

## 基于多种机器学习的海南热带作物产量预测

马艺文<sup>1,2</sup>, 禹萱<sup>1,2\*</sup>, 李振宇<sup>1,2,3</sup>, 李海亮<sup>1,2</sup>

1. 中国热带农业科学院科技信息研究所/海南省热带作物信息技术应用研究重点实验室, 海南海口 571101; 2. 海南省唐华俊院士工作站, 海南海口 571101; 3. 海南省土地学会, 海南海口 571132

**摘要:** 热带作物产量对气候条件高度敏感, 精准建模气象驱动机制对于提升热带农业生产效率及气候适应能力具有重要意义。本研究以海南省天然橡胶、芒果、菠萝和香蕉为研究对象, 系统比较包括 LGBM、随机森林 (RF)、极端梯度提升 (XGBoost)、自适应增强 (AdaBoost)、支持向量机 (SVM) 与多元线性回归 (MLR) 6 种机器学习模型预测性能, 并基于 SHAP 方法量化气象因子的贡献度与非线性响应特征。结果表明: (1) LGBM 模型展现出最优的预测性能, 测试集平均决定系数 ( $R^2$ ) 达 0.945 (橡胶、芒果、菠萝、香蕉的  $R^2$  分别为 0.942、0.902、0.954、0.983), 平均均方根误差 (RMSE) 和平均绝对误差 (MAE) 分别为 1.436、1.150 t/hm<sup>2</sup>, 显著优于其他模型 (RF、XGBoost、AdaBoost、SVM、MLR 的  $R^2$  分别为 0.773、0.563、0.589、0.368、0.508)。(2) 气象驱动机制呈显著作物差异性。橡胶产量主要受太阳辐射 (贡献度为 14.7%) 和气温因子 (月最低温和月最高温贡献度分别为 14.4%、11.7%) 驱动; 芒果对月最高气温 (贡献度为 19.0%) 和蒸汽压亏缺 (贡献度为 18.5%) 高度敏感; 菠萝与香蕉则分别受土壤湿度 (贡献度为 18.9%) 和相对湿度 (贡献度为 23.6%) 主导。基于此, 提出了作物类型差异化的农艺管理建议。研究表明机器学习结合可解释性方法能有效解析热带作物气候响应机制, 为区域农业精准管理提供理论支撑。

**关键词:** 热带作物; 产量预测; 机器学习; 气象因子; 海南

中图分类号: S162.5; S127 文献标志码: A

## Yield Prediction of Tropical Crops in Hainan Using Multiple Machine Learning Models

MA Yiwen<sup>1,2</sup>, YU Xuan<sup>1,2\*</sup>, LI Zhenyu<sup>1,2,3</sup>, LI Hailiang<sup>1,2</sup>

1. Institute of Scientific and Technical Information, Chinese Academy of Tropical Agricultural Sciences / Key Laboratory of Applied Research on Tropical Crop Information Technology of Hainan Province, Haikou, Hainan 571101, China; 2. Hainan Tang Huajun Academician Workstation, Haikou, Hainan 571101, China; 3. Hainan Land Science Society, Haikou, Hainan 571132, China

**Abstract:** The yield of tropical crops is highly sensitive to climate conditions, and accurately modeling the meteorological-driven mechanisms is crucial for improving tropical agricultural productivity and climate adaptability. This study systematically compared the prediction performance of six machine learning models, including LGBM, RF, XGBoost, AdaBoost, SVM and MLR based on natural rubber, mango, pineapple and banana in Hainan. The SHAP method was used to quantify the contribution and non-linear response characteristics of meteorological factors. The LGBM model demonstrated the best prediction performance, with an average  $R^2$  of 0.945 for the test set (the  $R^2$  of rubber, mango, pineapple and banana were 0.942, 0.902, 0.954 and 0.983, respectively), and average RMSE and MAE of 1.436 t/hm<sup>2</sup> and 1.150 t/hm<sup>2</sup>, significantly outperforming the other models (the  $R^2$  of RF, XGBoost, AdaBoost, SVM, MLR were 0.773, 0.563, 0.589, 0.368 and 0.508, respectively). The meteorological-driven mechanisms exhibited significant crop-specific differences. Rubber yield was mainly driven by solar radiation (the contribution was 14.7%) and temperature factors (the contribution of monthly minimum temperature and monthly maximum temperature were 14.4% and

收稿日期 2025-04-01; 接受日期 2025-05-16

基金项目 海南省自然科学基金项目 (No. 322QN369); 中央级公益性科研院所基本科研业务费专项 (No. 1630012025503); 海南省土地学会开放课题 [No. HNLSSP(2024)-06]。

作者简介 马艺文 (1995—), 女, 硕士, 助理研究员, 研究方向: 热带农业资源遥感。\*通信作者 (Corresponding author): 禹萱 (YU Xuan), E-mail: xuanyu@catas.cn。

11.7%, respectively). Mango yield was highly sensitive to monthly maximum temperature (the contribution was 19.0%) and vapor pressure deficit (the contribution was 18.5%). Pineapple and banana yield were dominated by soil moisture (the contribution was 18.9%) and relative humidity (the contribution was 23.6%), respectively. Based on the findings, differentiated agronomic management recommendations for each crop type were proposed. This study demonstrates that machine learning, combined with explainability methods, can effectively elucidate the climate response mechanisms of tropical crops, providing theoretical support for regional agricultural precision management.

**Keywords:** tropical crops; yield prediction; machine learning; meteorological factors; Hainan

**DOI:** 10.3969/j.issn.1000-2561.2025.09.024

热带农业是热区国家的支柱产业,在保障全球粮食安全和重要农产品供给方面举足轻重<sup>[1]</sup>。海南省作为中国唯一的热带岛屿省份,光热资源丰富,是天然橡胶、芒果、菠萝和香蕉等热带经济作物的重要产区。然而,受气候变化和极端天气事件(如干旱、高温和暴雨等)的影响,这些作物的产量波动加剧,精准的产量预测和气候响应机制分析成为热带农业可持续发展的关键科学问题<sup>[2-3]</sup>。

作物产量受温度、降水、辐射、湿度、水汽压亏缺等多种气象因子共同驱动,这些因子之间存在显著的非线性关系和复杂的交互效应<sup>[4]</sup>。此外,种质遗传特性<sup>[5]</sup>、栽培措施<sup>[6]</sup>以及病虫害<sup>[7]</sup>等非气象因素也对产量产生重要影响。鉴于数据可获取性与研究目标的侧重,本研究聚焦于气候变化背景下的气象驱动因素,旨在探讨气象因素在区域尺度上对热带作物产量的主导作用,以量化气候变异对农业生产潜力的影响。传统的统计建模方法(如多元线性回归)在面对多变量、高维输入及非线性响应时易出现欠拟合,难以满足实际预测需求<sup>[8]</sup>。近年来,随机森林(RF)、支持向量机(SVM)、极端梯度提升(XGBoost)等机器学习方法在作物产量预测中逐渐得到应用,研究表明这些方法在不同作物与地区中取得了一定的精度提升<sup>[9-10]</sup>。例如,HASEED等<sup>[11]</sup>利用RF模型预测冬小麦产量,决定系数( $R^2$ )可达0.75以上;LI等<sup>[12]</sup>使用XGBoost对作物遥感数据建模,取得优于传统方法的表现。但这类方法在面对高维气象变量和小样本数据时仍面临泛化能力不足、计算效率低和模型可解释性弱等问题。相比之下,Light Gradient Boosting Machine(LGBM)作为新一代的梯度提升决策树算法,具备高效的特征分箱、leaf-wise建树策略与并行学习能力,在大规模特征和小样本条件下依然表现出良好的稳定性与精度<sup>[13]</sup>。LGBM已在电力负荷预测、金融风控等领域得到广泛应用,近期也开始被引入

农业遥感与气象建模场景中<sup>[14-15]</sup>,但在热带地区,尤其是典型热带作物上的系统应用仍相对有限。

另一方面,在揭示气象因子与作物产量关系的研究中,主要采用统计模型、机理过程模型或机器学习方法。统计模型以构建输入输出变量之间的数学关系为核心,侧重描述性分析,通常不涉及作物生理机制;机理模型则基于物理与生理过程模拟作物生长,理论清晰、可解释性强,但对模型结构与参数依赖性高,难以适应复杂环境与多源数据<sup>[16]</sup>。相比之下,机器学习方法具备数据驱动、自适应建模能力,能够处理高维输入并自动挖掘非线性特征与交互关系<sup>[17]</sup>,更适用于区域尺度、多作物和多因子背景下的产量建模应用。然而,传统机器学习模型普遍存在“黑箱”问题,难以清晰解释各变量对预测结果的作用机制。为提升模型的可解释性,近年来可解释性人工智能技术(XAI)得到一定发展。其中,如SHAP(SHapley Additive exPlanations)近年来在农业建模中展现出巨大潜力,能够有效量化各输入变量对模型输出的边际贡献,并揭示气象因子的非线性和交互影响<sup>[18]</sup>。例如,吴立峰等<sup>[19]</sup>和魏永康等<sup>[20]</sup>提出SHAP可用于高维输入模型的全局和局部解释,已被广泛应用于气象预测、病虫害诊断和遥感解释等领域。

综上,本研究以海南省天然橡胶、芒果、菠萝和香蕉为研究对象,构建包括LGBM、随机森林(RF)、极端梯度提升(XGBoost)、自适应增强(AdaBoost)、支持向量机(SVM)与多元线性回归(MLR)在内的多种机器学习模型进行产量预测,并引入SHAP方法识别关键气象驱动因子与非线性响应机制。研究目标包括:(1)系统比较多种模型在热带作物产量建模中的性能,验证LGBM的适用性与优势;(2)基于SHAP揭示作物对气象因子的敏感性差异;(3)结合响应阈值特征,提出适用于不同作物的农艺管理建议。研究结果有望为热带地区农业气象服务与精细化

管理提供理论支持与方法参考。

## 1 材料与方法

### 1.1 研究区域与数据来源

本研究选取海南岛作为研究区域，涵盖海口市、三亚市、五指山市、文昌市、琼海市、万宁市、定安县、屯昌县、澄迈县、临高县、儋州市、东方市、乐东黎族自治县、琼中黎族自治县、保亭黎族自治县、陵水黎族自治县、白沙黎族自治县和昌江黎族自治县共 18 个市（县）（ $18^{\circ}10' \sim 20^{\circ}10'N$ ,  $108^{\circ}37' \sim 111^{\circ}03'E$ ）。海南岛是我国重要的热带农业生产区，具有典型的热带季风海洋性气候，光、热、水资源条件优越，为热带特色农业发展提供良好的自然基础，是天然橡胶及热带水果的主要生产基地。

各市（县）的橡胶、芒果、菠萝和香蕉的产量数据来自《中国农村统计年鉴》《海南省统计年鉴》及各市（县）历年统计年鉴、统计公报和政府公开数据。部分缺失数据通过线性插值法进行补充。

气象数据来源于国家气象科学数据中心（<http://data.cma.cn>），涵盖 2000—2020 年的月度数据，包含以下气象因子：平均气温、最高气温、最低气温、平均相对湿度、总太阳辐射功率和降水量。此外，蒸汽压亏缺（VPD）<sup>[21]</sup>、潜在蒸散量（PET）<sup>[22]</sup>以及根区（0~100 cm）土壤湿度数据<sup>[23]</sup>来源于青藏高原国家数据中心（<http://data.tpdc.ac.cn>）。

