显改善。从拟合能力来看，4 月(7 月)模型的 R^2 达 0.846 (0.780)，高于 ERA5 的 0.798 (0.711)，表明该模型能够更好地刻画研究区域的气温空间分布规律。在业务指标方面，4 月(7 月)模型 2°C 准确率为 72.78% (74.21%)，相比 ERA5 提升了 9.71 (5.53)个百分点，其业务可用性显著提高。

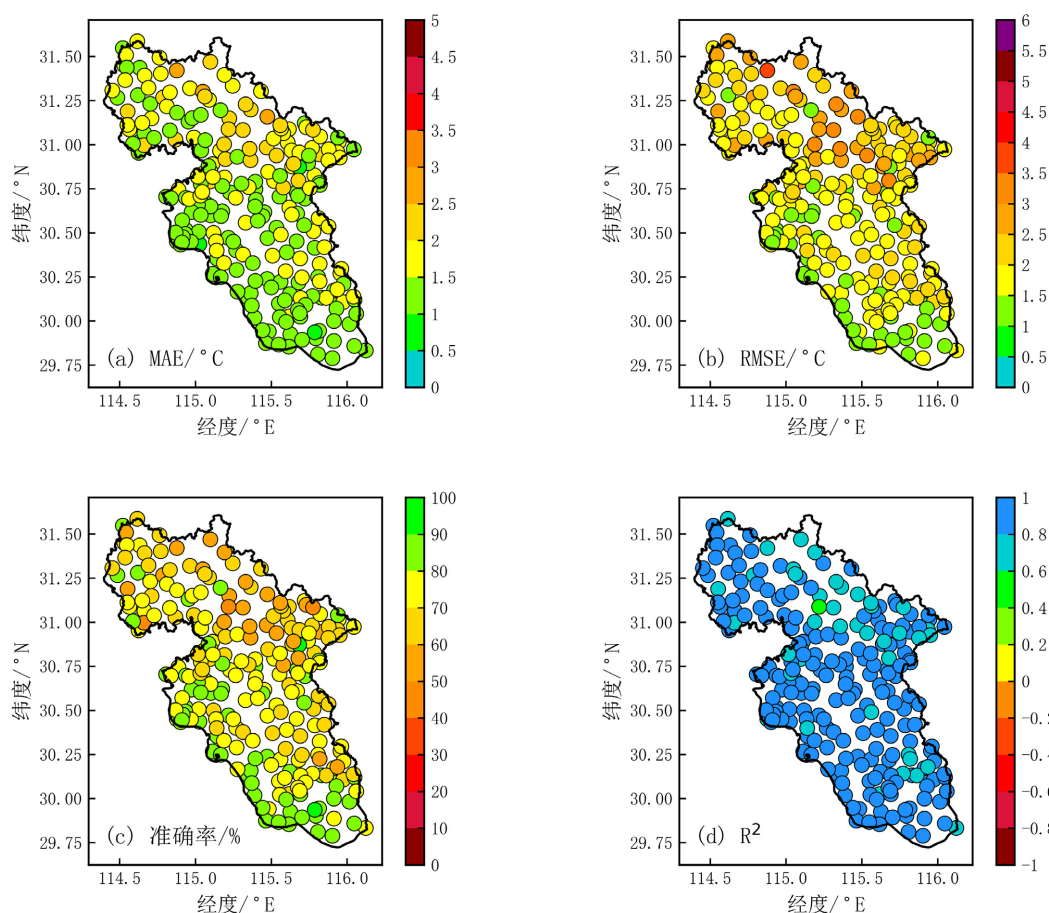


Figure 2. Spatial distribution of model evaluation metrics in April

图 2. 4 月模型评估指标的空间分布

基于各站点的评估结果,分别绘制了4月模型与ERA5评估指标的空间分布图(图2、图3)。模型与ERA5整体都呈现出误差平原较低,山区明显增大的特征。平原地区二者MAE普遍小于 1.5°C ,RMSE小于 2.0°C ,准确率高于70%, R^2 普遍大于0.8;而在大别山等地形起伏较大的地区,误差明显升高,MAE多介于 $1.5^{\circ}\text{C}\sim 2.5^{\circ}\text{C}$ 之间,准确率偏低。从图3可以看出,ERA5在山区出现多个明显的高误差站点,部分站点MAE和RMSE突破 3°C 、准确率显著降低、 R^2 明显下滑甚至个别站点接近0;与之相比,模型在山区站点误差增幅显著小于ERA5,且绝大多数站点准确率都高于60%, R^2 几乎都在0.6以上,空间分布更为稳定。7月评估指标的分布特征与4月类似。从4月和7月模型与ERA5的指标差值空间分布(图4)可以更清楚地看到,模型相对ERA5的改进主要集中在北部大别山区,表明引入地形因子可以有效校正ERA5在复杂山地处的偏差。但模型对于平原地区的改进幅度有限,尤其是7月南部平原地区,部分站点甚至出现误差略高于ERA5的现象,这可能是由于水体热效应、城市活动会在一定程度上弱化海拔对气温的调控作用。此外,夏季湍流混合充分,地形引发的局地小尺度温度差异可能被抵消,使得地形因子可发挥的订正空间被压缩。

