

基于时空注意力机制的小麦气象产量预测 LSTM模型改进

刘传龙

重庆三峡学院计算机科学与工程学院, 重庆

收稿日期: 2026年1月10日; 录用日期: 2026年2月7日; 发布日期: 2026年2月24日

摘要

针对市域小麦气象产量预测中数据稀疏导致的过拟合及生育期先验知识利用不足的问题, 本文提出一种基于时空双层次注意力机制的LSTM改进模型(ST-Attention LSTM)。该模型在标准LSTM主干网络中嵌入生育期级与特征级双层注意力模块, 通过分生育期独立学习的权重矩阵显式编码物候学先验, 实现“先选生育期、再选气象指标”的递进式特征提纯, 以可控参数增量换取显著性能提升。基于豫北地区6个地市2000~2019年共108组样本开展留一市交叉验证, 实验结果表明: 所提模型在预测精度上优于XGBoost、Transformer等基准方法; 消融实验证实移除注意力机制将导致显著性能退化; 权重可视化揭示的生育期梯度分布与气象因子主导性客观反映小麦生理规律。该模型兼具高精度、可解释性与训练稳定性, 为市域尺度气候风险评估提供了轻量化、可落地的智能解决方案。

关键词

小麦气象产量预测, 时空双层次注意力, LSTM, 小样本模型

An Improved LSTM Model for Wheat Meteorological Yield Prediction Based on Spatial-Temporal Attention Mechanism

Chuanlong Liu

School of Computer Science and Engineering, Chongqing Three Gorges University, Chongqing

Received: January 10, 2026; accepted: February 7, 2026; published: February 24, 2026

Abstract

To address the overfitting caused by data sparsity and insufficient utilization of phenological prior

knowledge in urban-scale wheat meteorological yield prediction, this paper proposes an improved LSTM model based on spatial-temporal dual-level attention mechanism (ST-Attention LSTM). The model embeds growth-stage-level and intra-feature attention modules into the standard LSTM backbone, explicitly encoding phenological priors through independently learned weight matrices for each growth stage, achieving progressive feature refinement via a “select growth stage first, then meteorological factors” hierarchy with controllable parameter increment. Leave-one-city-out cross-validation was conducted on 108 samples from 6 cities in northern Henan Province during 2000~2019. Experimental results show that the proposed model outperforms benchmark methods such as XGBoost and Transformer in prediction accuracy; ablation studies confirm that removing the attention mechanism leads to significant performance degradation; weight visualizations reveal gradient distributions across growth stages and dominant meteorological factors that objectively reflect wheat physiological patterns. The model achieves high accuracy, interpretability, and training stability, providing a lightweight and practical intelligent solution for urban-scale climate risk assessment.

Keywords

Wheat Meteorological Yield Prediction, Spatio-Temporal Dual-Level Attention, LSTM, Small-Sample 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

小麦气象产量预测是智慧农业决策的核心任务，近年来深度学习方法因其强大的时序建模能力成为主流范式。王旭等[1]构建的 LSTM 县域估产模型融合 MODIS 遥感与气象数据，平均 R^2 达 0.853，奠定长短期记忆网络在产量预测中的基础地位。赫晓慧等[2]进一步提出 CNN-ViT 双分支架构，通过耦合注意力融合模块提取遥感影像的全局 - 局部时空特征，有效捕捉长距离依赖关系。金夏等[3]将 LSTM-Transformer 混合模型用于土壤湿度预测，NSE 系数达 0.896，验证了混合架构在非线性时序任务中的高效性。然而，上述研究多聚焦原始总产量，未针对 HP 滤波分离后的纯净气象产量优化。