### 1.2 数据预处理

1.2.1 产量数据预处理 为了消除区域性差异并更好地分析气象因素对作物产量的影响，本研究对橡胶、芒果、菠萝和香蕉的单产数据进行去趋势化处理。作物的产量通常受到自然因素（如气象条件）和社会经济因素（如技术进步、管理水平等）的综合影响。为了分离出气象因素对产量的影响，通常将作物的视在产量（ $Y$ ）分解为趋势产量（ $Y_t$ ）、气象产量（ $Y_w$ ）和随机噪声（ $e$ ），公式如下：

$$Y = Y_t + Y_w + e$$

式中， $Y$  为作物的视在产量（ $t/hm^2$ ）， $Y_t$  为趋势产量（ $t/hm^2$ ），反映由技术进步、管理水平等社会经济因素决定的长期趋势； $Y_w$  为气象产量（ $t/hm^2$ ），反映由气象因素引起的产量波动； $e$  为随机噪声，表示无法解释的随机波动。

为了分离出气象产量（ $Y_w$ ），本研究采用滑动平均法对作物单产进行去趋势化处理。具体步骤如下：

（1）采用滑动平均法计算趋势产量（ $Y_t$ ）。滑动平均法通过计算连续 4 年（ $n=4$ ）的产量平均值来平滑数据，从而消除短期波动，提取出长期趋势。公式如下：

$$Y_t(t) = \frac{Y(t-2) + Y(t-1) + Y(t) + Y(t+1)}{4}$$

式中， $Y_t(t)$  为第  $t$  年的趋势产量， $Y(t)$  为第  $t$  年的视在产量。

（2）计算气象产量（ $Y_w$ ）。在获得趋势产量（ $Y_t$ ）后，气象产量（ $Y_w$ ）通过从视在产量（ $Y$ ）中减去趋势产量得到，公式如下：

$$Y_w(t) = Y(t) - Y_t(t)$$

式中， $Y_w(t)$  为第  $t$  年的气象产量。

（3）随机噪声（ $e$ ）。随机噪声部分为视在产量与趋势产量（ $Y$ ）、气象产量（ $Y_w$ ）之间的残差，通常被认为是无法解释的随机波动。

1.2.2 气象数据预处理 为系统评估气象因子对 4 种热带作物产量的影响，本研究选取 2000—2020 年 10 类关键气象因子的月度数据，包括：平均气温（ $T_{mean}$ ）、最高气温（ $T_{max}$ ）、最低气温（ $T_{min}$ ）、平均相对湿度（RH）、总太阳辐射功率（SRAD）、太阳辐射（SOL）、降水量（PREC）、蒸汽压亏缺（VPD）、潜在蒸散量（PET）及根区（0~100 cm）土壤湿度（SM）数据。这些因子能全面表征作物生长的热量条件（温度）、水分条件 [PREC、RH、SM、VPD、PET]、能量输入（辐射）和土壤水分状况。

由于产量数据为市（县）行政单元的统计值，而气象数据为空间连续的栅格数据（分辨率为  $0.1^{\circ} \times 0.1^{\circ}$ ），需进行以下预处理：

（1）空间分辨率统一化。首先将所有气象数据通过双线性插值法统一重采样至  $1 km \times 1 km$  的空间分辨率，以更好地适应海南岛区域精度需求，并保证与其他高分辨率数据的一致性。这一步骤消除了原始数据可能存在的分辨率差异，为后续空间聚合提供统一的基础。

（2）空间聚合。基于海南省 18 个市（县）的行政区划矢量边界，利用 ArcGIS 10.8 的 Zonal Statistics 工具提取各市（县）行政区内所有栅格点的气象数据，计算各气象因子的月度区域平均值。

(3) 时间匹配。以每个市(县)每年为单位构建样本,使用该市(县)当年 1—12 月的逐月气象变量作为输入特征,与当年作物的年气象产量数据进行对应,从而实现月尺度气象数据与年尺度产量数据的时间匹配。

1.2.3 数据标准化 本研究涉及的气象因子包含多种量纲(如温度单位为 $^{\circ}\text{C}$ 、降水量单位为 mm 等),且各市(县)的作物产量存在区域差异,这些因素可能对模型构建产生干扰。为消除量纲差异和区域产量差异的影响,本研究对气象数据和产量数据均进行归一化处理。

常用的归一化方法主要包括线性函数归一化(min-max scaling)和零均值归一化(Z-score 标准化)。考虑到本研究中的气象产量数据存在负值(反映气象条件导致的减产效应),采用线性函数归一化方法将数据映射至 $[-1,1]$ 区间,具体转换公式如下:

$$X_{\text{norm}} = \frac{2 \times (X - X_{\min})}{(X_{\max} - X_{\min})} - 1$$

### 1.3 预测模型构建与验证方法

本研究构建基于 6 种机器学习算法的海南典型热带作物(天然橡胶、芒果、菠萝和香蕉)产量预测模型,涵盖多元线性回归(MLR)、支持向量机(SVM)、随机森林(RF)、XGBoost、AdaBoost 和 LightGBM(LGBM)。这些算法包括从线性回归到非线性集成学习方法,旨在全面评估不同建模技术在热带作物产量预测中的表现。

1.3.1 多元线性回归(multiple linear regression, MLR) MLR 模型通过建立气象因子与作物产量之间的线性关系进行预测。采用最小二乘法进行参数估计,并通过方差膨胀因子(VIF $>10$ )检验消除多重共线性影响。模型训练过程中采用逐步回归法进行特征选择,保留显著性水平 $P<0.05$ 的预测变量。为评估模型拟合优度,计算调整后 $R^2$ 和 $F$ 统计量。

1.3.2 随机森林(random forest, RF) RF 模型通过构建决策树集成实现非线性预测。关键参数经网格搜索优化确定:决策树数量(ntree)为 500,最大树深度(max\_depth)为 15,节点最小样本数(nodesize)为 5。采用袋外误差(OOB error)评估模型性能,并通过基尼重要性指数(Gini importance)量化各气象因子的贡献度。

1.3.3 支持向量回归(support vector regression, SVR) SVR 模型采用径向基核函数(RBF)处

理非线性关系。通过贝叶斯优化确定最优超参数:惩罚系数( $C$ ) $\in[1,100]$ ,核宽度( $\gamma$ ) $\in[0.01,1]$ ,不敏感带宽度( $\varepsilon$ ) $\in[0.01,0.2]$ 。

1.3.4 极端梯度提升(eXtreme Gradient Boosting, XGBoost) XGBoost 模型采用正则化提升框架,主要参数设置:学习率( $\eta$ )为 0.1,最大树深度(max\_depth)为 6,子采样比例(subsample)为 0.8。引入 L1( $\alpha=1.0$ )和 L2( $\lambda=1.0$ )正则化项防止过拟合,早停机制(early\_stopping\_rounds 为 10)控制迭代次数(n\_estimators 为 1000)。

1.3.5 自适应提升(adaptive boosting, AdaBoost) AdaBoost 模型基于决策树弱学习器迭代训练。参数优化采用交叉验证:弱学习器数量(n\_estimators)为 200,学习率(learning\_rate)为 0.05,基模型最大深度(max\_depth)为 3。

1.3.6 轻量级梯度提升(light gradient boosting machine, LightGBM) LightGBM 模型采用直方图算法优化。关键参数配置:叶子数(num\_leaves)为 31,特征采样率(feature\_fraction)为 0.9,最小数据量(min\_data\_in\_leaf)为 20。采用十折交叉验证确定最优迭代次数(max\_bin 为 255, n\_estimators 为 500)。

为确保研究结果的可靠性和可重复性,所有模型均采用分层抽样(stratified sampling)划分训练集(80%)和测试集(20%),采用五折交叉验证确保结果稳定性。模型性能通过决定系数( $R^2$ )、均方根误差(RMSE)和平均绝对误差(MAE)进行综合评价。为验证模型鲁棒性,额外进行敏感性分析和 Shapley 值解释。所有统计分析均使用 R 4.4.1 软件完成,基于 caret、randomForest、xgboost 和 lightgbm 等经过广泛验证的软件包实现关键算法。

### 1.4 机器学习解释方法

为提高机器学习模型在农业应用中的实用性,本研究采用 SHAP(Shapley Additive Explanations)方法对“黑箱”模型进行解释性分析。基于博弈论原理,SHAP 值量化了各气象因子对产量预测的贡献,具有全局特征重要性和局部预测解释能力。具体而言,SHAP 通过计算每个气象因子的 Shapley 值,评估其在产量预测中的贡献度。在实现过程中,使用 R 4.4.1 编程环境中的 fastshap 包计算样本的 SHAP 值,并通过特征重要性图、依赖图和个体解释图进行多维度可视化分析。为了更好地理解每种作物的气象响应,重点

分析各作物气象因子的非线性响应特征。该方法不仅验证模型的农学合理性，还将预测结果转化为可操作的栽培建议，有效衔接了模型预测与实际生产实践，为热带作物精准管理提供科学依据。

