机器学习揭示土壤养分对玉米产量形成驱动作用——以云南地区为例

孙佳泽[1,2], 曲明山[3], 杨积忠[4], 娄垂新[4], 李光伟[2], 张钟莉莉[5*]

(1. 中国农业大学信息与电气工程学院,北京 100083,中国; 2. 北京市农林科学院信息技术研究中心,北京 100097,中国; 3. 北京市农业技术推广站,北京 100029,中国; 4. 云南供销产业投资有限公司,云南昆明 650000,中国; 5. 北京市农林科学院智能装备技术研究中心,北京 100097,中国)

摘 要: [**目的/意义**] 云南省玉米种植区地形破碎、气候垂直分异明显,土壤养分与产量均存在较强空间差异。本研究旨在明确主要土壤养分对玉米产量差异的影响,提高复杂山地环境下产量预测结果的可解释性,构建融合空间信息与土壤养分特征的机器学习模型,为分区施肥和耕地质量提升提供依据。[**方法**] 基于2005—2020年云南省测土配方施肥土壤基础养分数据和5 km玉米产量栅格数据,选取有机质、全氮、有效磷、速效钾及经纬度变量,采用粒子群优化-极端梯度提升模型(Particle Swarm Optimization-Extreme Gradient Boosting, PSO-XGBoost)开展产量预测,并与多种机器学习模型进行对比。为检验空间相邻样本可能导致的精度高估风险,进一步设置空间分块交叉验证和去除经纬度的消融实验;同时引入沙普利加和解释(SHapley Additive exPlanations, SHAP)方法分析各变量贡献及主导因子的空间分布。[**结果和讨论**] PSO-XGBoost在随机划分下决定系数R^2为0.91,优于LightGBM(Light Gradient Boosting Machine)、CatBoost(Categorical Boosting)和未优化极端梯度提升模型等;空间分块验证下R^2为0.847,说明模型在更严格的空间验证条件下仍保持较好的预测能力。去除经纬度后R^2降至0.698,表明空间背景信息对云南玉米产量预测具有重要作用。SHAP分析显示,经纬度反映的区域环境差异贡献较高,有机质、全氮、有效磷和速效钾均对产量形成具有影响,且不同区域主导因子存在差异。[**结论**] 云南玉米产量受土壤养分与空间环境共同影响,单一养分难以完全解释产量差异。PSO-XGBoost结合SHAP可用于识别区域产量差异及主要养分约束,为云南山地玉米分区施肥和耕地质量管理提供参考。

关键词: 玉米产量预测;粒子群优化-极端梯度提升模型;沙普利加和解释;土壤养分;空间异质性;机器学习

中图分类号: S24;TP181 **文献标识码:** A **文章编号:** SA202508029

引用格式:孙佳泽, 曲明山, 杨积忠, 娄垂新, 李光伟, 张钟莉莉. 机器学习揭示土壤养分对玉米产量形成驱动作用——以云南地区为例[J]. 智慧农业(中英文), 2026, 8(4): 180-191. DOI: 10.12133/j.smartag.SA202508029

SUN Jiaze, QU Mingshan, YANG Jizhong, LOU Chuixin, LI Guangwei, ZHANG Zhonglili. Machine Learning Reveals the Driving Effects of Soil Nutrients on Maize Yield Formation: A Case Study of Yunnan, China[J]. Smart Agriculture, 2026, 8(4): 180-191. DOI: 10.12133/j.smartag.SA202508029 (in Chinese with English abstract)

0 引 言

玉米是中国仅次于水稻和小麦的主要粮食作物之一,目前全国玉米种植面积超过3.3×10^7 hm^2,年产量约占粮食总产的35%,在保障国家粮食安全、支撑畜牧业发展,以及稳定农民生计方面具有重要作用[1]。在玉米生长过程中,氮、磷、钾,以及有机质等土壤养分共同影响植株生长、籽粒形成和产量稳定性,养分供应不足或比例失衡均可能造成减产。云南省地形起伏大,山地、丘陵和坝区交错分布,土壤成因和耕作条件差异明显,这使得玉米产区土壤肥力和产量水平均呈现较强空间分异[2,3]。已有研究发现,昆明地区玉米产量与土壤

收稿日期:2025-08-27

基金项目:云南省重大科技专项计划项目(202202AE090013);国家玉米产业技术体系(CARS-02);北京市农林科学院优秀青年科学家培养(YXQN202304-C)

作者简介:孙佳泽,博士研究生,研究方向为智慧水肥调控技术与装备。E-mail:sunjiaze@cau.edu.cn

*通信作者:张钟莉莉,博士,副研究员,研究方向为智慧水肥调控技术与装备,E-mail:zhangzll@nercita.org.cn

copyright©2026 by the authors (登录www.smartag.net.cn免费获取电子版全文)

有效磷（Available Phosphorus, AP）和速效钾（Available Potassium, AK）含量呈显著正相关，而与碱解氮的相关性相对较弱。由此可见，在云南部分产区，磷、钾等养分的有效供应可能是影响玉米增产的重要环节。因而，从区域尺度识别土壤养分状况及其对产量差异的影响，不仅有助于明确主要限制因子，也可为分区施肥和缩小区域产量差距提供依据。

农作物产量估算和土壤肥力评价早期多依托田间试验和经验统计模型。这类方法在单一试验点、特定生育期或相对稳定的生产条件下较易应用，但一旦扩展到多年、多区域场景，模型参数和经验关系往往难以直接迁移。同样，线性模型的限制也较明显，它通常只能表达变量间的平均响应关系，难以反映产量与土壤、气象、地形等因子之间的非线性耦合[4]。但相比之下，作物生长机理模型能够描述作物发育和产量形成过程，但模型运行需要较完整的气象、土壤、品种和田间管理数据，区域尺度应用时容易受到高精度输入数据不足的制约[5]。云南山区的情况更为复杂，坡地、坝区和山地交错分布，小农经营条件下施肥、管理和种植制度差异较大，气候与地形因素又在较短距离内快速变化，使得产量变异来源更加分散。基于这一背景，采用能够整合多源信息并处理非线性关系的建模方法，是提高玉米产量预测精度和模型稳定性的必要途径[6]。

近年来，机器学习方法逐渐被用于作物产量预测和数字土壤制图。相比克里金插值、线性回归等传统方法，机器学习对变量关系的预设较少，更适合处理土壤、气象、遥感和空间位置等多源数据[7]。目前，随机森林（Random Forest, RF）、支持向量机（Support Vector Machine, SVM）和梯度提升树类模型在农业预测研究中应用较多[8]。其中，极端梯度提升（Extreme Gradient Boosting, XGBoost）兼具非线性拟合能力和正则化控制优势，在区域作物估产中表现较好。赵培钦等[9]基于中国西南地区大理、丽江和昭通等地的多点试验数据，对多种机器学习模型进行比较，发现XGBoost对玉米产量的预测效果优于RF和SVM，模型决定系数（Coefficient of Determination, R^2）达到0.98。这些研究表明，机器学习能够在多源数据支持下提升区域尺度作物产量预测能力。但模型效果并不只取决于算法类型。以XGBoost为例，学习率、树深、采样比例等超参数会直接影响模型的拟合程度和泛化能力。如果单纯依靠人工经验调参，不仅耗时，而且容易错过更优的参数组合。因此，粒子群优化算法（Particle Swarm Optimization, PSO）等智能优化方法开始被用于模型参数寻优，通过在设定搜索空间内自动筛选较优参数，提高XGBoost模型的预测精度和稳定性[10, 11]。