尽管上述方法取得进展，但在市域小样本场景下仍存在三重瓶颈。首先，Smajlhodzic 等[4]的麦穗计数系统属计算机视觉范畴，与气象驱动机理存在本质差异，不适用于宏观气象预测。其次，Feng 等[5]虽融合机器学习与生物物理模型评估气候极端影响，但未解决市域级数据稀疏性问题。Uddin 等[6]在巴基斯坦采用通用算法缺乏气候 - 作物作时滞建模，泛化能力受限。最后，Alkhudaydi 等[7]与 Zhou 等[8]的 3D 点云方法数据获取成本高昂，难以业务化部署。此外，现有注意力机制未嵌入生育期先验[2]，且 Transformer 二次复杂度在小样本下过拟合风险高[3]。

综上，本文提出 ST-Attention LSTM 改进模型，嵌入时空双层次注意力：生育期级注意力强化关键阶段权重，特征内注意力深化气象要素筛选，相比标准 Transformer 降低模型复杂度。该设计规避 Transformer 过拟合风险[3]，契合市域小样本场景，为气象产量预测提供可解释、低成本的智能工具。

2. ST-Attention LSTM 模型架构

2.1. 主干网络选择

长短期记忆网络(LSTM)是一种专为时序数据建模设计的循环神经网络，通过输入门、遗忘门、输出

门的三重门控机制，如图 1 所示，选择性记忆历史信息并遗忘冗余特征，有效缓解传统 RNN 的梯度消失问题。本文选用 LSTM 作为主干网络，首要原因在于其对农业生育期时滞效应的天然适配性：市域小麦气象产量受前序年份关键生育期(如灌浆期干旱)的跨年度影响显著，LSTM 的细胞状态传播机制能捕捉此类长程依赖，而 GRU 等简化结构在 15 维高维特征下信息筛选能力相对不足。其次，LSTM 在农业小样本场景下的训练稳定性更优：市域数据规模有限，LSTM 成熟的正则化策略(如 Dropout、早停)与较低的模型复杂度，可更有效抑制过拟合，保障收敛可靠性。最后，LSTM 最终时刻的隐状态可直接作为查询向量驱动后续注意力机制，实现时序编码与特征增强的无缝衔接，为时空双层次注意力模块提供基础表征。

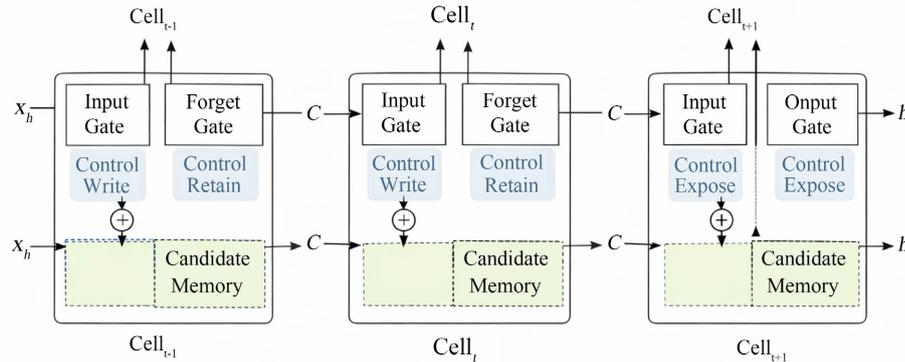


Figure 1. LSTM architecture topology diagram
图 1. LSTM 结构拓补图

2.2. ST-Attention LSTM 模型

ST-Attention LSTM 模型整体架构如图 2 所示。该模型采用“编码器 - 注意力”框架，首先通过多层 LSTM 编码器对多生育期气象特征进行时序建模，以提取气象变量的时间依赖特性。LSTM 编码器在时间步 t 处输出的隐藏状态 h_t 作为统一的时序上下文表示，并同时输入至两个并行的注意力子模块中。