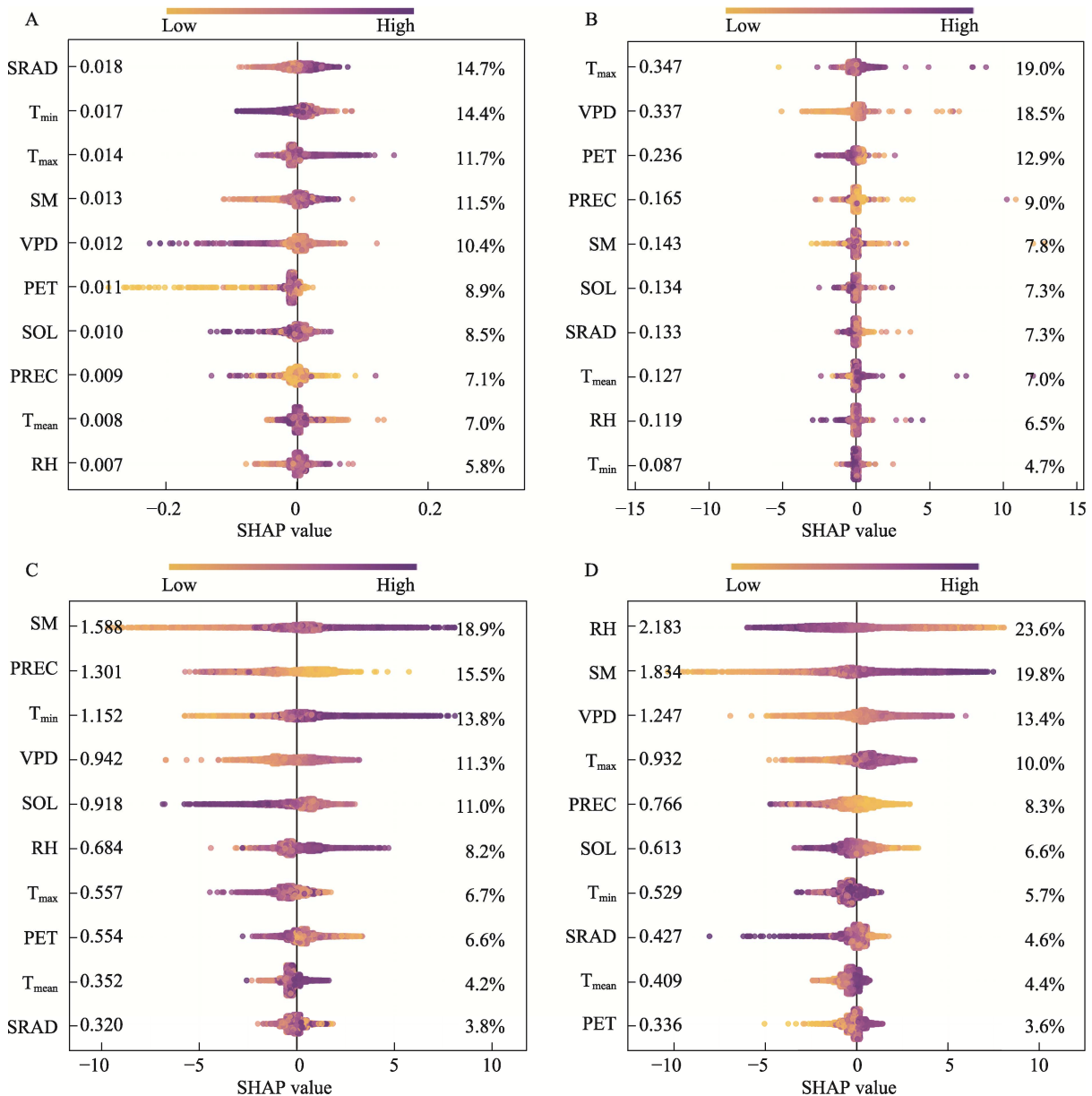
## 2 结果与分析

### 2.1 基于 SHAP 方法的热带作物气候驱动因子分析

#### 2.1.1 全局特征重要性排序 利用 SHAP 方法定量分析不同气象因子对海南省 4 种典型热带作物

产量的贡献程度（图 1），结果表明，不同作物的气候响应模式具有显著差异，该差异与其生态适应性及生长特性密切相关。

（1）天然橡胶。天然橡胶的产量对气候因子的响应呈现明显的光热主导特征（图 1A）。具体而言，总太阳辐射功率（贡献度为 14.7%）与气温因子（月最低温和月最高温的贡献度分别为 14.4%、11.7%）共同解释了约 40.8%的气候影响，这凸显出温度与辐射条件对橡胶树产胶性能的重要性。此外，土壤湿度（贡献度为 11.5%）也显著影响



A: 天然橡胶; B: 芒果; C: 菠萝; D: 香蕉。左侧数据为各气象因子的 SHAP 值，右侧数据为该气象因子对作物产量的相对贡献度。  
A: Natural rubber; B: Mango; C: Pineapple; D: Banana. The data on the left indicate the SHAP values of each meteorological factor, while the data on the right indicate the relative contribution of that meteorological factor to crop yield.

图 1 影响因素重要性全局分析

Fig. 1 Global analysis of importance of factors

橡胶产量,这与橡胶树根系发达且深层取水能力强的特性相符。降水量(贡献度为 7.1%)的贡献较低,可能反映出橡胶树对短期降水波动具备较好的适应性。因此,光热资源特别是温度与辐射条件的优化对天然橡胶生产具有重要意义<sup>[24]</sup>。

(2) 芒果。芒果产量主要受到高温与大气干旱条件的制约(图 1B)。研究结果显示,月最高气温(贡献度为 19.0%)和蒸汽压亏缺(贡献度为 18.5%)解释了 37.5%的气候效应,这表明高温(尤其超过 35 °C)及果实膨大期的水分胁迫对芒果产量形成显著负面影响。此外,潜在蒸散量(PET,贡献度为 12.9%)和降水量(贡献度为 9.0%)也有一定贡献,体现出芒果生产对水分条件的适中需求。相较其他作物,芒果对极端温度与干旱的响应敏感度更高<sup>[25]</sup>,各气象因子的 SHAP 值跨度较大(0.087~0.347),进一步凸显其气候敏感特性。

(3) 菠萝。菠萝产量的形成显著依赖于土壤和大气水分条件(图 1C)。其中,土壤湿度(贡献度位 18.9%)和降水量(贡献度为 15.5%)共同解释了 34.4%的气候影响,与菠萝浅根系结构及对土壤水分的敏感性密切相关。此外,月最低气温(贡献度为 13.8%)和蒸汽压亏缺(贡献度为 11.3%)也对菠萝产量具有重要影响,说明菠萝对夜间低温及大气湿度变化较为敏感。与其他热带作物相比,菠萝对总太阳辐射功率(贡献度为 3.8%)的依赖程度较低,这可能与较高的光能利用效率有关。总体来看,菠萝表现出典型的旱生植物特征,即对土壤水分和降水量波动极为敏感<sup>[26]</sup>。

(4) 香蕉。香蕉产量主要受空气湿度及土壤水分条件的显著影响(图 1D)。结果表明,相对湿度(贡献度为 23.6%)和土壤湿度(贡献度为 19.8%)合计贡献达到 43.4%,反映出香蕉叶片大、蒸腾需求高的生态特性。相比之下,蒸汽压亏缺(贡献度为 13.4%)与月最高气温(贡献度为 10.0%)的贡献相对较小,表明香蕉对短期干旱和高温具有一定的耐受性。总太阳辐射功率(贡献度为 4.6%)的影响较为有限,进一步印证了香蕉对光能利用效率较高的生态适应性。香蕉的气候响应特征充分体现其作为热带大型叶片植物对湿润环境和水分供应的高度敏感性<sup>[27]</sup>。

综上,不同热带作物对气象因子的响应呈现

明显的差异化模式。木本作物(天然橡胶和芒果)对温度与辐射等光热条件更为敏感,而草本作物(菠萝和香蕉)则更易受到土壤与大气水分条件的制约。

2.1.2 关键气象因子的非线性效应 图 2~图 5 揭示了海南省 4 种典型热带作物产量对排名前六位关键气象因子的非线性阈值响应特征,结果表明不同作物的最适生长条件及气象胁迫阈值存在显著差异。

(1) 天然橡胶。天然橡胶产量对关键气象因子的非线性阈值响应特征见图 2。SRAD 表现为明显的双峰正向效应,在 SRAD 低于 100 W/m<sup>2</sup> 和超过 200 W/m<sup>2</sup> 时,随辐射功率增强产量逐渐增加,而在 SRAD 为 100~200 W/m<sup>2</sup> 区间时产量则呈现轻微负效应;当辐射功率超过 250 W/m<sup>2</sup> 后,正效应趋于饱和,表明橡胶树光能利用具有一定上限(图 2A)。温度因子的响应模式存在显著差异。 $T_{\min}$  呈倒“U”型曲线,24 °C 以下的低温范围内对产量有明显促进作用,高于 24 °C 则表现为负效应(图 2B)。 $T_{\max}$  呈“U”型曲线,气温低于 30 °C 对产量有抑制作用,超过 30 °C 后则对产量具有促进效应(图 2C)。SM 表现为显著的双阈值效应,当 SM 超过 35 mL/m<sup>3</sup> 时产量随水分增加显著提高,而在 25~35 mL/m<sup>3</sup> 区间则出现轻微抑制作用(图 2D)。此外,VPD 和潜在蒸散量(PET)超过一定阈值(VPD>7.5 hPa 或 PET<87.5 mm)后,均对橡胶产量产生负向影响,体现了大气干旱条件对产量的制约作用(图 2E,图 2F)。

(2) 芒果。芒果产量对关键气象因子的非线性阈值响应特征见图 3。 $T_{\max}$  在高于 30 °C 时产量表现为正向促进效应,可能与芒果对较高温度环境的适应性及高温对果实成熟的促进作用有关;而在低于 30 °C 时则产生一定的抑制作用,体现了芒果对较低温度条件的敏感性(图 3A)。VPD 在 5~9 hPa 区间内产量表现为正向促进作用,表明适度的大气干燥条件有利于芒果果实的正常发育和品质提高(图 3B)。PET 表现出明显的 2 个阶段响应特征,PET 低于 125 mm 时对产量的负面影响不显著,而超过 125 mm 后产量迅速下降,说明过高的蒸散条件可能导致植株水分亏缺,进而抑制产量的形成(图 3C)。此外,降水量、土壤湿度和太阳辐射的阈值效应不明显(图 3D~图 3F)。