在农业决策中，模型不仅要预测准确，还需要解释高产或低产形成的主要原因。复杂集成模型通常具有较强的拟合能力，但内部决策过程不够直观，这在一定程度上限制了其在施肥管理和田块诊断中的应用[12]。因此，解释机器学习方法逐渐被引入农业建模。其中，沙普利加和解释（SHapley Additive exPlanations, SHAP）以博弈论中的Shapley值为基础，可以将模型预测结果分解为各输入特征的贡献值，从而判断不同变量对预测结果的影响方向和贡献大小，并同时提供全局变量重要性和局部样本解释[13-15]。已有研究开始将解释机器学习用于作物产量和土壤属性分析。JONES等[16]在约2 000 hm^{2}灌溉棉田中构建XGBoost产量模型，结果表明R^2为0.83、均方根误差（Root Mean Square Error, RMSE）为0.78 b/hm^2，并基于SHAP量化土壤碱性、盐分、钠化和地形等约束因子对棉花产量的空间贡献，从而识别田块尺度上的主要限产因子。WANG等[17]结合LightGBM、SHAP与数据增强方法估算冬小麦产量，模型R^2为0.63、RMSE为514.18 kg/hm^2、平均绝对误差（Mean Absolute Error, MAE）为8.79%，并识别出拔节期叶面积指数（Leaf Area Index, LAI）和植被温度状态指数（Vegetation Temperature Condition Index, VTCI）、返青期LAI是产量形成的主要解释变量。JOSHI等[18]进一步将SHAP、局部可解释模型无关解释（Local Interpretable Model-Agnostic Explanations, LIME）和积分梯度（Integrated Gradients, IG）用于双向长短期记忆网络（Bidirectional Long Short-Term Memory, Bi-LSTM）冬小麦产量模型解释，模型R^2最高达0.88，解释结果表明生育后期增强型植被指数（Enhanced Vegetation Index, EVI）、温度和降水对产量预测具有关键作用；此外，ABEKOON等[19]在土壤N、P、K预测中发现，最优人工神经网络（Artificial Neural Network, ANN）模型测试集R^2达0.97，SHAP和LIME能够同时揭示变量重要性、作用方向及不同生长阶段的局部差异。总体来看，解释机器学习方法已从单纯的特征重要性排序，逐步发展到对变量贡献方向、交互作用和局部预测来源

的综合刻画，其中SHAP在产量形成因子识别和空间异质性解释方面具有较强优势。

上述研究表明，机器学习方法能够在多源数据支持下提高作物产量预测能力，SHAP等解释方法也可用于识别不同变量对产量形成的贡献方向和作用强度。然而，已有研究多集中于平原农区、单一作物产量估算或模型方法验证，针对云南山地玉米种植区土壤养分差异、空间环境背景与产量形成关系的解释性研究仍相对不足。云南地形破碎、气候垂直分异明显，玉米产量不仅受有机质、氮、磷、钾等土壤养分影响，也受到地形、气候和区域管理条件等空间背景因素共同制约。因此，有必要在产量预测的基础上进一步分析变量贡献，建立兼顾预测精度和结果解释能力的综合方法。本研究利用云南省测土配方施肥土壤养分数据和玉米产量栅格数据，将有机质（Organic Matter, OM）、全氮（Total Nitrogen, TN）、AP和AK，以及空间位置变量作为主要输入因子，构建粒子群优化（Particle Swarm Optimization, PSO）的XGBoost产量预测模型。在此基础上，引入SHAP方法量化不同变量对模型输出的贡献，并比较各因子贡献强度及其空间分布差异。通过上述分析，识别云南不同区域玉米产量差异的主要影响因素，以期为山地玉米分区施肥、耕地质量提升和稳产增产提供参考。

1 材料与方法

1.1 研究区概况

研究区为云南省，位于中国西南部，地理范围为21°8′—29°15′N、97°31′—106°11′E。云南省地形以高原山地为主，平均海拔约2 000 m，整体呈西北高、东南低的地势格局，山地、丘陵和坝区交错分布。受海拔梯度和地形切割影响，省内气候类型复杂，立体气候特征明显，主要表现为亚热带季风气候。多年平均气温约16～22 ℃，年降水量多在800～1 200 mm之间，降水总体由南向北递减，雨热同期，干湿季分明。

1.2 数据获取与处理

1.2.1 玉米产量数据

玉米产量数据来源于中国科学院空天信息创新研究院发布的2000—2022年全球5 km格网化粮食产量数据集。本研究选取其中2005—2020年云南省玉米产量数据作为研究基础，并结合云南省范围及玉米主产区掩膜进行裁剪，剔除研究区外和非玉米主产区像元。为减少不同年份产量栅格之间的空间偏移影响，对各年度数据进行了统一掩膜裁剪、空间配准和像元中心匹配处理，使多年份产量数据在空间位置上保持一致，为后续与土壤养分数据的匹配提供基础。

1.2.2 土壤养分数据

土壤养分数据来源于测土配方施肥土壤基础养分数据集，原始数据为点位样本，包含经纬度坐标及主要养分指标信息。结合研究目标，选取OM、TN、AP和AK这4项指标作为核心变量。上述指标是测土配方施肥工作中常用的耕层土壤养分评价指标，可分别反映土壤肥力基础、氮素供应水平，以及磷钾有效性状况，与玉米产量形成关系密切[20]。

样点空间分布覆盖全省16个地级行政区。为保证数据时空代表性，选取与玉米主产区重叠的样点，并以样点年度均值代表该地区年度土壤养分状况。为与产量影像空间一致，使用反距离加权（Inverse Distance Weighting, IDW）法对土壤点数据进行空间插值，分辨率设定为5 km，以生成2005—2020年连续的土壤养分空间分布栅格。

1.3 空间配准与特征提取

为实现玉米产量与土壤养分的空间匹配，本研究建立了基于经纬度的双向空间配准流程：

1）以统一网格（5 km×5 km）为基础，计算栅格中心点经纬度坐标；

2）提取每个像元在对应年份的OM、TN、AP、AK均值作为输入特征；

3）以产量栅格值作为目标变量Y，形成经纬度-土壤养分-产量的时空样本表。

该方法在栅格级别实现了土壤属性与产量信息的空间一体化融合，为后续机器学习建模提供了高精度的配准输入数据。

1.4 PSO-XGBoost模型构建

本研究采用PSO与XGBoost相结合的建模框架，用于构建玉米产量预测模型。XGBoost以其高效的梯度提升策略和特征自动选择能力，广泛应用于农业预测任务中；PSO算法则用于优化其超参数，提升模型的泛化能力。

XGBoost的基本形式为加性模型，如公式（1）所示：

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), \quad f_k \in F \tag{1}$$