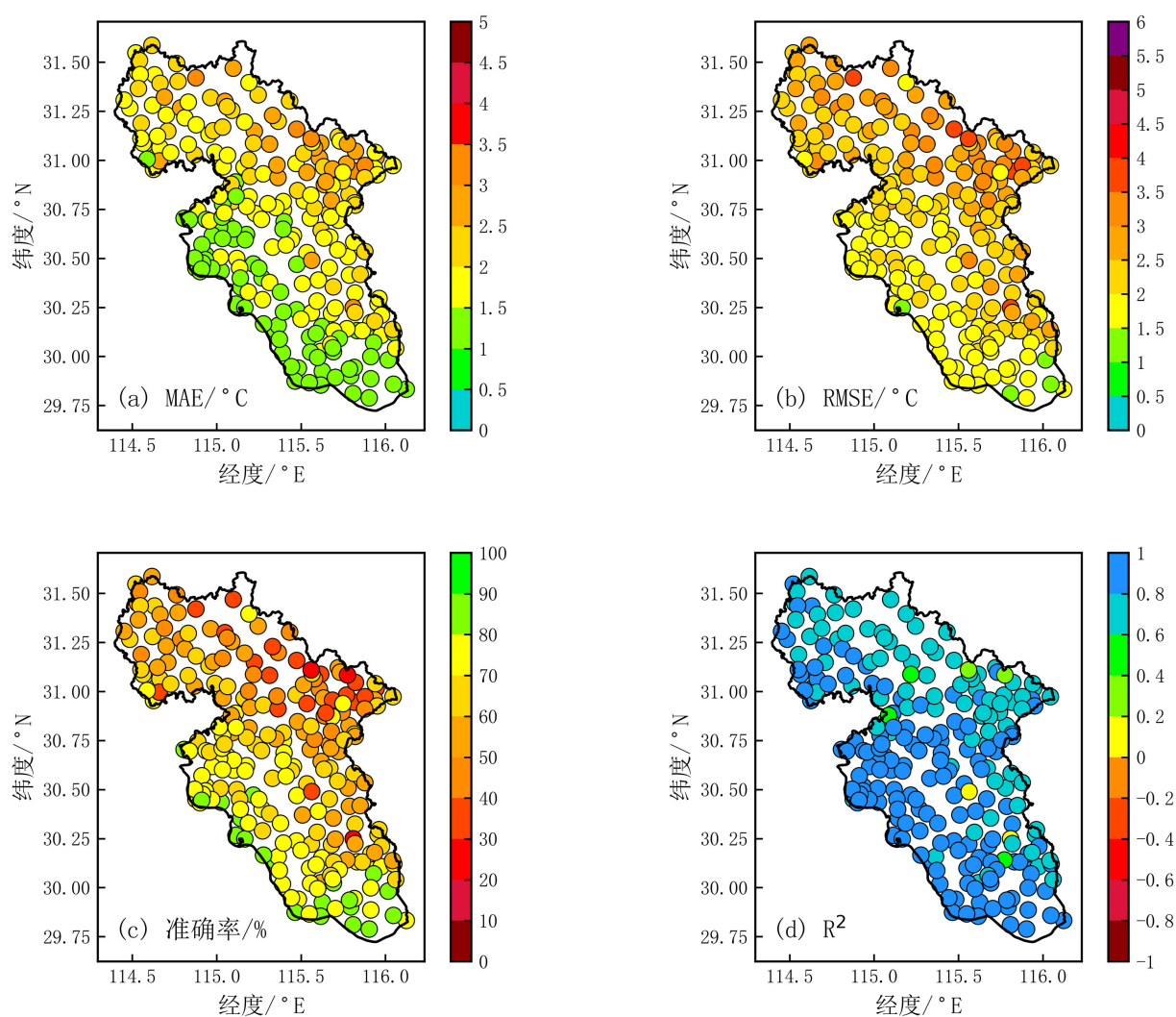


Figure 3. Spatial distribution of ERA5 evaluation metrics in April

图 3. 4月 ERA5 评估指标的空间分布