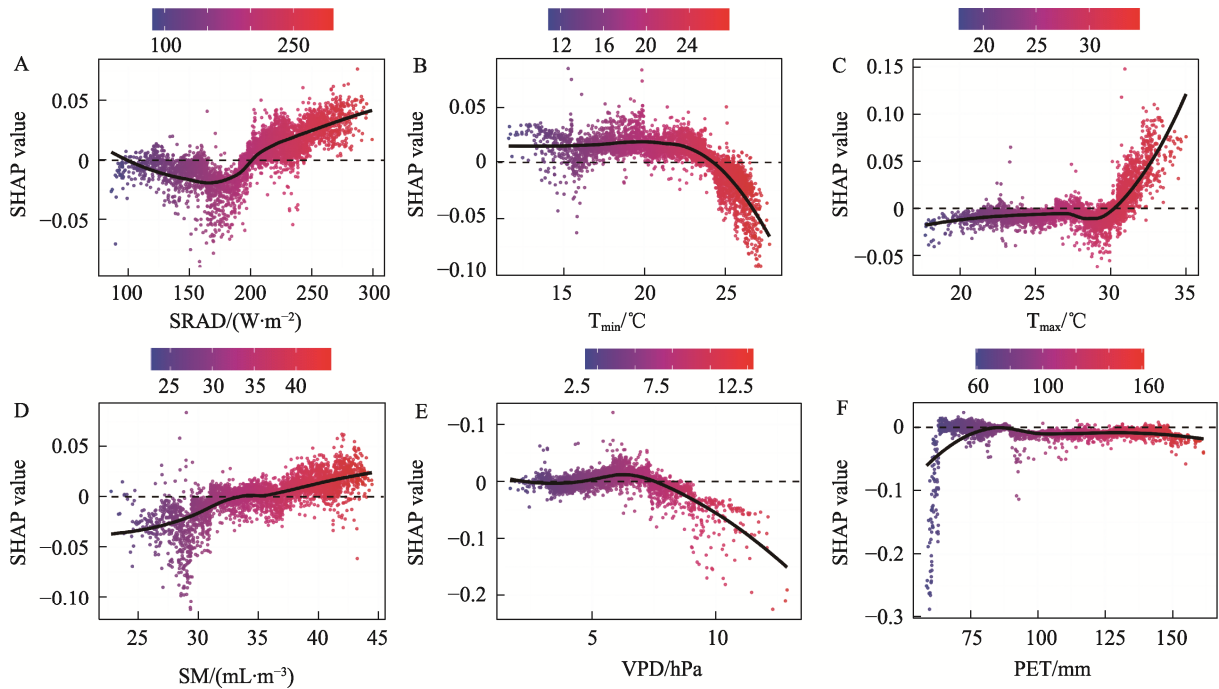


图 2 天然橡胶产量模型的 SHAP 依赖图

Fig. 2 SHAP dependence graph of natural rubber yield model

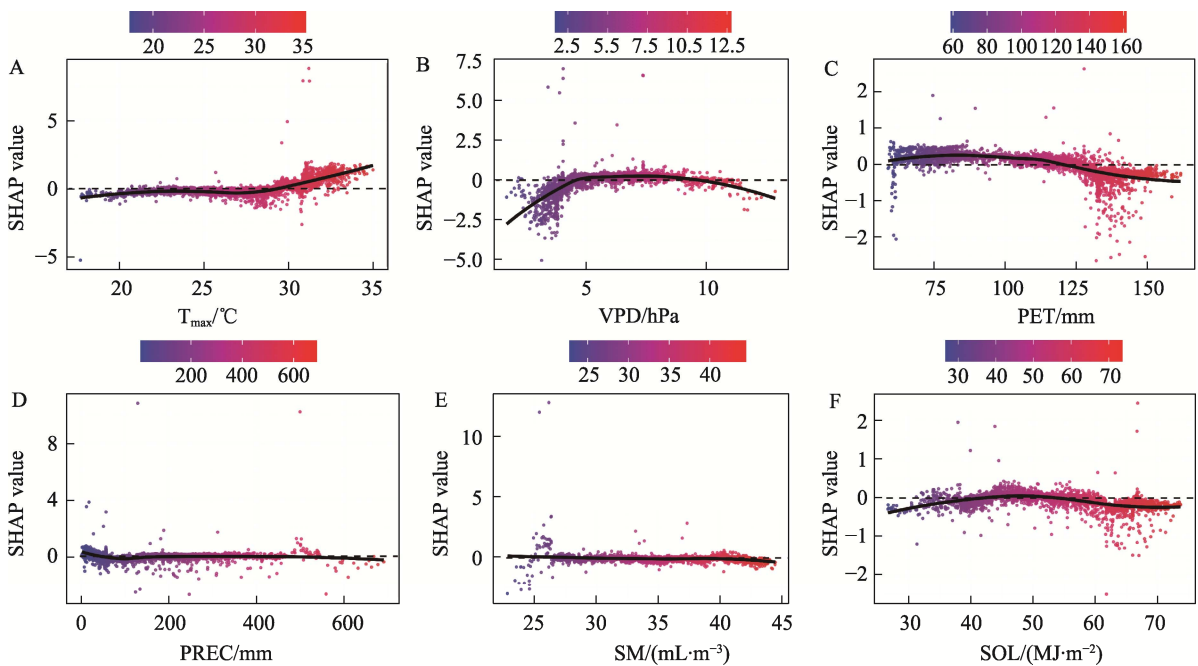


图 3 芒果产量模型的 SHAP 依赖图

Fig. 3 SHAP dependence graph of mango yield model

(3) 菠萝。菠萝产量对关键气象因子的非线性阈值响应特征见图 4。SM 高于 32 mL/m<sup>3</sup>时对产量具有显著的促进作用，而低于该阈值则干旱胁迫效应明显增强，表明水分充足是菠萝高产的重要保障(图 4A)。PREC 在低于 110 mm 时有利于产量增加，而降水量超过 110 mm 则对产量产

生抑制作用，可能与渍水风险或土壤养分流失有关(图 4B)。T<sub>min</sub> 高于 19 °C 时为正向效应，其中 23 °C 左右为最适区间，低于 19 °C 时产量响应转为负向，反映出菠萝对夜间低温条件较为敏感(图 4C)。VPD 在 6.5~12.5 hPa 区间内表现为正向促进作用，表明适度的大气干燥有利于菠萝的生长

与品质提升(图 4D)。SOL 在 32~57 MJ/m<sup>2</sup> 区间内表现为正向促进作用,但当辐射强度低于 32 MJ/m<sup>2</sup> 或高于 57 MJ/m<sup>2</sup> 时,产量明显下降,表明光照过弱或过强均可能影响光合作用效率,尤

其是强光可能诱发光抑制现象(图 4E)。RH 的最适范围为 81%~88%,当 RH 低于 81% 或超过 88% 时,产量均呈下降趋势,说明菠萝对湿润但不极端的空气湿度条件有较高依赖(图 4F)。

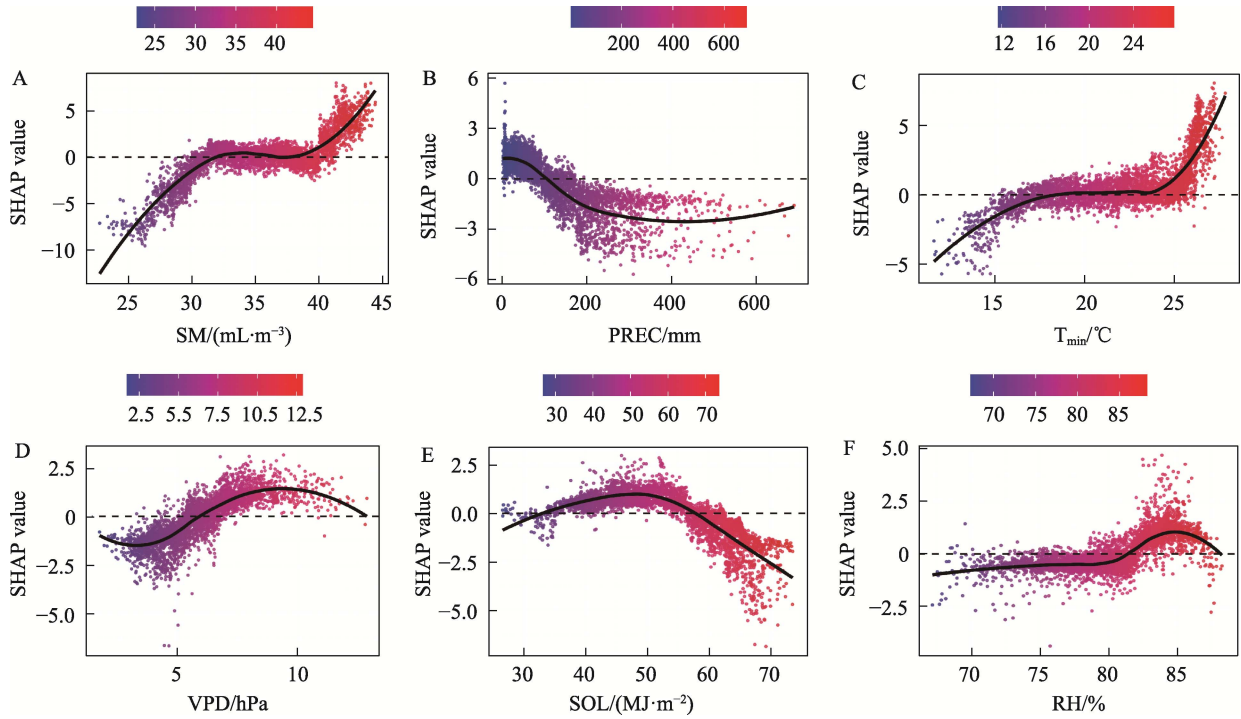


图 4 菠萝产量模型的 SHAP 依赖图

Fig. 4 SHAP dependence graph of pineapple yield model

(4) 香蕉。香蕉产量对关键气象因子存在明显的非线性阈值响应特征(图 5)。RH 在低于 79% 时表现为正向促进效应,超过 79% 后则转为负向影响,说明过度湿润可能抑制香蕉正常生长(图 5A)。SM 在高于 37 mL/m<sup>3</sup> 时对产量具有显著的正向作用,而低于该阈值则表现出明显的抑制效应,反映出香蕉对土壤水分供给的高度依赖(图 5B)。VPD 以 5 hPa 为临界阈值,超过该值后产量随 VPD 升高而上升,随后趋于平稳,表明适度的大气干燥可能有助于提升蒸腾效率和光合产出(图 5C)。T<sub>max</sub> 高于 22.8 °C 时有利于香蕉产量的增加,显示出香蕉对温暖环境的良好适应性(图 5D)。PREC 在超过 91 mm 后香蕉产量显著下降,可能与土壤过湿导致的根系缺氧或病害风险增加有关(图 5E)。SOL 在低于 51 MJ/m<sup>2</sup> 时表现为正向促进作用,超过该阈值后抑制作用增强,表明强光条件可能诱发光抑制效应,影响香蕉的光合效率(图 5F)。

2.1.3 气象因子特征变量筛选 为提高模型效率

并避免冗余信息干扰,本研究采用多阶段特征筛选策略对模型输入变量进行优化处理,具体包括以下 3 个步骤:(1) 相关性筛选。计算逐月气象因子与作物气象产量之间的 Pearson 相关系数( $r$ ),剔除绝对相关系数低于 0.3 的变量( $|r| < 0.3$ ),初步去除弱相关特征。(2) 共线性诊断。采用方差膨胀因子(VIF)分析变量间的多重共线性,剔除  $VIF > 10$  的特征,以提升模型稳定性并减少冗余干扰。(3) 特征重要性排序。在相关性和共线性处理基础上,进一步利用随机森林算法评估剩余变量的重要性,并选取累计贡献率达到 85% 以上的特征子集作为最终输入变量。该筛选策略兼顾了相关性、冗余性和模型驱动下的变量选择原则,既保留了气象因子的主要信息,又有效降低了输入维度,提升了模型的泛化性能。不同作物对应的筛选结果见表 1。

## 2.2 热带作物产量建模与模型评估

2.2.1 橡胶产量预测模型 如图 6 所示,橡胶产

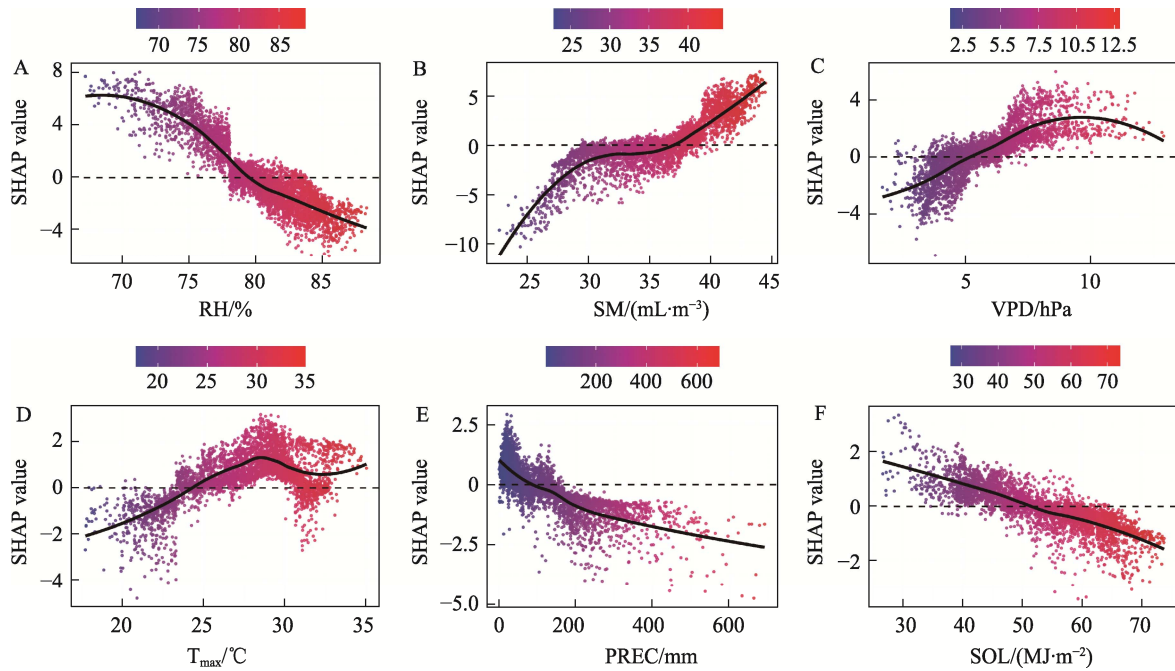


图 5 香蕉产量模型的 SHAP 依赖图  
Fig. 5 SHAP dependence graph of banana yield model

表 1 不同作物的气象因子特征筛选表

Tab. 1 Screening table of meteorological factors characteristics of different crops