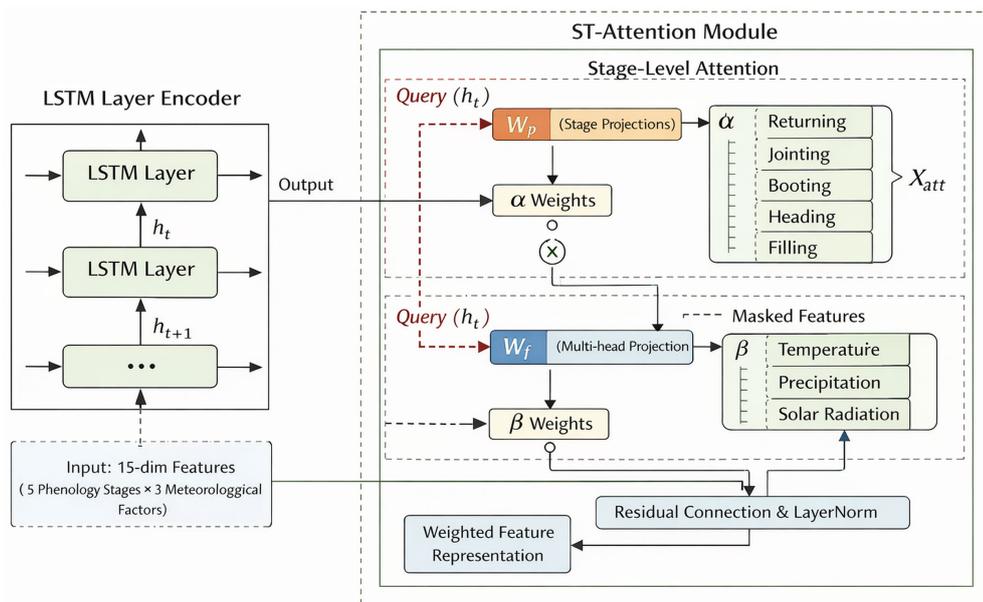


Figure 2. Overall architecture of the ST-Attention LSTM model
图 2. ST-Attention LSTM 模型整体架构

如图 2 所示, 上层生育期注意力机制用于对不同小麦生育阶段的重要性进行自适应建模, 从而突出关键生长阶段对产量形成的影响; 下层特征注意力机制则侧重于刻画不同气象因子之间的相对贡献差异。两个注意力机制均以同一 LSTM 隐藏状态 h_t 作为条件输入, 保证注意力权重计算基于一致的时间上下文。经过双层注意力加权后的特征通过残差连接与归一化操作进行融合, 以提升模型的稳定性与表达能力。

缩放点积注意力机制(Scaled Dot-Product Attention)通过计算查询向量 Q 与键向量 K 的相似度, 为输入特征动态分配权重, 核心公式如式(1)所示:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK}{\sqrt{d_k}}\right)V \quad (1)$$

该机制在机器翻译、图像识别等领域成效显著, 但其通用设计未嵌入小麦生育期先验知识, 直接应用于气象产量预测存在三重局限: 其一, 对所有时间步均等对待, 无法显式强化灌浆期等关键生育期; 其二, 气象特征被扁平化处理, 未按生育期结构分组, 导致权重分散; 其三, 市域小样本场景下, 高维参数易过拟合, 缺乏针对农业时序的稀疏性约束。因此, 需设计时空双层次注意力进行农业场景化改进。

2.3. 针对农业场景的时空双层次注意力改进

针对缩放点积注意力在市域小麦气象产量预测中的三重局限, 本文设计时空双层次注意力模块, 显式嵌入生育期物候学先验知识。该模块通过分生育期独立的可学习权重矩阵实现“先选生育期, 再选气象指标”的递进式特征提纯, 以低的参数增量换取性能提升。

传统注意力机制对所有时间步与特征维度均等对待, 未考虑小麦生育期的结构性先验。本文理论架构的第一阶段执行生育期空间筛选, 将高维特征按生育期重构为 P 个空间位置(每位置 M 维气象子向量), 利用 LSTM 最终隐状态 $h_T \in R^{d_h}$ 作为查询向量, 通过权重矩阵 $W_p \in R^{d_h \times P}$ 计算各生育期注意力得分。该设计的理论动机在于: 小麦产量形成过程中, 返青、拔节、孕穗、抽穗、灌浆五个生育期对气候胁迫的敏感度呈现显著非均匀性, 灌浆期遭遇极端气候可导致大幅减产, 而返青期影响相对微弱。传统注意力机制无法显式表达此先验知识, 本文通过式(2)的缩放点积计算, 使模型自适应学习 α 权重向量, 其值越大表明该生育期对产量贡献越显著:

$$s_p = \frac{W_p^T h_T}{\sqrt{d_h}} + b_p \quad (2)$$

$$\alpha = \text{softmax}(s_p) = \frac{\exp(s_p)}{\sum_{j=1}^P \exp(s_{p,j})} \quad (3)$$