式中：$\hat{y}_i$为第i个样本的预测产量；$\boldsymbol{x}_i$为输入特

征向量（经度、纬度、OM、TN、AP、AK）；f_k为第k棵回归树；F为回归树函数空间。

模型目标函数如公式（2）所示：

$$Obj = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{2}$$

式中：Obj为模型目标函数；n为训练样本数量；K为回归树数量；y_i和$\hat{y}_i$分别为第i个样本的实测产量和预测产量；f_k为第k棵回归树；$l\left(y_i, \hat{y}_i\right)$为平方误差损失函数；$\Omega\left(f_k\right)$为正则化项，用于防止过拟合。

PSO算法通过模拟粒子在搜索空间的群体协同进化来自动寻优模型超参数（如学习率、最大深度、子样本率等）。粒子速度与位置更新如公式（3）和公式（4）所示：

$$v_i^{t+1} = wv_i^t + c_1 r_1\left(p_i - \theta_i^t\right) + c_2 r_2\left(g - \theta_i^t\right) \tag{3}$$

$$\theta_i^{t+1} = \theta_i^t + v_i^{t+1} \tag{4}$$

式中：v_i、θ_i分别为第i个粒子的速度与位置；p_i为个体最优位置；g为全局最优位置；w为惯性权重；c_1、c_2为学习因子；r_1、r_2为［0，1］区间的随机数。本研究中，PSO的参数设置为：粒子数为20，最大迭代次数为15，惯性权重w为0.8，个体学习因子c_1和群体学习因子c_2均为1.5。PSO优化的XGBoost超参数包括最大树深max_depth、学习率learning_rate、样本采样比例subsample、特征采样比例colsample_bytree、最小叶节点样本权重min_child_weight和L2正则化系数reg_lambda，其搜索空间分别设定为3～8、0.01～0.20、0.60～1.00、0.60～1.00、1～10和0.00～10.00。模型固定参数为：目标函数objective=reg：squarederror，弱学习器数量n_estimators=4 000，评价指标eval_metric=rmse，随机种子random_state=42。参数优化过程中，样本按8∶2划分为训练集和测试集，并从训练集中进一步划分20%作为验证集；以验证集上的模型性能作为粒子适应度进行迭代寻优，最终获得最优参数组合为max_depth=8、learning_rate=0.01、subsample=0.824 9、colsample_bytree=1.0、min_child_weight=3和reg_lambda=0.0。最终模型结构如图1所示。

1.5 对比模型设置

为全面评价PSO-XGBoost模型在云南省玉米产量预测中的适用性，选取多种典型机器学习模型作为对比模型，包括RF、Light Gradient Boosting Machine（LightGBM）、CatBoost、XGBoost，以及岭回归（Ridge Regression）、Lasso回归（Lasso Regression）和K近邻回归（K-Nearest Neighbors, KNN）等。

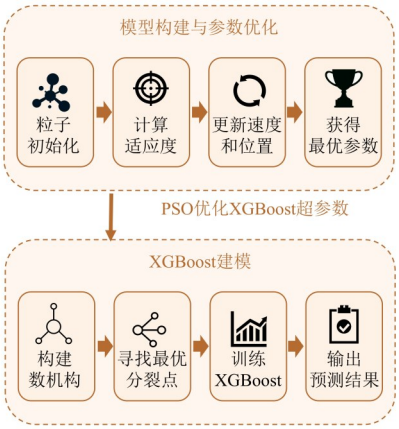


图1　PSO-XGBoost玉米产量预测模型结构图

Fig. 1　Structure of the PSO-XGBoost corn yield prediction model

其中，Ridge回归和Lasso回归属于典型线性回归扩展方法，分别通过L2和L1正则化抑制模型过拟合；RF通过构建多棵决策树并进行集成平均，具有较强的非线性拟合能力和抗过拟合能力；LightGBM是一种基于梯度提升框架的高效集成学习算法，具有训练速度快、内存消耗低的特点；CatBoost则通过有序提升和类别特征处理机制，有助于提高模型的鲁棒性与泛化性能。上述模型能够代表线性模型、单树模型、集成树模型和核方法等不同建模思路，可为PSO-XGBoost模型的性能评估提供多维参照。

1.6 模型训练与精度评价

模型输入特征为经度、纬度、OM、TN、AP、AK，输出变量为玉米产量。为保证模型评估的稳健性与泛化能力，本研究采用两类验证方案。首先，在随机划分条件下将样本划分为训练集和测试集，并在训练集中结合交叉验证进行模型训练与参数选择；其次，为检验模型在空间外推条件下的稳定性，根据样本经纬度采用空间网格分块方法将研究区划分为5个空间子区，每次以其中1个子区为验证集，其余子区为训练集，开展空间分块交叉验证。模型性能通过R^2、RMSE和MAE进行综合评价，如公式（5）～公式（7）所示。

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{5}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \tag{6}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2} \tag{7}$$

式中：n为评价样本数量；y_i为实测产量；$\hat{y}_i$为预测值；$\bar{y}$为样本均值。

1.7 模型解释性分析(SHAP)

为深入解析不同土壤养分要素及空间位置对玉米产量的驱动作用，本研究引入SHAP方法对PSO-XGBoost模型进行解释性分析。SHAP基于博弈论思想，通过计算各特征在不同预测样本中的边际贡献，定量反映每个输入变量对模型输出（产量预测值）的正负影响及重要性。其数学定义如公式（8）所示：

$$\phi_i=\sum_{S\subseteq N\{i\}}\frac{|S|!(|F|-|S|-1)!}{|F|!}\left[f\left(S\cup\{i\}\right)-f(S)\right] \tag{8}$$

式中：ϕ_i为第i个特征的SHAP值；F为全部特征的集合；S为不包含特征i的特征子集；$f(S)$表示仅使用子集S时模型的输出；$N\{i\}$表示全体特征集合中，排除第i个特征后的剩余特征集合；$f\left(S\cup\{i\}\right)$表示在S基础上加入特征i后模型的输出。

在具体实现上，考虑到本模型为XGBoost树模型，解释阶段采用SHAP库中的TreeExplainer进行计算，即基于树模型结构的Tree SHAP算法对模型进行解释。数据集首先按照test_size=0.2、random_state=42进行随机划分，SHAP分析仅针对独立测试集样本开展，以保证解释结果与模型泛化性能评价对象一致。解释过程中未额外构造背景样本集，而是直接基于训练完成的XGBoost模型进行SHAP值计算。最终得到测试集中每个样本在每个特征上的SHAP值矩阵，并同步保存基线值（base value）、测试集特征矩阵及对应SHAP数值，以保证分析过程可追溯和可复核，基于上述结果，本研究从全局解释、特征响应关系和空间异质性三个层面展开分析。

2 结果与分析

2.1 土壤养分数据时空分析

基于云南省2005—2020年测土配方施肥土壤基础养分数据集，对OM、TN、AP和AK进行统计与空间分析。统计结果（表1）显示，2005—2014年与2015—2020年相比，全省土壤OM由32.94 g/kg降至31.50 g/kg，呈轻微下降趋势；TN基本保持稳定（约1.8 g/kg），但区域差异显著。AK和AP含量明显上升，分别由136.1和21.3 mg/kg增至152.3和29.5 mg/kg，说明近年化肥投入增加促进了速效养分积累。总体上，稳定性养分（OM、TN）下降而速效性养分（AK、AP）上升，反映出土壤养分结构由“储备型”向“速效型”转变[21]。