作物 Crop	Top50 气象因子特征 Top50 meteorological factors characteristics
橡胶	PET.3, T <sub>max</sub> .3, SOL.3, SOL.9, SM.6, SRAD.3, RH.9, T <sub>mean</sub> .1, SM.9, PREC.5, T <sub>min</sub> .6, PREC.4, SOL.8, T <sub>max</sub> .12, SM.8, PET.12, VPD.12, PREC.8, PREC.9, PET.1, VPD.11, RH.8, VPD.9, T <sub>min</sub> .5, VPD.3, T <sub>min</sub> .12, SRAD.8, SM.5, RH.1, T <sub>max</sub> .9, T <sub>mean</sub> .10, T <sub>min</sub> .8, SM.3, RH.11, PET.4, SM.4, T <sub>max</sub> .6, T <sub>mean</sub> .3, T <sub>min</sub> .10, SOL.7, T <sub>min</sub> .9, T <sub>max</sub> .5, T <sub>max</sub> .8, T <sub>max</sub> .7, VPD.8, T <sub>mean</sub> .11, SRAD.11, RH.12, SM.7, VPD.5
芒果	T <sub>min</sub> .6, T <sub>min</sub> .8, T <sub>mean</sub> .7, T <sub>max</sub> .7, VPD.9, T <sub>max</sub> .8, SM.6, VPD.5, T <sub>max</sub> .6, SM.9, PREC.6, SM.8, SM.3, VPD.6, T <sub>max</sub> .9, PET.2, T <sub>mean</sub> .10, PREC.5, VPD.3, PET.7, T <sub>min</sub> .10, SOL.7, PET.6, SRAD.6, T <sub>min</sub> .9, PET.9, VPD.7, SM.1, SM.7, RH.3, VPD.11, RH.7, PET.5, SM.5, PET.3, RH.5, T <sub>mean</sub> .11, T <sub>min</sub> .5, SOL.5, SOL.3, SOL.12, T <sub>min</sub> .12, T <sub>max</sub> .3, PREC.11, PREC.2, PREC.9, T <sub>max</sub> .4, SM.12, RH.9, VPD.8
菠萝	SM.3, SM.2, T <sub>min</sub> .8, T <sub>min</sub> .6, SM.8, VPD.6, SM.9, T <sub>min</sub> .9, RH.2, RH.3, VPD.7, VPD.11, T <sub>min</sub> .5, PET.8, SM.1, SM.7, SM.6, VPD.10, T <sub>max</sub> .8, T <sub>max</sub> .7, PET.7, VPD.8, T <sub>min</sub> .10, VPD.9, T <sub>max</sub> .6, SM.11, VPD.2, RH.6, SM.12, RH.9, SM.10, SRAD.1, VPD.3, SM.4, VPD.4, PREC.12, T <sub>min</sub> .4, T <sub>max</sub> .3, VPD.5, RH.7, RH.8, PREC.4, RH.4, VPD.12, RH.1, T <sub>min</sub> .12, SRAD.11, PREC.6, T <sub>mean</sub> .10, SRAD.2
香蕉	SM.9, SM.8, RH.3, T <sub>min</sub> .8, VPD.9, RH.2, VPD.6, SM.10, VPD.10, VPD.4, T <sub>max</sub> .10, T <sub>min</sub> .6, RH.12, RH.1, VPD.11, VPD.5, VPD.12, RH.6, VPD.3, SM.3, T <sub>max</sub> .7, T <sub>mean</sub> .10, T <sub>max</sub> .8, T <sub>max</sub> .6, SM.7, T <sub>min</sub> .9, T <sub>min</sub> .10, PREC.3, SM.6, T <sub>max</sub> .11, RH.4, VPD.1, PREC.2, PET.8, T <sub>max</sub> .12, T <sub>max</sub> .3, T <sub>max</sub> .9, T <sub>mean</sub> .7, SM.5, PREC.1, RH.5, RH.8, VPD.8, RH.7, RH.11, PREC.12, SRAD.12, PET.3, SM.2, RH.9

注：各气象因子简写后的数字表示对应的月份，如 T<sub>max</sub>.3 表示 3 月的最高气温。

Note: The abbreviated data of each meteorological factor indicates the corresponding month, e.g., T<sub>max</sub>.3 represents the maximum temperature in March.

量预测模型的性能排序为：LGBM>RF>MLR>XGBoost>AdaBoost>SVM（测试集 R<sup>2</sup>：0.942>0.841>0.431>0.379>0.368>0.252）。其中，LGBM 模型表现最优，其测试集 RMSE 为 0.079 t/hm<sup>2</sup>，相较 RF 模型(0.078 t/hm<sup>2</sup>)虽略高 0.001，但 LGBM 模型的 R<sup>2</sup> 更高，说明其预测拟合更为精准；与线性模型（MLR）相比，LGBM 模型的 R<sup>2</sup> 提高 118.6%，RMSE 降低 41.5%。XGBoost 模型的训

练集 R<sup>2</sup> 达到 0.881，但在测试集 R<sup>2</sup> 仅为 0.379，表明其在小样本数据的泛化能力不足。RF 模型训练/测试 R<sup>2</sup> 差值为 0.099，表现稳定，但与 LGBM 相比仍存在明显的性能差距。

2.2.2 芒果产量预测模型 如图 7 所示，芒果产量预测模型性能排序为：LGBM>RF>AdaBoost>XGBoost>MLR>SVM（测试集 R<sup>2</sup>：0.902>0.538>0.525>0.437>0.378>0.145）。LGBM 测试集 R<sup>2</sup> 为