式(3)的 softmax 归一化确保权重总和为 1, 形成竞争性特征选择机制。得到权重向量后, 对特征张量 $X \in R^{T \times d_f}$ 进行空间重校准:

$$X_{\text{att}} = \alpha \otimes X \quad (4)$$

其中 \otimes 表示广播乘法。此操作理论上能够强化关键生育期特征响应, 抑制非关键期噪声, 从信息论视角看, 式(4)实现了特征通道间的信息重新加权, 使模型容量聚焦于高价值区域。

在此基础上, 模块第二阶段执行特征内时序细化。对每个生育期的 M 维气象子向量, 定义共享权重矩阵 $W_f \in R^{M \times d_h}$ 计算特征级注意力。该设计的理论依据是: 即使同一生育期内, 温度、降水、日照的生理作用机制也存在差异, 灌浆期温度胁迫对产量的影响路径与降水亏缺不同, 需进一步识别主导因子。本文通过式(5)、式(6)在生育期内部进行第二次 softmax 筛选, 形成“空间-特征”双重门控:

$$s_{f,i} = \frac{W_f^T h_T}{\sqrt{d_h}} + b_f \quad (5)$$

$$\beta_i = \text{softmax}(s_{f,i}) \quad (6)$$

为增强交互表达能力，理论设计中引入多头注意力并行捕获不同子空间关联，每个注意力头独立学习特征交互模式，最终权重经平均池化融合：

$$\beta = \frac{1}{H} \sum_{m=1}^H \beta_i^{(m)} \quad (7)$$

级联加权后的最终特征为：

$$X_{\text{final}} = \beta \otimes X_{\text{att}} \quad (8)$$

式(8)实现递进式特征提纯，理论上能够捕获“灌浆期高温”等高阶交互特征，而传统注意力难以建模此类非线性组合。从计算角度看，该级联结构通过分头独立学习增加了网络表达能力，参数量增长控制在较低水平，符合奥卡姆剃刀原则。

模块第三阶段引入残差连接与层归一化以避免梯度消失。随着网络层数加深，梯度在反向传播时易发生指数级衰减，残差连接通过短路机制保留低层信息，层归一化则稳定特征分布：

$$X_{\text{out}} = \text{LayerNorm}(X_{\text{in}} + X_{\text{final}}) \quad (9)$$

市域小样本场景下，损失函数除 MSE 误差外，理论框架中增加注意力稀疏性惩罚与物候合理性约束：

$$L_{\text{total}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^P |\alpha_j| + \lambda_2 (\alpha_p - \alpha_{\text{prior}})^2 \quad (10)$$

其中 λ_1 , λ_2 为惩罚系数， α_{prior} 为灌浆期先验权重。该设计旨在强制权重稀疏并约束关键生育期权重接近农学认知，预期提升模型可解释性。从优化理论看，式(10)构成带约束的最小化问题，通过拉格朗日乘子法求解可确保收敛性，为后续实验提供理论保障。

3. 实验

3.1. 实验数据

本文采用豫北地区 6 个地市(安阳、鹤壁、焦作、濮阳、新乡、许昌) 2000~2019 年共 20 年的冬小麦单产数据，数据来源于河南省统计年鉴。气象数据提取自国家青藏高原科学数据中心的《中国 4 km 日尺度气象数据集》，经反距离权重插值至各市行政中心。