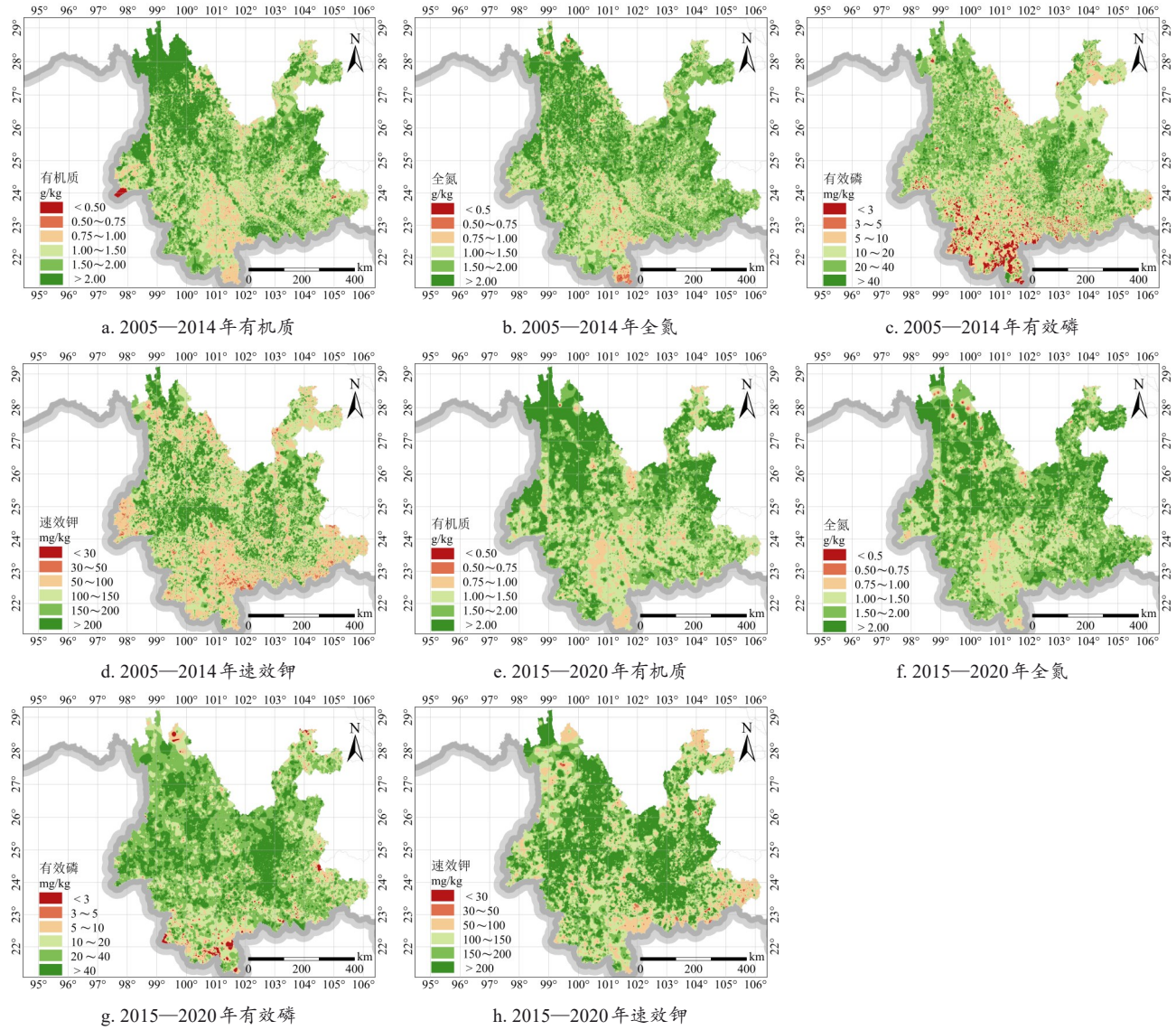
表1 土壤养分统计特征

Table 1 Statistical characteristics of soil nutrients

土壤养分		样本数/个	平均值	标准差	变异系数/%	5%～95%范围
2005—2014	有机质/(g/kg)	242 085	32.94	15.66	47.50	12.52～63.43
	全氮/(g/kg)	112 899	1.84	0.95	51.60	0.75～3.46
	速效钾/(mg/kg)	240 665	136.10	88.50	65.10	37.03～320.01
	有效磷/(mg/kg)	241 856	21.30	20.50	96.30	2.84～63.92
2015—2020	有机质/(g/kg)	55 886	31.50	14.20	45.03	11.43～58.71
	全氮/(g/kg)	32 005	1.67	0.80	47.73	0.24～3.12
	速效钾/(mg/kg)	53 537	152.30	88.70	58.22	43.04～335.03
	有效磷/(mg/kg)	54 519	29.50	26.60	90.08	3.32～87.84

空间分布结果（图2）表明，云南省不同土壤养分指标具有较为明显的区域分异特征，且两个时期之间存在一定演变趋势。总体上，OM和TN的高值区主要分布于滇东北及滇中北部地区，尤其在昭通、曲靖北部和昆明周边较为突出，而滇西南和滇南部分地区相对较低，表现出由北部和高海拔区域向南部和低海拔区域递减的格局。与2005—2014年相比，2015—2020年OM空间格局整体变化不大，但滇西地区及部分中南部地区低值斑块有所扩大；TN则总体呈增加趋势，在滇东北和滇中部分区域表现更为明显。相比之下，AP和AK对农业投入变化响应更为敏感。AP在2015—2020年较前一

时期整体升高，中高值区分布范围明显扩大；AK高值区也进一步扩展，主要集中于滇中和滇东北部分地区。总体来看，云南省耕层土壤养分呈现“OM、TN相对稳定或略降，而AP和AK明显上升”的空间演变特征，表明土壤养分结构由相对稳定的储备型养分向速效养分积累增强方向转变。



a. 2005—2014年有机质　b. 2005—2014年全氮　c. 2005—2014年有效磷

d. 2005—2014年速效钾　e. 2015—2020年有机质　f. 2015—2020年全氮

g. 2015—2020年有效磷　h. 2015—2020年速效钾

注：颜色由红到绿表示养分水平由低到高。该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333号标准地图制作，底图无修改。

图2　云南省2005—2020年主要土壤养分空间分布图

Fig. 2　Spatial distribution of major soil nutrients in Yunnan province (2005—2020)

2.2　模型性能评价

为验证所构建模型的稳定性与泛化性能，本研究选取多种典型的机器学习模型对云南玉米产量进行预测精度比较（表2）。结果显示，不同模型的回归性能指标存在明显差异。优化后的PSO-XGBoost模型在MSE、RMSE和MAE等指标上均表现最优，R^2达到0.91，RMSE为564.00 kg/hm^2，MAE为458.54 kg/hm^2，显著优于岭回归（R^2= 0.72）模型，说明该模型在复杂非线性关系拟合方面具有较强优势。相比之下，基于树模型的Random Forest及XGBoost均表现出较好的稳定性与精度，而集成算法LightGBM与CatBoost在训练速度与泛化能力方面也具有一定优势。总体来看，非线性集成算法优于线性模型，表明玉米产量受土壤与地理环境因素的影响具有显著的非线性特征。