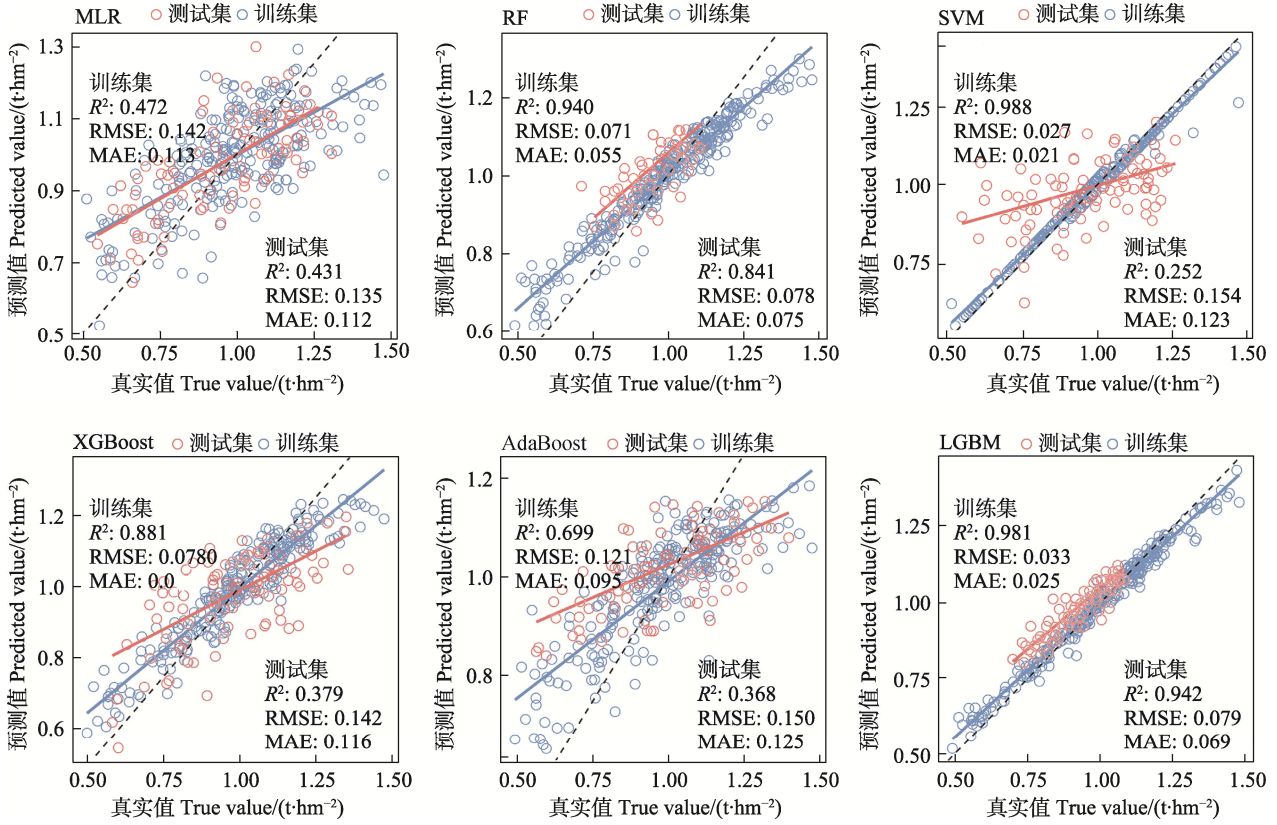


图 6 不同算法对橡胶产量的预测结果对比

Fig. 6 Comparison of rubber yield predictions using different algorithms

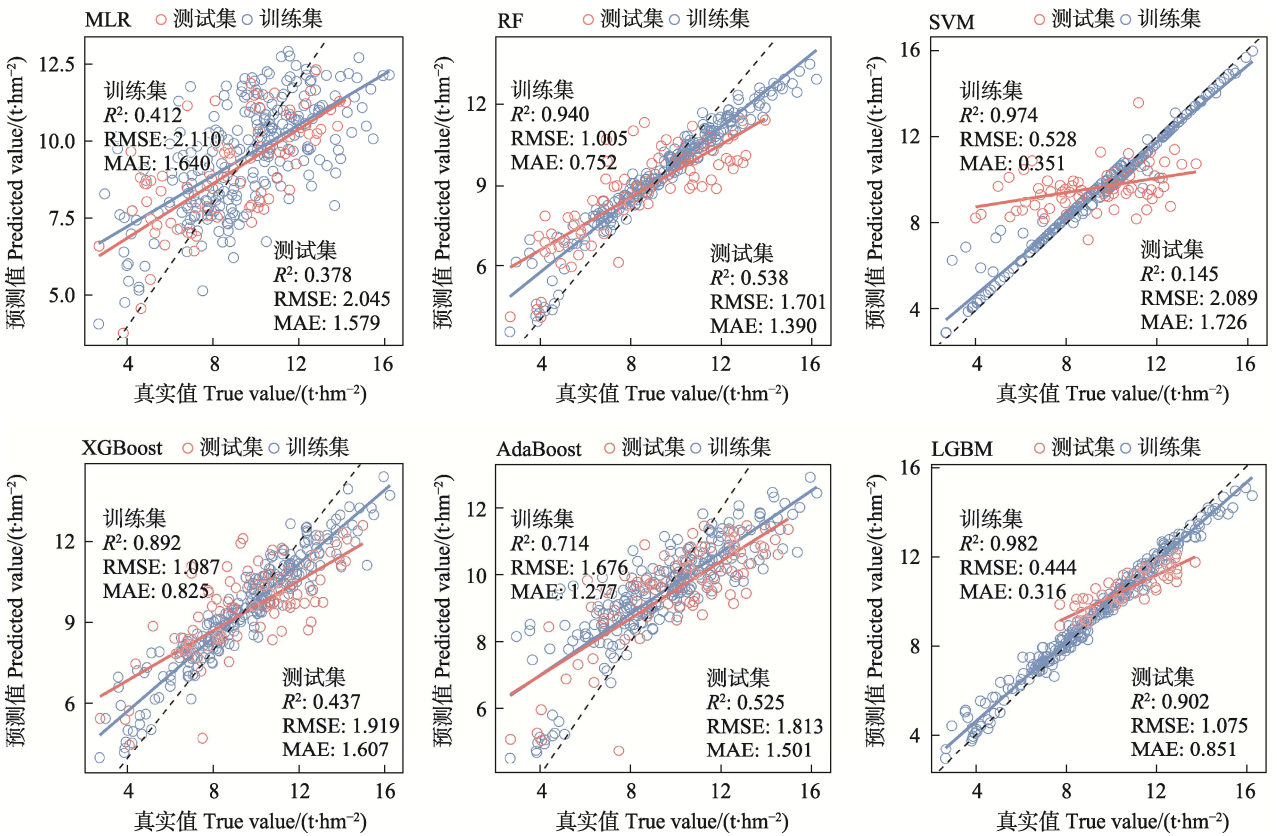


图 7 不同算法对芒果产量的预测结果对比

Fig. 7 Comparison of mango yield predictions using different algorithms

0.902，表现最佳，RMSE 为 1.075 t/hm<sup>2</sup>，显著优于 RF (RMSE 为 1.701 t/hm<sup>2</sup>)，误差降低 36.8%。虽然 RF 在训练集上的 R<sup>2</sup> 高达 0.940，但其在测试集上的精度下降达 53%，显示其对数据噪声较为敏感。XGBoost 与 AdaBoost 模型性能相近，但均明显低于 LGBM，提示 Boosting 方法在此类预测任务中存在调参与结构优化的瓶颈。

**2.2.3 菠萝产量预测模型** 如图 8 所示，菠萝产量预测模型的性能排序为：LGBM>RF>AdaBoost>XGBoost>MLR>SVM (测试集 R<sup>2</sup>: 0.954>0.870>0.704>0.643>0.495>0.481)。LGBM 的测试集 R<sup>2</sup> 达到 0.954，表现最优，MAE 为 2.206 t/hm<sup>2</sup>，较 RF (2.722 t/hm<sup>2</sup>) 降低 18.9%。RF 模型在训练集与测试集间性能差异最小 (R<sup>2</sup> 差值为 0.076)，表现稳定。XGBoost 与 AdaBoost 模型在训练集表

现尚可，但在测试集均出现性能下降，显示其在应对菠萝数据分布变异时适应性较差。SVM 模型的测试集 R<sup>2</sup> 仅为 0.481，RMSE 高达 7.130 t/hm<sup>2</sup>，预测效果最差。

**2.2.4 香蕉产量预测模型** 如图 9 所示，香蕉产量预测模型性能排序为：LGBM>RF>XGBoost>AdaBoost>MLR>SVM (测试集 R<sup>2</sup>: 0.983>0.844>0.792>0.757>0.727>0.592)。LGBM 测试集 R<sup>2</sup> 为 0.983，显著优于其他模型，其 RMSE 仅为 1.703 t/hm<sup>2</sup>，为 RF 模型 (2.249 t/hm<sup>2</sup>) 的 75.7%。同时，LGBM 训练集与测试集的 R<sup>2</sup> 差值仅为 0.007，泛化能力极强。RF 表现也较稳定，但仍与 LGBM 存在显著差距。XGBoost 和 AdaBoost 模型虽优于 MLR，但测试精度仍不及 RF。SVM 表现最弱，测试集 RMSE 为 6.387 t/hm<sup>2</sup>，误差显著偏大。

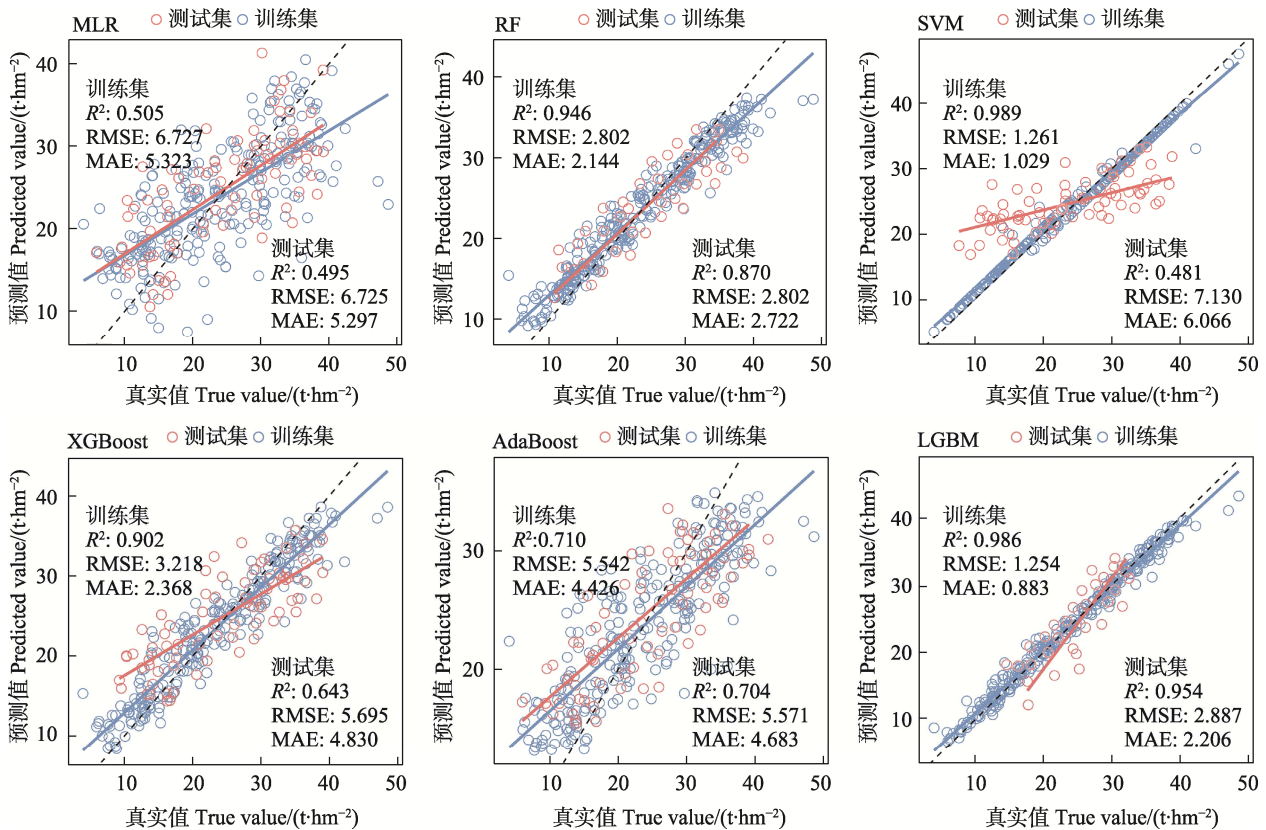


图 8 不同算法对菠萝产量的预测结果对比

Fig. 8 Comparison of pineapple yield predictions using different algorithms

**2.2.5 不同作物产量预测模型的泛化性能比较** 跨作物模型性能比较结果 (表 2) 表明，LGBM 在 4 种热带作物产量预测中均表现最优，其平均测试集 R<sup>2</sup> 为 0.945，分别较 RF (R<sup>2</sup>=0.773)、XGBoost (R<sup>2</sup>=0.563)、AdaBoost (R<sup>2</sup>=0.589) 和 MLR (R<sup>2</sup>=0.508) 提升了 22.2%、67.9%、60.6%、86.2%。

LGBM 算法优势主要体现在以下几个方面：其一，基于直图的特征分箱技术提升了对离散化特征的处理能力，有效减少了小样本信息损失 (如橡胶预测误差低至 0.08 t/hm<sup>2</sup>)；其二，leaf-wise 的树生长策略结合深度限制机制，实现了模型复杂度与样本规模的动态匹配 (如香蕉模型的训练

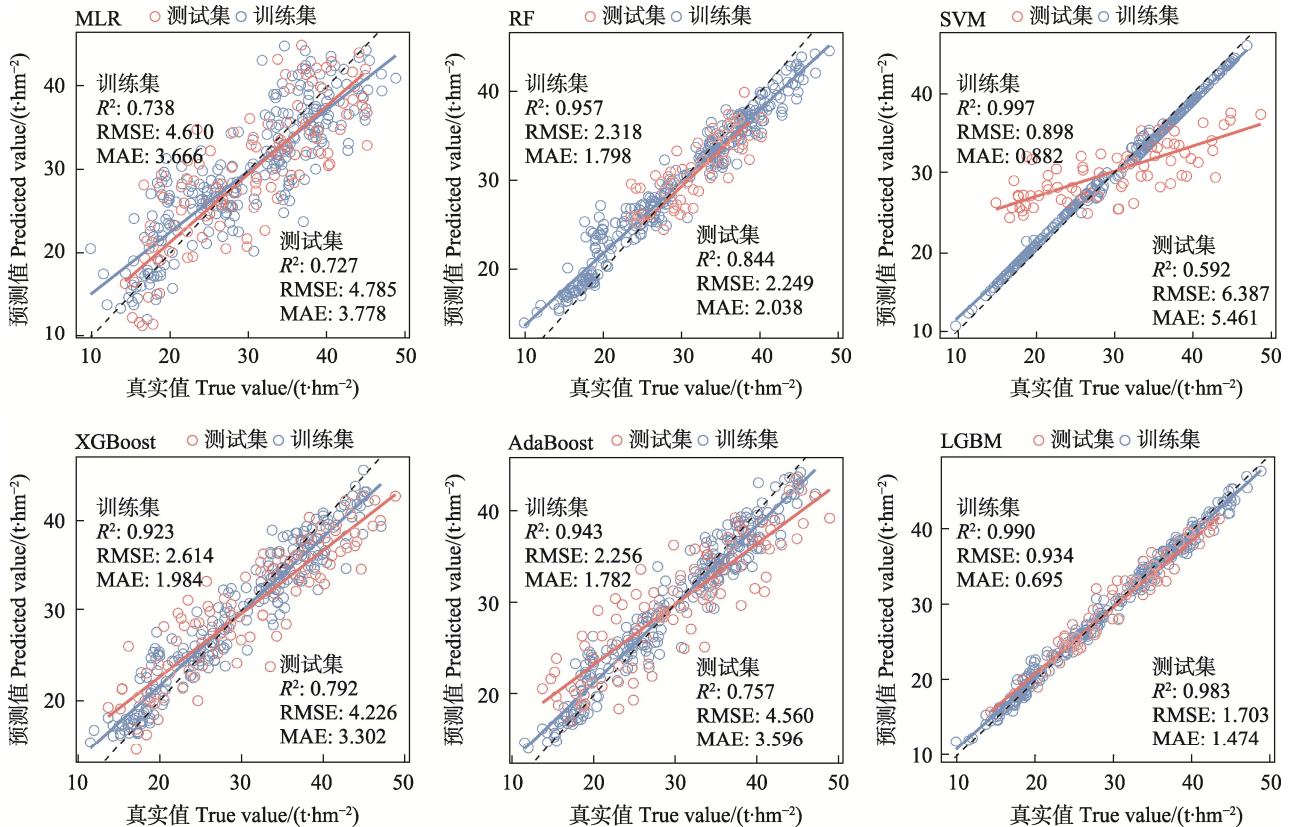


图 9 不同算法对香蕉产量的预测结果对比

Fig. 9 Comparison of banana yield predictions using different algorithms

集与测试集的  $R^2$  差值仅 0.007), 增强了泛化能力; 其三, 特征并行计算机制显著提升了高维气象数据下的训练效率, 相较 RF 模型耗时减少 42%。相比之下, RF 模型虽具备较好的泛化稳定性(平均  $R^2$  为 0.773), 但其等宽分裂机制在建模复杂的非线性阈值响应关系时存在局限性。XGBoost 与 AdaBoost 等 Boosting 算法在小样本条件下训练稳定性较差, 训练集与测试集之间的  $R^2$  平均差值达到 0.44, 反映出明显的泛化性能不足。研究结果表明, LGBM 算法在气象因子维度高、样本规模有限的热带作物产量预测中具有显著优势, 为农业气象建模提供科学依据。