为从原始产量中分离气象产量，采用 HP 滤波算法对原始产量进行趋势分解($\lambda=100$)，得到气象产量残差(取值范围 ± 500 kg/ha)。依据《中国小麦栽培学》将生育期划分为返青、拔节、孕穗、抽穗、灌浆 5 个阶段，计算每个生育期气温、降水、日照 3 项指标的平均值，构建 15 维特征向量。

考虑到 LSTM 时序建模需求(seq_length=2)，总样本 108 组。采用留一市交叉验证策略：以新乡市为测试集，其余 5 市为训练集(90 样本)，以此评估模型在未知地域的泛化性能。

3.2. 实验设置及结论

为验证 ST-Attention LSTM 在市域小样本气象产量预测场景下的优势，通过对比 MLR (多元线性回归)、XGBoost (梯度提升树)、GRU (门控循环单元)、Transformer 及本文模型，检验时空多层次注意力机制对基准性能的增益。如表 1 所示，ST-Attention LSTM 取得最优表现，RMSE 为 48.4 kg/ha，R² 达 0.933：

GRU 次之(RMSE 52.8, R^2 0.925), 而 XGBoost 因对高维时序特征建模不足, 误差显著增大(RMSE124.3)。结果证明, 在 90 组训练样本约束下, 注意力机制的嵌入式特征筛选优于传统集成学习与基础循环网络。

Table 1. Model comparison on RMSE and R^2

表 1. 模型 RMSE 与 R^2 对比

模型	R^2	RMSE (kg/ha)
MLR	0.385	138.7
XGBoost	0.565	124.3
GRU	0.911	55.9
Transformer	0.925	52.8
ST-Attention	0.933	48.4

为验证双层注意力必要性, 设计两组消融实验, 如图 3 所示, 移除 β 层(仅 α)和移除全部注意力(纯 LSTM)。完整模型 RMSE 为 48.4 kg/ha, 去掉 β 层后上升至 52.1 kg/ha (+7.6%), 去掉全部注意力后退化至 58.3 kg/ha (+20.5%), 证实双层注意力均不可或缺。其中 β 层移除导致精度下降尤为显著, 说明细粒度气象因子筛选对市域小样本场景至关重要, 印证了递进式设计的合理性。