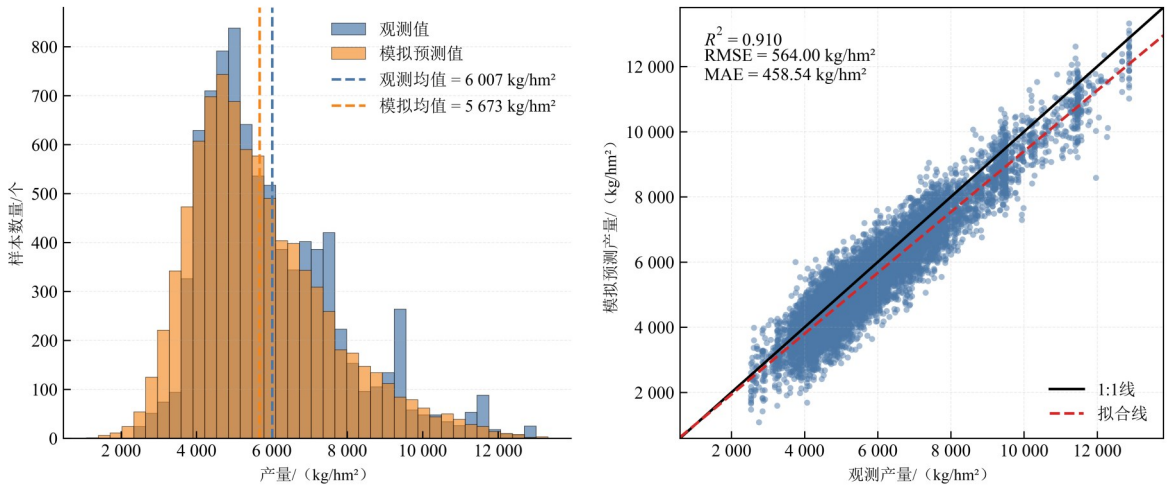
图3展示了PSO-XGBoost模型在测试集上的直方图与散点图结果，可见预测产量与实测产量的分

表2 不同机器学习模型在云南省玉米产量预测中的精度对比

Table 2 Comparison of prediction accuracy among different machine learning models for maize yield prediction in Yunnan province

模型	RMSE/(kg/hm²)	MAE/(kg/hm²)	R^2
Ridge Regression	1 564.87	1 248.30	0.72
Lasso Regression	1 635.42	1 305.76	0.70
Random Forest Regressor	1 036.58	802.41	0.84
Decision Tree Regressor	1 210.54	918.27	0.78
KNN Regressor	1 128.36	864.52	0.75
LightGBM	734.43	587.07	0.88
CatBoost	645.28	502.64	0.90
XGBoost	603.17	475.92	0.89
PSO-XGBoost(本模型)	564.00	458.54	0.91

布高度吻合，两者峰值均集中在4 200～5 400 kg/hm²区间，说明模型对主流产量区间的拟合能力较强。蓝色（实测）与橙色（预测）直方条几乎重叠，表明模型在不同产量层次均具备良好的预测一致性，未出现明显的系统性偏差。散点图中，绝大多数样本点分布在1∶1拟合线附近，回归拟合线（红色虚线）与理想线（黑色实线）几乎重合，验证了模型在空间尺度和样本层面均具有较高精度和稳健性。极少数低产区域样本点出现轻微偏离，可能与土壤贫瘠或局地气候极端事件导致的异常值有关。



a. 观测产量与预测产量频数分布　　b. 观测产量与预测产量拟合关系

图3 PSO-XGBoost模型预测结果的统计分布与拟合关系图

Fig. 3 Statistical distribution and fitting relationship of predicted yield by the PSO-XGBoost model

需要指出的是，由于模型输入变量中包含经纬度信息，而云南地区玉米产量及土壤养分分布具有明显的空间自相关特征，仅采用随机划分可能会在一定程度上高估模型的泛化能力。为此，本研究进一步增加了空间分块交叉验证和经纬度消融实验，以检验模型在更严格条件下的稳健性。结果如表3所示，在随机划分条件下，PSO-XGBoost模型的R^2为0.91；而在空间分块交叉验证条件下，R^2均值下降至0.847，说明随机划分确实会使模型精度略偏乐观，但模型在空间外推条件下仍保持较高的预测能力。进一步地，在去除经纬度变量后，模型R^2降至0.698，表明经纬度在模型中承担了重要的空间背景表征作用。也就是说，经纬度并非单纯的位置变量，而是综合反映了区域气候、地形、种植制度和管理条件等空间异质性信息。与此同时，去除经纬度后模型仍具有一定解释能力，说明土壤养分变量本身对玉米产量形成仍具有实质性贡献，但其作用需要放在空间环境背景中综合理解。

2.3 SHAP特征重要性分析

基于最优PSO-XGBoost模型的SHAP分析揭示

表 3　不同验证方案下 PSO-XGBoost 模型对云南省玉米产量预测的精度评价结果

Table 3　Accuracy evaluation of the PSO-XGBoost model for maize yield prediction in Yunnan province under different validation schemes

验证方案	输入特征	R^2	RMSE/(kg/hm^2)	MAE/(kg/hm^2)
随机划分	土壤养分＋经纬度	0.910	564.00	458.54
空间分块交叉验证	土壤养分＋经纬度	0.847	734.43	587.07
随机划分消融实验	仅土壤养分	0.698	1 032.41	829.31

了各预测因子的相对重要性和正负效应。SHAP 值重要性排序（图 4）中可见，北纬与东经的平均 SHAP 值最高，说明地理空间因素在产量差异中仍具有显著解释力，反映出区域尺度下产量差异存在明显的空间异质性。与此同时，四项土壤养分（OM、TN、AP 和 AK）的重要性总体接近，说明不同养分因子均对模型预测具有重要贡献，产量形成并非由单一养分因素决定。图 5 进一步揭示了各特征对模型输出的正负影响及其非线性关系：OM 与 TN 的高值（红色点）主要分布在正向区域，显示二者提升有助于增产；而 AP 和 AK 的贡献分布较为分散，说明其对模型输出的影响在不同样本间存在较强异质性。值得注意的是，经纬度变量在正负 SHAP 区间均有较广分布，说明空间位置与产量预测之间存在明显非线性关系，难以用简单单调关系概括，而是受到多环境因子的共同制约。总体而言，该结果揭示了云南省玉米产量形成的多因子驱动机制：空间环境提供背景约束，而土壤养分水平决定了区域内的可提升潜力，为后续的分区施肥与精准管理提供了科学依据。

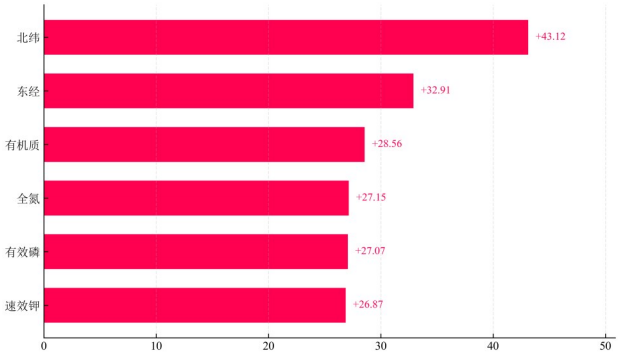


图 4　PSO-XGBoost 模型各输入特征的平均绝对 SHAP 值

Fig. 4　Input features by mean absolute SHAP values in the PSO-XGBoost model

图 6 展示了基于 SHAP 分析得到的主要驱动因子空间分布。整体来看，经纬度相关变量在样本点中占比较高。其中，北纬主导点分布范围较广，主要出现在滇中、滇西南及部分中部区域；东经主导点则在滇东、滇东南及部分边缘区域较为明显。这