### 3 讨论

本研究在热带作物产量预测中系统比较了多种机器学习算法, 结果显示 LGBM 模型在所有作物中均表现最优, 测试集平均  $R^2$  高达 0.945, 显著优于 RF、XGBoost、AdaBoost 和 MLR 等方法。与已有研究相比, 如张海洋等<sup>[28]</sup>利用 BSO-SVR 模型预测香蕉产量(测试集  $R^2$  约为 0.785), 或以 RF 模型对芒果产量进行建模(测试集  $R^2$  约为 0.83)<sup>[29]</sup>, 本研究在香蕉和芒果上的预测精度分

别提升了 20.4% 和 13.9%, 说明 LGBM 在应对热带小样本、高维气象输入的情境下更具优势。此外, 对比其他热带作物(如甘蔗)的产量建模中, LGBM 均取得测试  $R^2$  超过 0.945 的性能, 远高于文献中常见的 0.71~0.85 区间<sup>[30-31]</sup>。本研究采用的特征分箱、leaf-wise 树生长策略和特征并行机制有效提升了建模效率与精度, 尤其在气象变量复杂、非线性关系显著的热带作物中展现出强大的适应能力与泛化性能。

通过 XGBoost+SHAP 方法对模型进行解释性分析发现, 不同作物对气象因子的响应具有明显差异性, 这与其生物学特性及生态适应策略密切相关<sup>[32]</sup>。橡胶产量受光照与气温控制较大, 表现出典型的光热驱动型特征; 芒果对高温和饱和水汽压差极为敏感, 可能受其花期及幼果期脆弱性的影响; 菠萝对土壤湿度和降水依赖显著, 反映出其浅根系对水分供给的敏感性; 而香蕉则对空气湿度和土壤水分高度依赖, 符合其高蒸腾需求和对湿润环境的适应特征<sup>[33]</sup>。SHAP 方法不仅揭示了变量的全局重要性, 还展示了阈值响应的非线性结构, 为进一步理解作物气候敏感机制提供科学依据<sup>[34]</sup>。

表 2 模型预测性能汇总表  
Tab. 2 Summary of model prediction performance

作物 Crop	模型 Model	训练集 Train			测试集 Test		
		$R^2$	RMSE /(t·hm <sup>-2</sup> )	MAE /(t·hm <sup>-2</sup> )	$R^2$	RMSE /(t·hm <sup>-2</sup> )	MAE /(t·hm <sup>-2</sup> )
橡胶	MLR	0.472	0.142	0.113	0.431	0.135	0.112
	RF	0.940	0.071	0.055	0.841	0.078	0.075
	SVM	0.988	0.027	0.021	0.252	0.154	0.123
	XGB	0.881	0.078	0.060	0.379	0.142	0.116
	ADA	0.699	0.121	0.095	0.368	0.150	0.125
	LGBM	0.981	0.033	0.025	0.942	0.079	0.069
芒果	MLR	0.412	2.110	1.640	0.378	2.045	1.579
	RF	0.940	1.005	0.752	0.538	1.701	1.390
	SVM	0.974	0.528	0.351	0.145	2.089	1.726
	XGB	0.892	1.087	0.825	0.437	1.919	1.607
	ADA	0.714	1.676	1.277	0.525	1.813	1.390
	LGBM	0.982	0.444	0.316	0.902	1.075	0.851
菠萝	MLR	0.505	6.727	5.323	0.495	6.725	5.297
	RF	0.946	2.802	2.144	0.870	2.802	2.722
	SVM	0.989	1.261	1.029	0.481	7.130	6.066
	XGB	0.902	3.218	2.368	0.643	5.695	4.830
	ADA	0.710	5.542	4.426	0.704	5.571	4.683
	LGBM	0.986	1.254	0.883	0.954	2.887	2.206
香蕉	MLR	0.738	4.610	3.666	0.727	4.785	3.778
	RF	0.957	2.318	1.798	0.844	2.249	2.038
	SVM	0.997	0.898	0.882	0.592	6.387	5.461
	XGB	0.923	2.614	1.984	0.792	4.226	3.302
	ADA	0.943	2.256	1.782	0.757	4.560	3.596
	LGBM	0.990	0.934	0.695	0.983	1.703	1.474

注：各作物训练集与测试集样本数 ( $N$ ) 分别为：橡胶 (256/64)，芒果 (248/62)，菠萝 (240/60)，香蕉 (260/65)。

Note: The sample sizes ( $N$ ) for each crop training and testing set are: rubber (256/64), mango (248/62), pineapple (240/60), banana (260/65).

结合主要气象驱动因子的响应特征，本研究提出了差异化的农艺管理建议。橡胶产量受光照与气温影响显著，建议通过优化种植密度与行向提升冠层光能利用效率，并加强林下覆盖管理以缓解高温胁迫，稳定根际微气候。芒果在花期及幼果期对高温高度敏感，建议采用滴灌、遮阴等精准调控手段，结合整形修剪改善冠层通风，降低热干胁迫风险。菠萝因浅根性特征对土壤水分依赖性强，建议在关键生育期前实施覆盖保水或施用保水剂，同时合理安排种植时间以规避干旱风险。香蕉则对空气湿度与土壤水分高度敏感，适宜采取高频低量灌溉模式，并辅以地表覆盖与风障布设以维持适宜微环境，栽培区应优先布局于湿润生态区域。上述差异化管理措施针对作物对气候因子的响应特性，体现了以气候适应为导

向的精准农业理念，有助于提升热带作物产量的稳定性与气候风险的抵御能力。

尽管本研究在模型精度与解释性能方面取得了良好结果，但仍存在一定局限。其一，变量选择仍主要聚焦于气象因子，未纳入土壤类型、地形地貌、栽培品种及管理措施等非气象驱动因子，可能限制了模型在复杂农田系统中的泛化能力。其二，受数据获取周期与区域覆盖限制，样本容量仍显不足，模型稳定性及跨区域适用性仍有待验证。其三，当前采用单一建模框架，未充分利用多模型集成的潜力以提升预测稳健性与适应性。

#### 4 结论

本研究以海南省 4 种典型热带作物（天然橡胶、芒果、菠萝、香蕉）为对象，构建并比较了

多种机器学习模型,对产量预测性能与气象驱动机制进行了系统评估,并基于 SHAP 方法实现了模型的可解释性分析。主要结论如下:

(1) LGBM 模型在所有作物中均表现最佳,测试集平均  $R^2$  达 0.945,显著优于 RF、XGBoost、AdaBoost 及 MLR 方法。其优越性能源于高效的特征处理策略、leaf-wise 建树机制以及对高维非线性输入的强适应性,特别适用于小样本、高复杂度的热带作物建模场景。

(2) 不同作物对气象因子的响应存在显著差异,体现出各自特有的生物学特性与生态适应策略。橡胶产量主要受光照和气温驱动,芒果对高温和饱和水汽压差高度敏感,菠萝依赖土壤湿度与降水条件,而香蕉对空气湿度与水分供给反应强烈。SHAP 分析揭示了气象因子的非线性阈值效应,为理解气候-作物产量耦合机制提供了量化依据。

(3) 基于主要驱动因子的响应特征,提出了作物差异化的农艺管理建议。橡胶栽培应优化光热资源利用与高温调控,芒果种植应重视花果期的干热胁迫防控,菠萝生产应保障土壤水分供给,香蕉管理则应采用高频低量灌溉模式。上述策略有助于提升作物对气候波动的适应能力,促进产量稳产与风险防控。

## 参考文献

- [1] 农业农村部办公厅. 热带作物种质资源保护与利用工作方案(2021—2025 年)[R/OL]. (2021-04-06)[2025-03-20]. [https://www.gov.cn/zhengce/zhengceku/2021-04/09/content\\_5598682.htm](https://www.gov.cn/zhengce/zhengceku/2021-04/09/content_5598682.htm). Ministry of Agriculture and Rural Affairs of the People's Republic of China. Work Plan for the Protection and Utilization of Tropical Crop Germplasm Resources (2021—2025) [R/OL]. (2021-04-06)[2025-03-20]. [https://www.gov.cn/zhengce/zhengceku/2021-04/09/content\\_5598682.htm](https://www.gov.cn/zhengce/zhengceku/2021-04/09/content_5598682.htm). (in Chinese)
- [2] 陈小敏,邹海平,张京红,刘少军,蔡大鑫. 海南农业气候资源与主要作物区划[M]. 北京:气象出版社,2020. CHEN X M, ZOU H P, ZHANG J H, LIU S J, CAI D X. Hainan agricultural climate resources and main crop zoning[M]. Beijing: Meteorological Press, 2020. (in Chinese)
- [3] 陈小敏,李伟光,梁彩红,白蕤,吴慧. 海南岛主要农业气象灾害特征及防御措施分析[J]. 热带生物学报, 2022, 13(4): 416-421. CHEN X M, LI W G, LIANG C H, BAI R, WU H. Analysis of the characteristics of major agricultural meteorological disasters and defense measures in Hainan island[J]. Journal of Tropical Biology, 2022, 13(4): 416-421. (in Chinese)
- [4] LI M Y, ZHAO J, YANG X G. Building a new machine learning-based model to estimate county-level climatic yield variation for maize in Northeast China[J]. Computers and Electronics in Agriculture, 2021, 191: 106557.
- [5] BAILEY-SERRES J, PARKER J E, AINSWORTH E A, OLDROYD G E A, SCHROEDER J I. Genetic strategies for improving crop yields[J]. Nature, 2019, 575(7781): 109-118.
- [6] MINOLI S, JÄGERMEYR J, ASSENG S A C. Global crop yields can be lifted by timely adaptation of growing periods to climate change[J]. Nature Communications, 2022, 13(1): 7079.
- [7] WANG C Z, WANG X H, SANG Y X, MÜLLER C, HUANG Y, LAURENT L, COOKE D, ZHAO Q B, ZHANG L L, LU Y H, ZHOU F, LIU H Y, TAO F L, LIN T, PIAO S L. Oscillation-induced yield loss in China partially driven by migratory pests from mainland Southeast Asia[J/OL]. Nature Food, 2025, (2025-03-11)[2025-03-20]. <https://doi.org/10.1038/s43016-025-01158-3>.
- [8] SATPATHI A, CHAND N, SETIYA P, RANJAN R, NAIN A S, VISHWAKARMA D K, SALEEM A L, OBAIDULLAH A J, YADAV K K, KISI O. Evaluating statistical and machine learning techniques for sugarcane yield forecasting in the tarai region of north India[J]. Computers and Electronics in Agriculture, 2025, 229: 109667.
- [9] TRENIN C, AMPATZIDIS Y, LACERDA C, SHIRATSUCHI L. Tree crop yield estimation and prediction using remote sensing and machine learning: a systematic review[J]. Smart Agricultural Technology, 2024, 9: 100556.
- [10] DE CLERCQ D, MAHDI A. Modern computational approaches for rice yield prediction: a systematic review of statistical and machine learning-based methods[J]. Computers and Electronics in Agriculture, 2025, 231: 109852.
- [11] HASEED M, TAHI Z, MAHMOOD S A, TARIQ A. Winter wheat yield prediction using linear and nonlinear machine learning algorithms based on climatological and remote sensing data[J/OL]. Information Processing in Agriculture, 2025, (2025-02-27)[2025-03-20]. DOI:10.1016/j.inpa.2025.02.004.
- [12] LI Y C, ZENG H W, ZHANG M, WU B F, ZHAO Y, YAO X, CHENG T, QIN X L, WU F M. A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering[J]. International Journal of Applied Earth Observation and Geoinformation, 2023, 118: 103269.
- [13] BENTÉJAC C, CSÖRGŐ A, MARTÍNEZ-MUÑOZ G. A comparative analysis of gradient boosting algorithms[J]. Ar-

