吕林有,李佳慧,赵艳,等.融合随机森林-递归特征消除的冬黑麦产量预测模型的建立[J].华中农业大学学报,2026,45(3):56-67.
DOI:10.13300/j.cnki.hnlkxb.2026.03.005

融合随机森林-递归特征消除的冬黑麦产量预测模型的建立

吕林有[1,2],李佳慧[1],赵艳[2],马艳[1],何梓源[1]

1.辽宁工程技术大学环境科学与工程学院,阜新 123000; 2.辽宁省沙地治理与利用研究所,阜新 123000

摘要 针对传统模型难以解析行距-播种量互作非线性效应的问题,通过双因素田间梯度试验,系统测定冬黑麦关键生育期生物学性状与产量构成要素指标,解析关键生育期群体动态与产量构成的互作效应,构建融合随机森林-递归特征消除的冬黑麦产量预测模型。田间试验结果显示,40 cm行距×600 g/40 m²播种量处理组合能够优化冠层透光率、平衡"源-库"关系及提升抗逆性,实现冬黑麦高产稳产的协同调控;在冬黑麦完全成熟期,系统采集不同处理组的植株样本,并采用标准化试验流程对关键农艺性状,包括穗长、穗质量、穗节数、株高、根长、有效分蘖数、叶片数以及单株种子数进行量化表征,共获得540组有效数据,以单株种子数为预测目标,构建了包含播种行距、播种量、穗长、穗质量、穗节数、株高、根长、有效分蘖数、叶片数等原始变量及6个衍生特征的多维特征集,构建递归特征消除的随机森林(random forest-recursive feature elimination,RF-RFE)模型,模型表现出优异的预测性能(R^2=0.951,RMSE=2.28)和泛化能力(OOB-RMSE=4.33)。通过产量预测模型反演获得行距-播种量三维响应曲面,确定了高产参数为行距35～49 cm、播种量535～680 g/40 m²;边际效应揭示行距>45 cm时每增加1 cm减产1.23 kg,播种量>650 g/40 m²时边际收益递减。因此,建议生产实践中将行距控制在40～45 cm,播种量控制在600～650 g/40 m²,既可保证高产又留出操作容错空间。

关键词 冬黑麦;随机森林;递归梯度融合;产量预测;边际效应;三维响应曲面

中图分类号 S512.5 **文献标识码** A **文章编号** 1000-2421(2026)03-0056-12

冬黑麦作为全球重要的禾谷类作物,在农业生产和生态系统中具有不可替代的战略地位[1-2]。因其极强的耐寒性、耐旱性和土壤适应性,在全球高纬度及温带干旱地区广泛种植[3-5]。而冬黑麦产量的精准预测不仅能提高农业生产效率,还可以为保障区域粮食安全、促进农业可持续发展注入新动力。传统预测模型多依赖于经验统计模型与作物生长模型[6-8]。经验统计模型存在无法精准捕获冬黑麦形态指标与农艺操作间的非线性关系并且极易忽略变量间的多重共线关系[9-10];而作物生长模型通过数学模拟光合作用、养分分配等生理过程进行产量预测,预测过程需要校准大量品种参数,获取成本极高,难以整合易变的农业措施,无法动态操控冬黑麦种植农艺[11-12]。而植株形态性状与农艺操作参数作为产量形成的直接表型,尚未在预测框架中被系统整合,这些指标具有易观测、低成本、强生物学解释性的优势,能够直接反映植物生长状态与环境的互作效应,它们与最终产量的映射关系呈现高维度、非线性及多重共线关系,亟需开发与之适配的智能预测算法。

机器学习在数据驱动的学习机制下持续优化模型性能,其非线性特征特别适合于处理农业系统中的复杂关系[13]。当前已有众多学者将智能算法应用于农业系统[14-15]。例如,李涛宇等[16]利用机器学习随机森林和XGBoost模型,对冬小麦农田氮淋失进行定量分析;Johnson等[17]采用多元线性回归(multiple linear regression,MLR)以及贝叶斯神经网络(Bayesian neural network,BNN)和基于模型的递归分割(recursive partitioning,RP)预测种植区大麦、油菜等作物产量;刘峻明等[18]基于随机森林算法,依据统计数据提取不同穗分化期温湿度、降水及小麦种植区经纬度、高程等空间因素,对冬小麦产量进行了回归预测;Huber等[19]提出了一种通道方法,将遥感

收稿日期:2025-11-23
基金项目:国家重点研发计划项目(2024YFD1501405);辽宁省科技攻关专项(2023JH1/10400001)
吕林有,E-mail:llyou515@126.com

图像处理为基于特征的表征形式，使用梯度提升决策树XGBoot对美国大豆产量进行预测，与基于深度学习模型相比，该方法获得了较高的预测精度；Nyéki等[20]基于时空数据，对比5种智能预测算法，对玉米产量进行对比测试，同时估算了特定地点土壤肥力的空间分布对玉米籽粒产量的影响；Cai等[21]整合不同来源的数据，预测了澳大利亚多年小麦产量，同时采用LASSO作为baseline，利用3种主流的机器学习方法（支持向量机、随机森林和神经网络）建立了多种经验模型进行收益率预测。