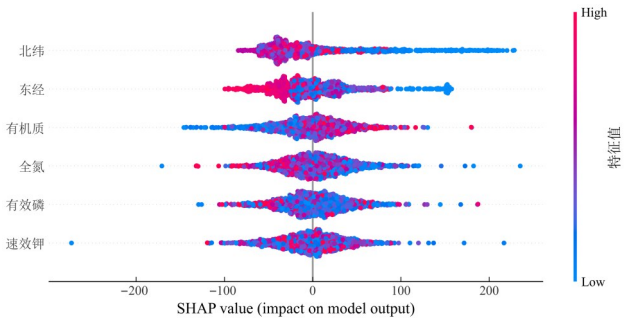


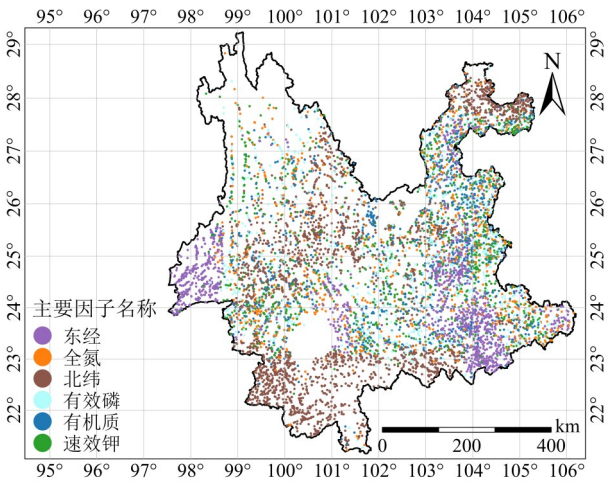
图 5　各输入特征 SHAP 值的蜂群分布及其对玉米产量预测的正负效应

Fig. 5　SHAP beeswarm distribution of input features and their positive and negative effects on maize yield prediction

表明，云南玉米产量预测对空间背景信息较为敏感。经纬度在模型中并不是直接农学因子，更可能综合反映气候带、地形起伏、海拔梯度和区域管理差异等环境背景。土壤养分因子也表现出一定的区域集中特征。OM 和 TN 主导点分布相对分散，在滇东北、滇中及部分高产区有所体现，说明土壤基础肥力和氮素供应仍是这些区域产量形成的重要条件。AP 和 AK 主导点更多出现在滇南、滇西南及部分低肥力区域，提示这些地区磷钾有效性可能对产量提升形成限制。总体来看，云南山地玉米产量的主导因素具有明显区域差异，施肥管理不宜采用统一方案。后续应结合不同区域的主要限制因子，分别加强有机质培育、氮素供应调控和磷钾养分平衡管理，为分区施肥和重点改良区域识别提供依据。

3　讨　论

本研究基于云南省 2005—2020 年测土配方施肥土壤基础养分数据和玉米产量数据，构建了融合土壤养分、空间位置和机器学习方法的玉米产量预测模型，并利用 SHAP 方法分析主要影响因子的贡献特征。结果表明，研究期内云南省耕层土壤养分发生了较明显变化：OM 和 TN 含量总体下降，AP 和 AK 含量则有所升高。这一变化可能与长期耕作强度增加、秸秆还田不足，以及化肥投入偏重速效养分有关，也与国内部分地区耕地养分演变规律基



注:该图基于自然资源部标准地图服务网站下载的审图号为GS(2019)3333标准地图制作,底图无修改。

图6 云南省玉米产量主导驱动因子空间分布图(以绝对SHAP值最大特征表征)

Fig. 6 Spatial distribution of dominant drivers of maize yield in Yunnan province, represented by the feature with the highest absolute SHAP value

本一致[22,23]。OM减少会削弱土壤结构稳定性、缓冲能力和保肥能力,并可能影响氮素矿化过程,进而对玉米根系发育和水分利用产生不利影响。与之相比,AK和AP的积累更多反映出部分区域施肥结构不够均衡,局部地区可能存在养分投入偏高的问题。

从空间格局看,云南省土壤养分分布存在较明显的区域差异。滇东北及滇中高海拔地区OM、TN含量相对较高,滇西南和滇南平坝区则整体偏低;AP和AK高值区主要分布在滇中及滇南部分红壤、紫色土区域。造成这种差异的原因并不单一,既包括气候和地形条件,也与种植制度、复种强度及施肥方式有关。例如,高海拔地区气温较低,土壤有机质分解较慢,有利于有机质积累[24];部分低海拔农区复种频率高、化肥投入集中,则更容易形成速效养分富集[25]。这说明云南土壤养分管理不宜采用统一模式,而应根据不同区域的肥力基础和生产条件开展分区调控。

SHAP特征重要性分析进一步显示,经纬度等空间位置变量在玉米产量预测中贡献较高,说明在云南省复杂山地环境下,气候、地形和区域管理背景仍是影响产量差异的重要因素。需要注意的是,经纬度本身不是直接农学因子,更适合理解为海拔梯度、热量条件、降水差异、种植制度和管理水平等空间背景信息的综合代理。除空间位置外,OM、TN、AP和AK也表现出较明显贡献,表明土壤养分仍是解释产量差异的重要变量。其中,OM与产量总体呈正向关系,说明改善土壤结构、增强保水保肥能力仍是稳产的重要基础;AP和AK在部分区域贡献较高,但在高含量区域可能出现边际效应减弱,提示局部地区应关注过量施肥和养分利用效率问题[26]。

与北方平原农区相比,云南山地农业系统的主导因子更分散,土壤养分、地形梯度、微气候和土地利用方式之间的耦合关系也更复杂。从农业管理角度看,土壤养分变化和模型解释结果共同指向两个问题:一是OM和氮素含量下降区域,应加强有机肥还田、秸秆资源化利用和绿肥轮作,以维持土壤碳氮平衡并改善基础肥力;二是AP和AK富集区域,应优化施肥结构,避免单纯增加速效养分投入带来的养分盈余和面源污染风险。结合产量预测模型和SHAP解释结果,可进一步识别不同区域的主要养分限制因素,为精准施肥和分区管理提供量化依据[27]。

从模型表现看,PSO-XGBoost在本研究中具有较好的适用性。表2结果显示,该模型在多种对比模型中取得最优综合预测效果,R^2达到0.91,高于线性回归、岭回归、Lasso、RF、SVR、KNN、LightGBM、CatBoost和原始XGBoost等模型。与传统线性模型相比,PSO-XGBoost更适合处理云南山区玉米产量形成中的非线性关系和多因子耦合特征;与RF和单一决策树模型相比,XGBoost的boosting机制能够通过误差迭代修正提升复杂关系的拟合能力。PSO的引入则减少了人工调参的不确定性,提高了参数搜索效率,也进一步改善了模型性能。不过,该模型仍存在一定局限。其预测效果对输入数据质量和样本空间代表性较为敏感,训练过程也相对复杂;同时,模型主要依赖数据驱动关系,对空间自相关和作物生长过程的显式表达仍然不足。因此,PSO-XGBoost更适用于省域尺度玉米产量预测、主导因子识别和区域施肥决策支持。若进一步开展田块尺度精细预测或作物生长过程模拟,还需引入更高分辨率数据、地形气象变量和作物生长过程信息。