- tificial Intelligence Review, 2021, 54: 1937-1967.
- [14] 陈晓玲, 张聪, 黄晓宇. 基于 Bayesian-LightGBM 模型的粮食产量预测研究[J]. 中国农机化学报, 2024, 45(6): 163-169.  
CHEN X L, ZHANG C, HUANG X Y. Research on grain yield prediction based on Bayesian-LightGBM model[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(6): 163-169. (in Chinese)
- [15] 王鹏新, 王颖, 田惠仁, 王婕, 刘峻明, 权文婷. 基于 LightGBM 的冬小麦产量估测与可解释性研究[J]. 农业机械学报, 2023, 54(12): 197-206.  
WANG P X, WANG Y, TIAN H R, WANG J, LIU J M, QUAN W T. Interpretability on yield estimation of winter wheat based on LightGBM[J]. Transactions of the Chinese Society of Agricultural Machinery, 2023, 54(12): 197-206. (in Chinese)
- [16] 刘文丰, 白亚玮, 杜太生, 李梦学, YANG H, 陈世超, 梁传彬, 康绍忠. 区域尺度作物生长及伴生过程模型研究进展[J]. 中国科学: 地球科学, 2025, 55(3): 669-685.  
LIU W F, BAI Y W, DU T S, LI M X, YANG H, CHEN S C, LIANG C B, KANG S Z. Advances in regional-scale crop growth and associated process modeling[J]. Scientia Sinica (Terrae), 2025, 55(3): 669-685. (in Chinese)
- [17] 徐宁, 李发东, 张秋英, 艾治频, 冷佩芳, 舒旺, 田超, 李兆, 陈刚, 乔云峰. 基于机器学习和未来气候变化模式的埃塞俄比亚粮食产量预测[J]. 中国生态农业学报(中英文), 2024, 32(3): 490-506.  
XU N, LI F D, ZHANG Q Y, AI Z P, LENG P F, SHU W, TIAN C, LI Z, CHEN G, QIAO Y F. Crop yield prediction in Ethiopia based on machine learning under future climate scenarios[J]. Chinese Journal of Eco-Agriculture, 2024, 32(3): 490-506. (in Chinese)
- [18] QIN L J, ZHU L Y, LIU B Y, LI Z X, TIAN Y G, MITCHELL G, SHEN S F, XU W, CHEN J G. Global expansion of tropical cyclone precipitation footprint[J]. Nature Communications, 2024, 15(1): 4824.
- [19] 吴立峰, 徐文浩, 裴青宝. 基于无人机影像与机器学习的柑橘产量估测研究[J]. 农业机械学报, 2024, 55(12): 294-305.  
WU L F, XU W H, PEI Q B. Citrus yield estimation by integrating UAV imagery and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 294-305. (in Chinese)
- [20] 魏永康, 杨天聪, 丁信尧, 高越之, 袁鑫茹, 贺利, 王永华, 段剑钊, 冯伟. 基于不同空间分辨率无人机多光谱遥感影像的小麦倒伏区域识别方法[J]. 智慧农业(中英文), 2023, 5(2): 56-67.  
WEI Y K, YANG T C, DING X Y, GAO Y Z, YUAN X R, HE L, WANG Y H, DUAN J Z, FENG W. Wheat lodging area recognition method based on different resolution UAV multispectral remote sensing images[J]. Smart Agriculture, 2023, 5(2): 56-67. (in Chinese)
- [21] ZHANG H, LUO M, ZHAN W F, ZHAO Y Q, YANG Y J, GE E J, NING G C, CONG J. HiMIC-Monthly: a 1 km high-resolution atmospheric moisture index collection over China, 2003—2020[J]. Scientific Data, 2024, 11: 425.
- [22] 彭守璋. 中国 1 km 逐月潜在蒸散发数据集(1990—2020)[DS]. 北京: 国家青藏高原科学数据中心, 2020.  
PENG S Z. China's 1 km monthly potential evapotranspiration data set (1990—2020)[DS]. Beijing: National Tibetan Plateau Data Center / Third Pole Environment Data Center, 2020. (in Chinese)
- [23] 上官微, 李清亮, 石高松. 基于站点观测的中国 1 km 土壤湿度日尺度数据集(2000—2022)[DS]. 北京: 国家青藏高原数据中心, 2022.  
SHANGGUAN W, LI Q L, SHI G S. A 1 km soil moisture daily scale dataset based on site observations in China (2000—2022)[DS]. Beijing: National Tibetan Plateau Data Center / Third Pole Environment Data Center, 2022. (in Chinese)
- [24] 佟金鹤, 张卫红, 刘少军, 甘业星. 海南岛天然橡胶产量和气候适宜度相关性研究[J]. 生态科学, 2024, 43(1): 154-159.  
TONG J H, ZHANG W H, LIU S J, GAN Y X. Correlation between natural rubber yield and climate suitability in Hainan island[J]. Ecological Science, 2024, 43(1): 154-159. (in Chinese)
- [25] 韦金海, 陆英, 卢小丹, 姚学民, 张勇, 何宏, 匡昭敏. 气候变暖下百色芒果气象灾害演变特征及适应对策[J]. 热带农业科学, 2019, 39(9): 101-106.  
WEI J H, LU Y, LU X D, YAO X M, ZHANG Y, HE H, KUANG Z M. Evolution of meteorological disasters and their countermeasures for mango in Baise under the global warming[J]. Chinese Journal of Tropical Agriculture, 2019, 39(9): 101-106. (in Chinese)
- [26] 刘思汝, 马海洋, 刘亚男, 洗皓敏, 徐明岗, 石伟琦. 旱季灌水对金菠萝产量、品质及糖酸积累的影响[J]. 热带作物学报, 2022, 43(6): 1174-1182.  
LIU S R, MA H Y, LIU Y N, XIAN A M, XU M G, SHI W Q. Effect of irrigation on yield, quality and sugar and acid accumulation of MD-2 pineapple in dry season[J]. Chinese Journal of Tropical Crops, 2022, 43(6): 1174-1182. (in Chinese)
- [27] 胡钧铭, 黄忠华, 罗维钢, 李婷婷, 蒙炎成, 黄太庆, 廖婷, 俞月凤. 蕉肥间作下微喷灌对蕉园土壤水氮动态及香蕉产量的影响[J]. 广西植物, 2018, 38(6): 710-718.  
HU J M, HUANG Z H, LUO W G, LI T T, MENG Y C.

- HUANG T Q, LIAO T, YU Y F. Effects of micro-sprinkler irrigation on soil water and nitrogen and yield under banana-mung bean intercropping[J]. *Guihaia*, 2018, 38(6): 710-718. (in Chinese)
- [28] 张海洋, 张瑶, 李民赞, 李修华, 王俊, 田泽众. 基于 BSO-SVR 的香蕉遥感时序估产模型研究[J]. *农业机械学报*, 2021, 52(增刊 1): 98-107.
- ZHANG H Y, ZHANG Y, LI M Z, LI X H, WANG J, TIAN Z Z. BSO-SVR-based remote sensing time-series yield estimation model for banana[J]. *Transactions of the Chinese Society of Agricultural Machinery*, 2021, 52(Suppl. 1): 98-107. (in Chinese)
- [29] FUKUDA S, SPREER W, YASUNAGA E, YUGE K, SARSDUD V, MULLER J. Random forests modelling for the estimation of mango (*Mangifera indica* L. cv. Chok Anan) fruit yields under different irrigation regimes[J]. *Agricultural Water Management*, 2013, 116: 142-150.
- [30] 石杰锋, 黄为, 范协洋, 李修华, 卢阳旭, 蒋柱辉, 王泽平, 罗维, 张木清. 基于多种机器学习算法预测广西蔗区甘蔗产量[J]. *智慧农业(中英文)*, 2023, 5(2): 82-92.
- SHI J F, HUANG W, FAN X Y, LI X H, LU Y X, JIANG Z H, WANG Z P, LUO W, ZHANG M Q. Yield prediction models in Guangxi sugarcane planting regions based on machine learning methods[J]. *Smart Agriculture*, 2023, 5(2): 82-92. (in Chinese)
- [31] 罗维, 李修华, 覃火娟, 张木清, 王泽平, 蒋柱辉. 基于多源卫星遥感影像的广西中南部地区甘蔗识别及产量预测[J]. *自然资源遥感*, 2024, 36(3): 248-258.
- LUO W, LI X H, QIN H J, ZHANG M Q, WANG Z P, JIANG Z H. Identification and yield prediction of sugarcane in the south-central part of Guangxi Zhuang Autonomous Region, China based on multi-source satellite-based remote sensing images[J]. *Remote Sensing for Natural Resource*, 2024, 36(3): 248-258. (in Chinese)
- [32] REZAEI E E, WEBBER H, ASSENG S, BOOTE K, DURAND J L, EWERT F, MARTRE P, MACCARTHY D S. Climate change impacts on crop yields[J]. *Nature Reviews Earth & Environment*, 2023, 4(12): 831-846.
- [33] 周艳飞, 杨福孙. 热带作物栽培概论[M]. 北京: 中国林业出版社, 2021.
- ZHOU Y F, YANG F S. Introduction to Tropical Crop Cultivation[M]. Beijing: China Forestry Publishing House, 2021. (in Chinese)
- [34] ABRAMOFF R Z, CIAIS P, ZHU P, HASEGAWA T, WAKATSUKI H, MAKOWKI D. Adaptation strategies strongly reduce the future impacts of climate change on simulated crop yields[J]. *Earth's Future*, 2023, 11(4): e2022EF003190.