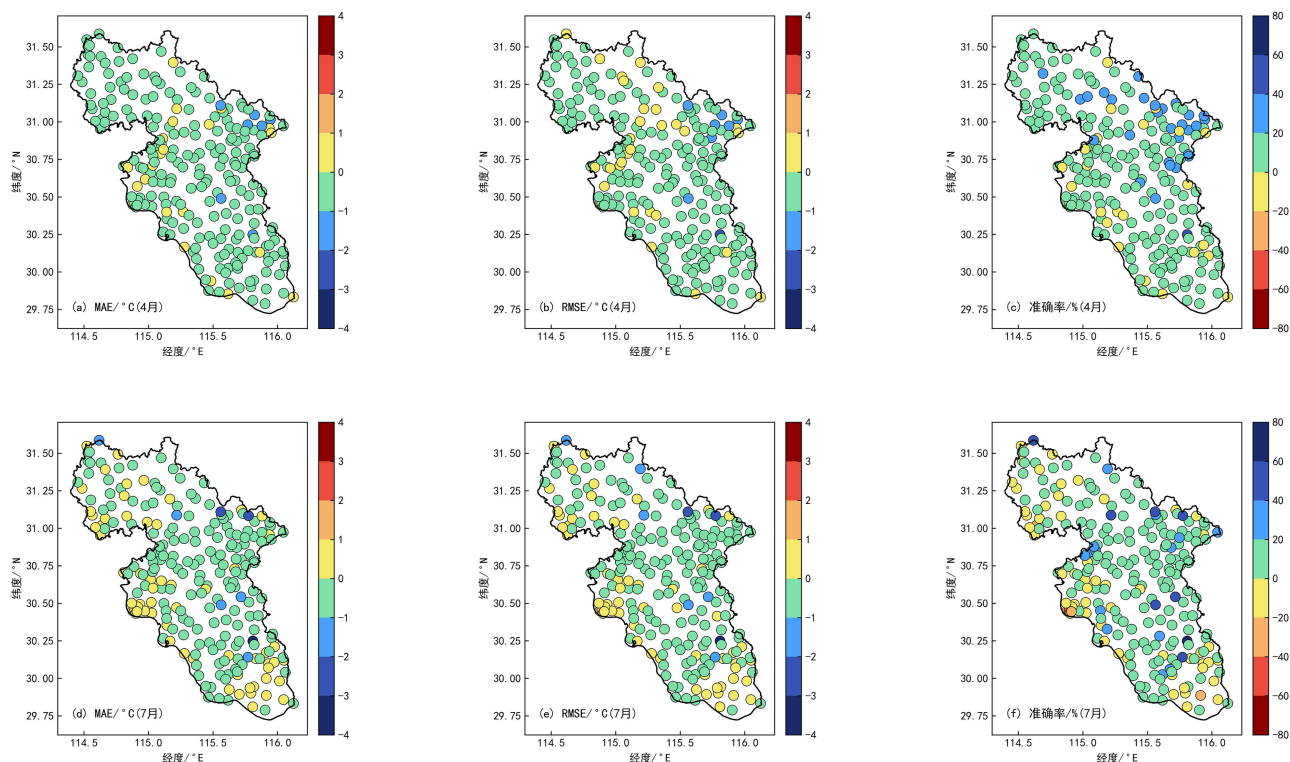


Figure 4. Spatial distribution of differences between model and ERA5 evaluation metrics in April and July