需要指出的是,本研究采用的玉米产量数据空间分辨率为5 km,在云南复杂山地条件下仍存在一定局限。云南坡地多、地形起伏大,坡位、坡向和田块管理方式在较短距离内即可发生明显变化,5 km栅格难以完整反映这类小尺度差异,也可能带来一定混合像元误差,从而影响局部高产或低产斑

块的识别。因此，本研究结果更适用于省域尺度下玉米产量空间格局和主导因子的宏观分析，对田块尺度精细差异的刻画仍显不足。后续研究可结合更高分辨率产量数据、地形因子、气象过程和农户管理信息，进一步开展分区建模和尺度校正，以提高复杂山地环境下模型预测的精细化程度和实际应用价值。

4　结　论

基于云南省2005—2020年耕层土壤养分监测数据，结合空间统计分析与机器学习方法，分析了土壤养分变化特征及其对玉米产量的驱动作用，主要结论如下：

1）2005—2020年，云南省土壤养分总体呈现“OM、TN略有下降，AP和AK上升”的变化趋势。其中，OM由32.94 g/kg降至31.50 g/kg，TN略有下降，而AK和AP分别升至152.3和29.5 mg/kg。空间上，滇东北及滇中高原区养分水平较高，滇西南及滇南地区相对较低，表现出明显的区域分异特征。

2）PSO-XGBoost模型在云南省玉米产量预测中表现较好。随机划分条件下，模型R^2达到0.91，RMSE为564.00 kg/hm^2，MAE为458.54kg/hm^2，整体优于线性模型和其他机器学习模型，表明该方法能够较好地刻画山区复杂环境下土壤养分与产量之间的非线性关系。

3）SHAP解释结果表明，有OM、TN、AP和AK是驱动玉米产量变化的核心变量，其中OM和TN对高产区贡献更大，而AP和AK在低肥力区作用更为显著。空间分布上，滇东北地区主要受氮素和有机质主导，滇西南地区则更多受AP和AK影响，反映出云南玉米产量形成机制具有明显的空间异质性。

本研究表明，PSO-XGBoost结合SHAP方法能够有效揭示云南省玉米产量形成的主要养分驱动因素，可为区域尺度分区施肥和耕地质量提升提供参考依据。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] CHEN F F, QIU H G, ZHAO Y L, et al. Impact of new maize variety adoption on yield and fertilizer input in China: Implications for sustainable food and agriculture[J]. Agricultural Systems, 2024, 218: 104004.

[2] SU B C, LIU R, LU Z Z, et al. Mapping key soil properties of cropland in a mountainous region of southwestern China[J]. Agronomy, 2024, 14(7): 1417.

[3] 毛昭庆, 张晓娇, 袁媛, 等. 云南省玉米种植区域比较优势分析[J]. 农学学报, 2025, 15(7): 84-92.
MAO Z Q, ZHANG X J, YUAN Y, et al. Comparative advantage analysis of corn planting regions in Yunnan Province[J]. Journal of Agriculture, 2025, 15(7): 84-92.

[4] 咸阳, 宋江辉, 王金刚, 等. 基于环境变量筛选与机器学习的土壤养分含量空间插值研究[J]. 农业机械学报, 2024, 55(10): 379-391.
XIAN Y, SONG J H, WANG J G, et al. Spatial interpolation of soil nutrients content based on environmental variables screening and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10): 379-391.

[5] MERONI M, WALDNER F, SEGUINI L, et al. Yield forecasting with machine learning and small data: What gains for grains [J]. Agricultural and Forest Meteorology, 2021, 308/309: 108555.

[6] CHLINGARYAN A, SUKKARIEH S, WHELAN B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review[J]. Computers and Electronics in Agriculture, 2018, 151: 61-69.

[7] HE Y S, LI M H, ZHANG W, et al. Environmental impacts of crop production systems in subtropical plateau regions: Case study of Yunnan, China[J]. Scientific Reports, 2024, 14: 30254.

[8] HUANG H, LIU Y L, LIU Y F, et al. The spatial distribution and driving mechanism of soil organic matter in hilly basin areas based on genetic algorithm variable combination optimization and shapley additive explanations interpretation[J]. Remote Sensing, 2025, 17(7): 1186.

[9] 赵培钦, 刘长斌, 郑婕, 等. 面向干旱条件下的冬小麦估产HLM模型改进研究[J]. 智慧农业(中英文), 2025, 7(2): 106-116.
ZHAO P Q, LIU C B, ZHENG J, et al. Improvement of HLM modeling for winter wheat yield estimation under drought conditions[J]. Smart Agriculture, 2025, 7(2): 106-116.

[10] FENG P Y, WANG B, LIU D L, et al. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique[J]. Agricultural and Forest Meteorology, 2020, 285/286: 107922.

[11] MA C Y, YE Z L, ZI Q Y, et al. Machine-learning-based multisite corn yield prediction integrating agronomic and meteorological data[J]. Agronomy, 2025, 15(8): 1978.

[12] LISCHEID G, WEBBER H, SOMMER M, et al. Machine learning in crop yield modelling: A powerful tool, but no surrogate for science[J]. Agricultural and Forest Meteorology, 2022, 312: 108698.

[13] ZHONG L, GUO X, DING M, et al. SHAP values accurately explain the difference in modeling accuracy of convolution neural network between soil full-spectrum and feature-spectrum[J]. Computers and Electronics in Agriculture, 2024, 217: 108627.

[14] MOHAN R N V J, RAYANOOTHALA P S, SREE R P. Next-gen agriculture: Integrating AI and XAI for precision crop yield predictions[J]. Frontiers in Plant Science, 2025, 15: 1451607.

[15] SHAMS M Y, GAMEL S A, TALAAT F M. Enhancing crop recommendation systems with explainable artificial intelligence: A study on agricultural decision-making[J]. Neural Computing and Applications, 2024, 36(11): 5695-5714.

[16] JONES E J, BISHOP T F A, MALONE B P, et al. Identifying causes of crop yield variability with interpretive machine learning [J]. Computers and Electronics in Agriculture, 2022, 192: 106632.

[17] WANG Y, WANG P X, TANSEY K, et al. An interpretable approach combining Shapley additive explanations and LightGBM based on data augmentation for improving wheat yield estimates [J]. Computers and Electronics in Agriculture, 2025, 229: 109758.

[18] JOSHI A, PRADHAN B, CHAKRABORTY S, et al. An explainable Bi-LSTM model for winter wheat yield prediction[J]. Frontiers in Plant Science, 2025, 15: 1491493.