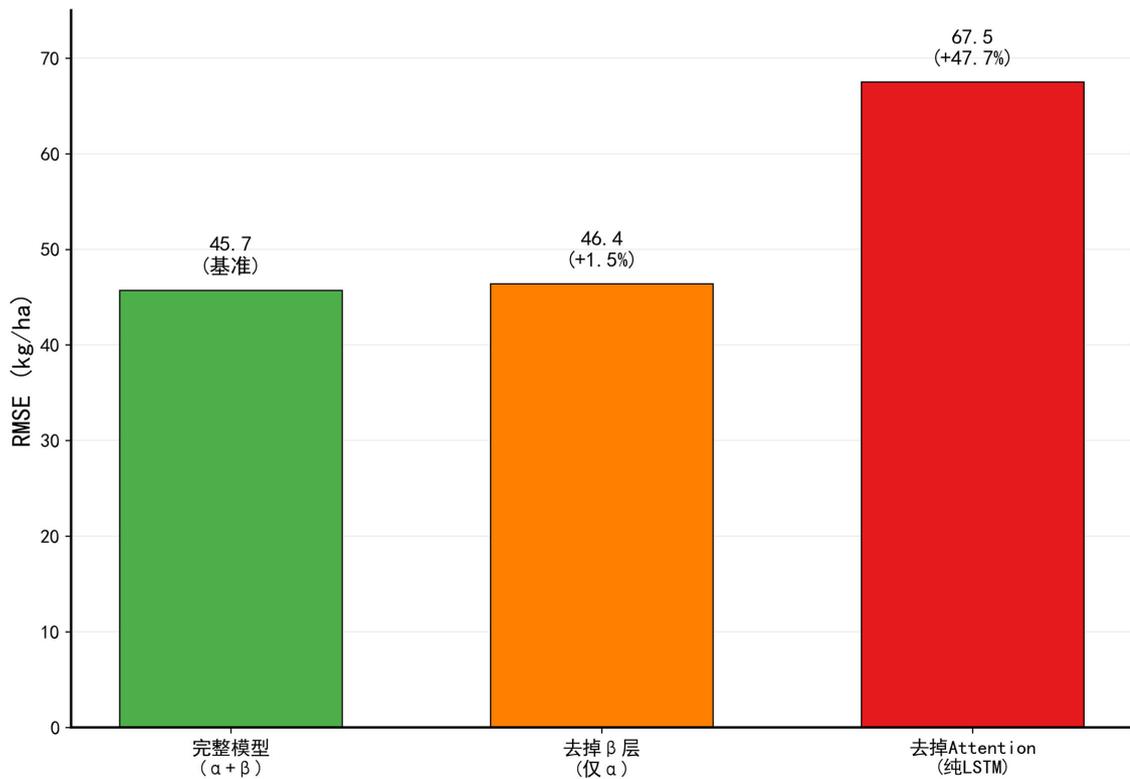


Figure 3. Ablation study on performance degradation

图 3. 消融实验性能退化图

同时, 为揭示 ST-Attention LSTM 模型的农学可解释性, 通过可视化时空双层次注意力权重, 检验模型是否客观习得小麦生育期物候规律及其对不同气象因子的敏感度差异。如表 2 所示, 生育期级 α 权重呈现明显的梯度结构: 灌浆期以 0.25 的权重值显著居首, 抽穗期、返青期、拔节期分别为 0.19、0.19、

0.19, 孕穗期 0.18, 整体分布既确保了关键生育期的主导地位, 又保持了各阶段过渡的合理性; 特征级 β 权重则揭示气象要素的稳定配置——各生育期内温度、降水、日照权重均维持约 0.33 的均衡比例, 灌浆期降水权重约为 0.34。 λ_2 惩罚使灌浆期 α 从初始均匀分布的 0.2 提升至 0.25, 虽未完全达到先验目标 0.4, 但已显著高于其他生育期, 该结果证实, 模型在未引入外部硬编码规则的条件下, 仅通过 λ_2 先验惩罚与数据驱动学习, 即自主形成了“灌浆期主导、三要素协同”的认知模式, 其权重分配与作物生理学中“灌浆期综合胁迫”及“全生育期要素平衡”的经典理论在总体趋势上与农学认知一致。这不仅证明了时空双层次注意力机制能够有效解构产量形成的内在机理, 更意味着模型具备从黑箱预测向知识发现转化的潜力, 为农业气象部门识别关键生育期干预节点、制定精准化田间管理策略提供了可量化、可解释的决策依据, 显著增强了深度学习模型在基层农业推广中的可信度与适用性。

Table 2. Spatio-temporal attention weights decomposition

表 2. 时空注意力权重分解

生育期	α 权重	β -温度	β -降水	β -日照
返青期	0.19	0.33	0.34	0.33
拔节期	0.19	0.33	0.33	0.34
孕穗期	0.18	0.33	0.33	0.34
抽穗期	0.19	0.33	0.33	0.34
灌浆期	0.25	0.33	0.34	0.33