图 4. 4 月和 7 月模型与 ERA5 评估指标差值的空间分布

选取 2025 年 4 月 21 日 12 时, 对站点观测、模型预报与 ERA5 的地面气温空间分布情况(图 5)进行比较。可以看到, ERA5 虽然可以呈现出大范围的南北气温梯度变化, 但表现为平滑的带状分布, 无法反映地形等因素造成的局地温度差异。相比之下, 模型能较为准确地再现研究区域的温度空间分布细节, 高温、低温中心位置和量级都与站点观测基本一致。对比结果说明模型不仅在统计指标上表现优异, 在温度场的空间特征与局地细节刻画上也具备显著优势, 能更好满足站点尺度的精细化服务需求。

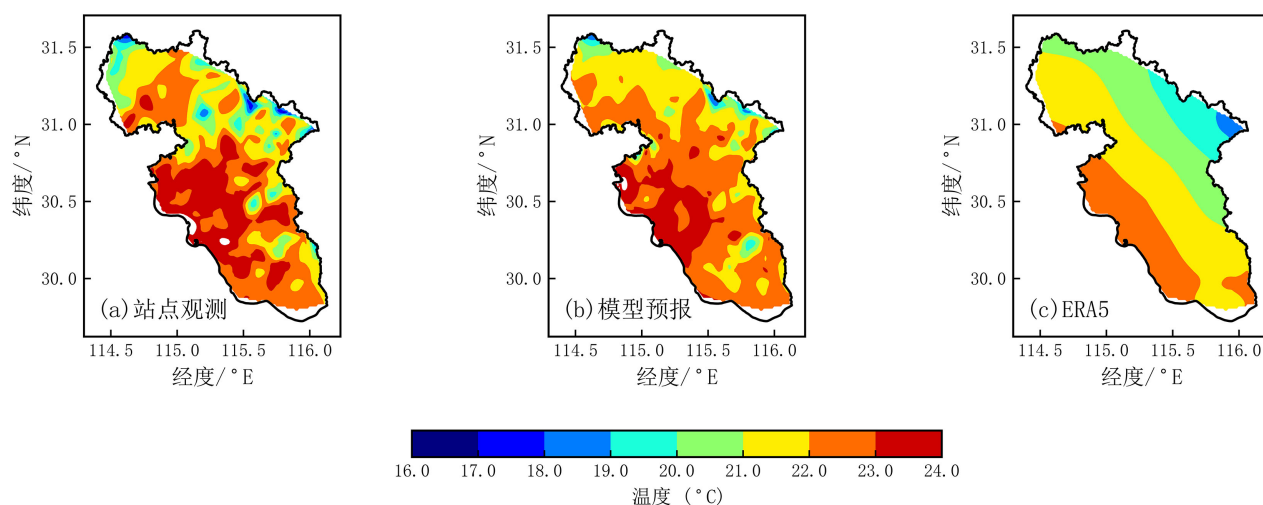


Figure 5. Temperature distribution from station observations, model, and ERA5

图 5. 站点观测、模型预报和 ERA5 的温度分布图

5. 结论与讨论

本研究基于 LightGBM 算法, 采用 ERA5 高空再分析数据、地面气象站点观测数据、高精度地形数据, 通过构建物理意义明确的特征体系, 建立了适用于黄冈地区的高精度气温订正模型。主要结论如下。

(1) 特征重要性分析表明, DEM 表征的气温垂直递减效应对 2 m 气温的空间分布起着关键作用, 1000 hPa 温度可以提供低层大尺度热力背景; 太阳辐射主导气温的日变化, 而近地面风场通过影响湍流混合进一步影响局地气温。