[19] ABEKOON T, SAJINDRA H, RATHNAYAKE N, et al. A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P, and K nutrient content in cabbage cultivation

[J]. Smart Agricultural Technology, 2025, 11: 100879.
[20] WANG N, AI Z P, ZHANG Q Y, et al. Impacts of nitrogen (N), phosphorus (P), and potassium (K) fertilizers on maize yields, nutrient use efficiency, and soil nutrient balance: Insights from a long-term diverse NPK omission experiment in the North China Plain[J]. Field Crops Research, 2024, 318: 109616.
[21] 农业农村部办公厅关于做好2022年退化耕地治理工作的通知[J]. 中华人民共和国农业农村部公报, 2022(7): 85-88.
Circular of the General Office of the Ministry of Agriculture and Rural Affairs on improving degraded farmland management in 2022[J]. Gazette of the Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2022(7): 85-88.
[22] LI X Y, SHANG B B, WANG D Y, et al. Mapping soil organic carbon and total nitrogen in croplands of the Corn Belt of Northeast China based on geographically weighted regression Kriging model[J]. Computers & Geosciences, 2020, 135: 104392.
[23] WANG J, ZHU B, ZHANG J B, et al. Mechanisms of soil N dynamics following long-term application of organic fertilizers to subtropical rain-fed purple soil in China[J]. Soil Biology and Biochemistry, 2015, 91: 222-231.
[24] SONG D P, LI H, LIU S J, et al. A geostatistic investigation of the comprehensive evaluation of fertility and spatial heterogeneity of forest soil nutrients in hilly and mountainous regions of Southern China[J]. Arabian Journal of Geosciences, 2019, 12 (9): 292.
[25] WANG T, KANG F F, CHENG X Q, et al. Spatial variability of organic carbon and total nitrogen in the soils of a subalpine forested catchment at Mt. Taiyue, China[J]. Catena, 2017, 155: 41-52.
[26] YANG K, MA Y N, TANG Q. Visualisation of key thresholds of crop production potential and their future distribution patterns in China under climate change scenarios[J]. Journal of Geovisualization and Spatial Analysis, 2025, 9(2): 33.
[27] 农业部关于印发«耕地质量保护与提升行动方案»的通知[J]. 中华人民共和国农业农村部公报, 2015(11): 13-18.

Machine Learning Reveals the Driving Effects of Soil Nutrients on Maize Yield Formation: A Case Study of Yunnan, China

SUN Jiaze[1,2], QU Mingshan[3], YANG Jizhong[4], LOU Chuixin[4], LI Guangwei[2], ZHANG Zhonglili[5*]

(1. *College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*; 2. *Information Technology Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*; 3. *Beijing Agricultural Technology Extension Station*, *Beijing* 100029, *China*; 4. *Yunnan Supply and Marketing Industrial Investment Co.* , *Ltd.* , *Kunming* 650000, *China*; 5. *Intelligent Equipment Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: [Objective] Maize production in Yunnan-representative mountainous regions is strongly shaped by rugged terrain, fragmented cropland, and steep vertical climatic gradients, which drive pronounced spatial heterogeneity in soil nutrients and yield. Unlike homogeneous plains, yield variation here stems from the combined effects of soil fertility, climate, and topography rather than any single nutrient factor, making traditional statistics or isolated soil indicators inadequate. To this end, an interpretable machine learning framework integrating soil nutrients and spatial information was developed to enhance maize yield prediction and identify the dominant drivers of yield disparities across sub-regions of such landscapes. **[Methods]** Soil testing and formula fertilization data from Yunnan province during 2005—2020 were combined with 5 km gridded maize yield data. Four soil nutrient indicators were selected, including soil organic matter, total nitrogen, available phosphorus, and available potassium. Longitude and latitude were also used as input variables. In this study, these two variables were not treated simply as location labels. They were included because spatial position can partly reflect regional environmental gradients that are difficult to describe using soil nutrient data alone, such as differences in elevation, heat and moisture conditions, terrain fragmentation, and cropping background. Based on these variables, particle swarm optimization was used to tune the main hyperparameters of the XGBoost model, forming a PSO-XGBoost yield prediction model. The optimized model was compared with LightGBM, CatBoost, random forest, support vector regression, and the original XGBoost model. Considering that nearby samples may share similar environmental conditions, random partitioning alone may overestimate model performance. Therefore, spatial block cross-validation was further used to examine the model's ability to predict yield under spatial separation. In addition, longitude and latitude were removed in an ablation experiment to test the contribution of spatial background information. Finally, SHAP was used to interpret the trained model, quantify the contribution of each variable, analyze the response of yield prediction to soil nutrients and spatial factors, and identify the dominant limiting factors in different maize-growing areas. **[Results and Discussions]** PSO-XGBoost showed the best performance among the tested models. Under random data partitioning, the model achieved an R^2 of 0.91, indicating that it could effectively capture the relationship among soil nutrients, spatial background, and

maize yield. Compared with LightGBM, CatBoost, and the unoptimized XGBoost model, PSO-XGBoost gave more accurate and stable predictions. This result suggested that particle swarm optimization improved the parameter configuration of XGBoost and made it more suitable for yield prediction in complex mountainous regions. Under spatial block cross-validation, the R^2 decreased to 0.847. Although this value was lower than that obtained from random partitioning, it still showed good predictive performance under a stricter validation strategy. This also indicated that the model did not rely only on the similarity between neighboring samples, but retained a certain ability to predict yield in spatially separated areas. The ablation experiment further confirmed the importance of spatial information. After longitude and latitude were removed, the model R^2 dropped sharply to 0.698. This decline showed that maize yield patterns in Yunnan cannot be well explained by soil nutrient indicators alone. Spatial variables probably contain information related to elevation-induced climatic differences, regional hydrothermal conditions, terrain fragmentation, and differences in the cropping environment. The SHAP results led to a similar conclusion. Longitude and latitude had relatively large contributions, suggesting that regional environmental gradients played an important role in yield formation. Among the soil nutrient variables, soil organic matter, total nitrogen, available phosphorus, and available potassium were all related to maize yield, but their roles differed across Yunnan. In some areas, yield was more closely associated with organic matter and nitrogen supply, while in others phosphorus or potassium appeared to be more limiting. This spatial difference suggested that the main constraints on maize production were not the same across the province. For this reason, a province-wide fertilization scheme would be difficult to apply effectively. Nutrient management in Yunnan should instead be adjusted according to local soil conditions and the environmental background of different mountainous regions. **[Conclusions]** Maize yield in Yunnan province is affected by both soil nutrient conditions and broader spatial environmental factors. Soil fertility is still an important basis for yield formation in mountainous farmland, but it does not explain the spatial yield differences on its own. By combining PSO-XGBoost with SHAP, this study provides a way to improve yield prediction while also tracing the main factors related to regional yield variation. The framework can be used to identify areas where nutrient constraints are more likely to occur, as well as areas where terrain, climate background, or other regional environmental factors may have a stronger influence. These results can support more targeted fertilization, cultivated land quality improvement, and sustainable maize production in the mountainous agricultural areas of Yunnan.

Keywords: maize yield prediction; PSO-XGBoost; SHAP; soil nutrients; spatial heterogeneity; machine learning

Foundation items: Yunnan Provincial Major Science and Technology Special Project (202202AE090013); China Agriculture Research System of Maize (CARS-02); Training Program for Excellent Young Scientists of Beijing Academy of Agriculture and Forestry Sciences (YXQN202304-C)

First author: SUN Jiaze, E-mail: sunjiaze@cau.edu.cn
***Corresponding author:** ZHANG Zhonglili, E-mail: zhangzll@nercita.org.cn