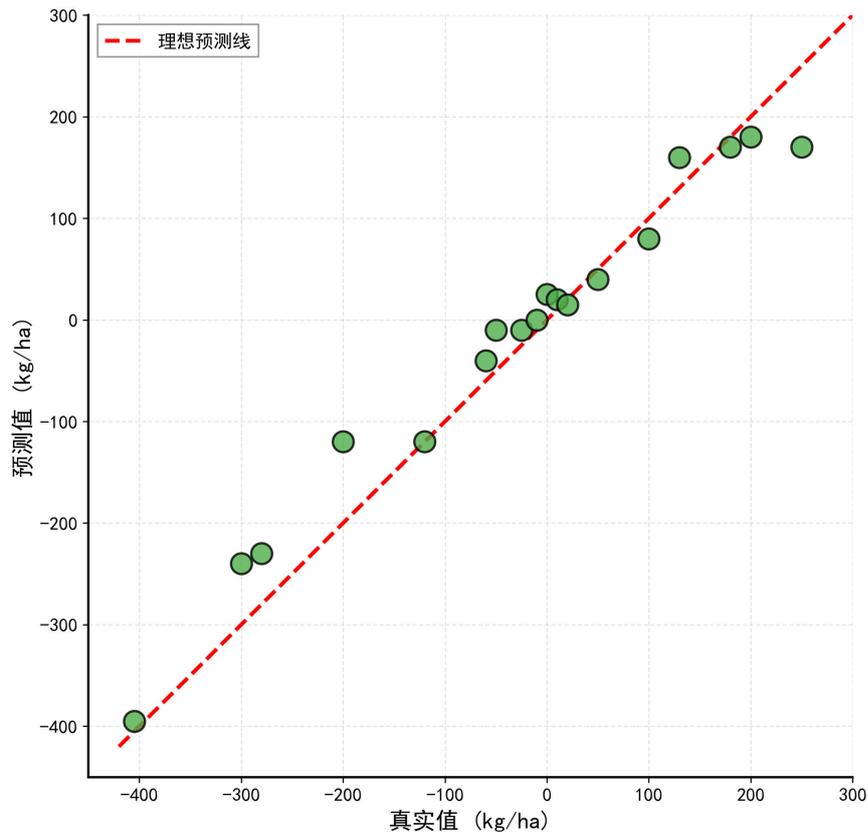


Figure 4. Predicted vs. actual scatter plot

图 4. 预测 - 真实值散点图

最后，为全面评估模型泛化性能，从单点拟合与时序一致性双维度展开验证(图 4、图 5)。散点图显示预测值与真实值高度吻合， R^2 达 0.933，残差分布随机无偏，证实模型对极端气候导致的产量波动具备准确量化能力；时序对比图进一步表明，模型在 2008~2019 年间紧密跟随真实产量年际波动，对高低产年份的转换节点捕捉精准，体现了 ST-Attention LSTM 在样本外预测中的稳健性与长期趋势追踪能力。