(2) 多种算法模型对比试验表明, LightGBM 算法的性能最优, 可高效捕捉气温与多源要素之间的复杂非线性关系。

(3) 引入地形因子的 LightGBM 模型可显著提升整体预报精度, 相较于 ERA5 的气温产品, 模型误差更低、拟合效果更优、可靠性更强。

(4) 模型与 ERA5 均表现出山区误差明显高于平原的特点, 但模型的各项评估指标均明显优于 ERA5, 尤其在山区的改进效果突出, 表明引入高分辨率的地形参数能够有效弥补模式对地形刻画不足的问题, 提升复杂下垫面区域的气温精准度。

(5) 模型在温度场空间细节与局地特征刻画上优势明显。相较于 ERA5 过于平滑的温度分布, 模型能够再现高低温中心位置, 且其温度分布与站点观测的一致性更高。

综上, 本研究构建的 LightGBM 气温订正模型, 能在保证较高预报精度的同时, 更贴合地面站点尺度的气象监测与服务需求, 尤其可为大别山区及同类复杂地形区的精细化气温预报提供了可行技术方案。

模型虽然对于高低温中心的空间位置把握较好, 但对于极端温度量级的模拟能力尚有不足。从空间与季节的对比结果可以看到, 模型在南部水域附近订正效果有限, 表明水域的调节作用、人类活动、大气湍流等会共同调制地形因子的订正贡献。后续研究可引入高分辨率下垫面类型、云量、辐射等因子, 进一步改善模型的预报效果。

基金项目

湖北省气象局年轻科技人员专项(项目编号: 2025Q19)。

参考文献

- [1] 黄建平, 陈斌. 人工智能技术在未来改进天气预报中的作用[J]. 科学通报, 2024, 69(17): 2336-2343.
- [2] Taylor, A.A. and Leslie, L.M. (2005) A Single-Station Approach to Model Output Statistics Temperature Forecast Error Assessment. *Weather and Forecasting*, **20**, 1006-1020. <https://doi.org/10.1175/waf893.1>
- [3] 任萍, 陈明轩, 曹伟华, 等. 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, 2021, 78(6): 1002-1020.
- [4] Ham, Y.-G., Kim, J.-H. and Luo, J.-J. (2019) Deep Learning for Multi-Year ENSO Forecasts. *Nature*, **573**, 568-572. <https://doi.org/10.1038/s41586-019-1559-7>
- [5] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., *et al.* (2023) Learning Skillful Medium-Range Global Weather Forecasting. *Science*, **382**, 1416-1421. <https://doi.org/10.1126/science.adi2336>
- [6] 王珊珊, 雷彦森, 方鸿斌, 等. 基于 CatBoost 算法的长江中游冬季降水相态预报方法研究[J]. 气象, 2022, 48(9): 1153-1161.
- [7] 谭江红, 陈伟亮, 王珊珊. 一种机器学习方法在湖北定时气温预报中的应用试验[J]. 气象科技进展, 2018, 8(5): 46-50.
- [8] 杨秀梅, 黄武斌, 李天江, 等. 两种机器学习模型在甘肃小时气温预报中的应用[J]. 大气科学学报, 2025, 48(3): 476-485.
- [9] 冯景瑜, 慕秀香, 张莹莹, 等. 基于地形因素的吉林省 ECMWF 气温预报订正方法研究[J]. 气象灾害防御, 2021, 28(3): 12-17.

-
- [10] 周秋雪, 冯良敏, 陈朝平, 等. 一种适用于复杂地形下最高气温订正的机器学习方法[J]. 高原气象, 2025, 44(6): 1589-1599.
- [11] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., *et al.* (2020) The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**, 1999-2049. <https://doi.org/10.1002/qj.3803>
- [12] Uhe, P., Lucas, C., Hawker, L., Brine, M., Wilkinson, H., Cooper, A., *et al.* (2025) Fathomdem: An Improved Global Terrain Map Using a Hybrid Vision Transformer Model. *Environmental Research Letters*, **20**, 034002. <https://doi.org/10.1088/1748-9326/ada972>