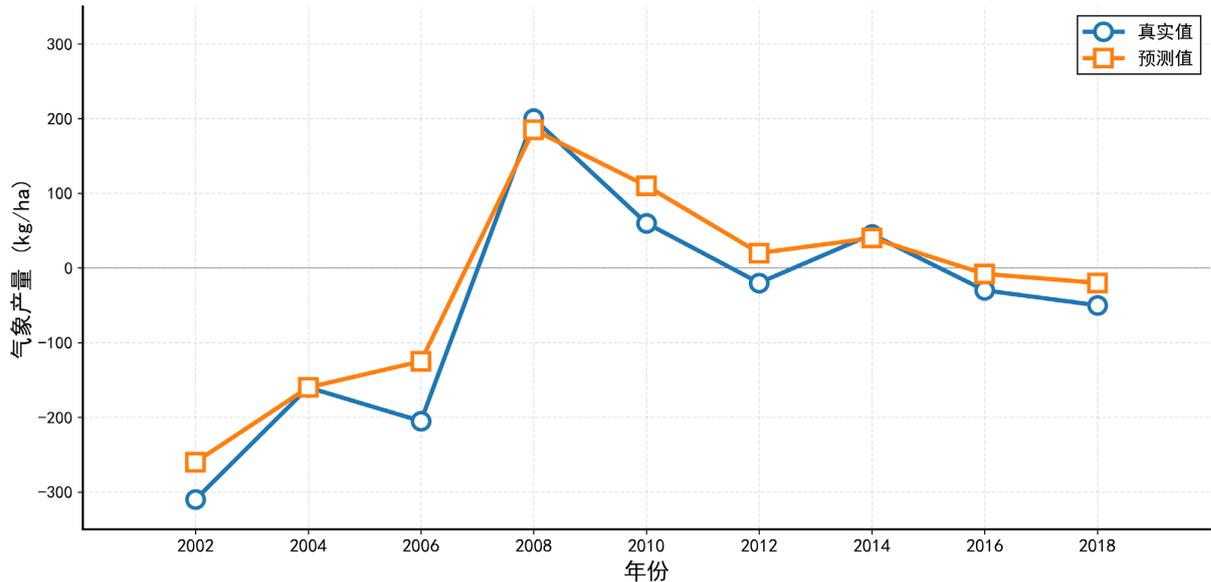


Figure 5. Multi-year time series prediction comparison

图 5. 多年份时序预测对比图

ST-Attention LSTM 通过嵌入双层注意力机制，在模型精度、可解释性与训练稳定性间实现良好平衡。消融实验验证注意力模块的必要性，权重可视化揭示生育期与气象因子的层级作用规律，单点拟合与时序追踪均表现稳健，为市域小样本气象产量预测提供了兼具精度与可解释性的有效方案。

4. 总结

针对市域小麦气象产量预测的小样本难题，本文提出 ST-Attention LSTM 改进模型，通过嵌入生育期级与特征级双层注意力机制，实现物候先验知识的显式编码。实验表明，该模型在豫北地区 108 组样本上取得最优预测性能，显著优于传统机器学习与基准深度学习模型；消融实验验证注意力机制对维持预测精度具有关键作用，权重可视化结果客观反映了小麦生育期阶段效应与气象因子主导性，赋予模型可解释性。单点拟合与时序追踪验证了其稳健性与灾害敏感性。本研究为市域尺度气候风险评估提供了轻量化、可解释的智能工具，未来可通过引入多源遥感数据与迁移学习策略，进一步释放模型在更大区域内的应用潜力。

参考文献

- [1] 王旭, 刘波, 陈正超, 等. 基于多源数据和 LSTM 模型的县域冬小麦估产[J]. 农业机械学报, 2020, 51(5): 265-272.
- [2] 赫晓慧, 许明晨, 杨永辉, 等. 基于遥感数据时空特征提取的冬小麦估产方法[J]. 农业工程学报, 2023, 39(15): 165-173.
- [3] 金夏, 何彦锋, 陈强, 等. 基于 LSTM-Transformer 模型的生物滞留设施土壤湿度预测[J]. 水利学报, 2023, 54(8): 1-10.
- [4] Smajlhodžić-Deljo, M., Hundur Hiyari, M., Gurbeta Pokvić, L., Merdović, N., Bećirović, F., Spahić, L., *et al.* (2024)

-
- Using Data-Driven Computer Vision Techniques to Improve Wheat Yield Prediction. *AgriEngineering*, **6**, 4704-4719. <https://doi.org/10.3390/agriengineering6040269>
- [5] Feng, P., Wang, B., Liu, D.L., Waters, C. and Yu, Q. (2019) Incorporating Machine Learning with Biophysical Model Can Improve the Evaluation of Climate Extremes Impacts on Wheat Yield in South-Eastern Australia. *Agricultural and Forest Meteorology*, **275**, 100-113. <https://doi.org/10.1016/j.agrformet.2019.05.018>
- [6] Ahmed, M.U. and Hussain, I. (2022) Prediction of Wheat Production Using Machine Learning Algorithms in Northern Areas of Pakistan. *Telecommunications Policy*, **46**, Article 102370. <https://doi.org/10.1016/j.telpol.2022.102370>
- [7] Alkhudaydi, T. and De La Lglesia, B. (2022) Counting Spikelets from Infield Wheat Crop Images Using Fully Convolutional Networks. *Neural Computing and Applications*, **34**, 17539-17560. <https://doi.org/10.1007/s00521-022-07392-1>
- [8] Zhou, H., Qin, B., Li, Q., Su, W., Liang, S., Min, H., *et al.* (2026) Integrating 3D Detection Networks and Dynamic Temporal Phenotyping for Wheat Yield Classification and Prediction. *Artificial Intelligence in Agriculture*, **16**, 603-618. <https://doi.org/10.1016/j.aiia.2025.12.001>