在精准农业与作物生长模型研究领域[22]，随机森林算法因其独特的集成学习机制和优异的特征处理能力，已成为农作物产量预测的重要建模工具[23-25]。该算法通过构建多棵决策树的集成模型，能够同时处理连续型和类别型变量，其内置的袋外（out-of-bag，OOB）误差估计机制，可有效规避过拟合风险，显著提升模型的泛化能力[26]。王俊等[27]基于2023年的Sentinel和MODIS NDVI数据及县级产量信息，应用随机森林模型从县级扩展至像素级，以江苏射阳无人农场为例对水稻产量进行预测，并探讨不同特征对模型学习机制的影响。徐浩等[28]基于WheatGrow模型输入输出数据建立生长季变量的随机森林模型（RF-GS）与月份变量的随机森林模型（RF-Mon），并利用均方根误差（RMSE）评价了模型的性能。

然而，现有研究在禾谷类作物，特别是冬黑麦的栽培调控方面存在以下关键问题亟待解决：其一是缺乏对行距和播种量互作非线性效应的系统解析与量化评估；二是尚未构建完整的农艺性状等多指标与最终产量响应的多维智能预测模型；三是传统方法难以处理高维度、非线性且具有多重共线性的农艺数据，无法为精准栽培提供量化决策依据。

本研究通过行距梯度与播种量梯度双因素田间试验，系统解析冬黑麦关键生育期群体生长动态与农艺性状的互作关系，并基于田间试验数据，设计构建融合RF-RFE冬黑麦产量预测模型，结合特征工程，分析冬黑麦产量预测过程中特征量重要分数并评估模型参数；通过训练的RF-RFE（random forest-recursive feature elimination）模型进行三维响应曲面反演，系统分析行距与播种量互作效应对产量形成的调控机制，结合边际效应量化行距和播种量并进行优化，旨在为冬黑麦精准栽培提供可靠的决策工具，同时为粮饲作物精准栽培提供理论支持与技术指导。

1　材料与方法

1.1　试验地概况

试验地位于辽宁省阜新市彰武县冬黑麦试验基地（42° 37′30″ ～42° 50′00″N，122° 11′15″ ～122°30′00″ E），年平均气温7.2 ℃，最高温度37.4 ℃，最低温度－30.4 ℃，平均风速3.8 m/s，最大风速38 m/s，年均降水量510.3 mm，平均相对湿度61%，平均无霜期156 d。植株处理前按条件统一管理，统一选择沙地土壤进行培养。

1.2　试验设计

本研究以冬黑麦为试验材料，设置30、40、50 cm行距及480、600、720 g/40 m^2播种量梯度处理，采用完全随机区组设计，于2024年9月14日播种，共设置30个小区，每小区面积40 m^2（8 m × 5 m），设置3种不同播种处理，分别按照行距30、40、50 cm进行播种；行距30 cm小区16垄，40 cm行距小区12垄，50 cm行距小区10垄。试验设置480、600、720 g/40 m^2 3个播种量水平，每个水平重复3次，并分别应用于不同行距处理，以降低环境变异对试验结果的影响。

播种后于2024年10月23日作物苗期，2025年3月20、21日作物返青期，2025年4月25日作物拨节期，2025年5月23日作物抽穗期，2025年6月12日作物灌浆期和2025年6月24日作物成熟前进行统一灌溉。播种时统一施有机肥，确保土壤基础肥力均匀。2025年4月11日、2025年5月13日追尿素肥，施用量15～22.5 g/m^2。通过控制行距与播种量的交互效应，结合标准化水肥管理，探究不同种植模式对作物生长及产量的影响。

1.3　农艺性状测量

测定不同生育阶段的农艺性状以阐明群体结构对生长发育及产量形成的影响。于2024年11月8日苗期，从东到西，每小区取30 cm段测定样方内总株数、单株分蘖数、根长、叶片数及器官干物质积累，以评估早期群体建成与个体发育特征；于2025年3月23日越冬期统计活苗率，解析行距与播种量对植株抗寒性的调控效应；于2025年4月10日至6月10日冬黑麦返青至灌浆期，每周监测所有冬黑麦植株株高并计算按照不同行距和不同播种的冬黑麦株高均值，以揭示群体冠层扩展规律；于2025年6月25日、7月4日和7月9日冬黑麦成熟期测定每小区每10穗的穗长、穗节数、穗质量平均值，并测量每20穗种子

总质量、百粒种子质量以及种子数量等产量构成要素；同时结合每5 d间隔的穗部成熟度监测，明确不同处理对库源关系及收获适期的影响。

1.4 光照间接评估方法

本研究基于作物冠层几何学与生物分配理论，采用几何透光指数（geometric light transmittance index，GTI）、冠层密度指数（canopy density index，CDI）与避荫响应指数（shade avoidance response index，SRI），分别从冠层空间几何、生物量密度和植株形态可塑性3个互补维度，对田间光环境及其生物学效应进行间接但系统的量化评估。行距决定植株间的水平间隔，而株高反映冠层垂直高度，二者之比值为GTI，用于表征冠层基部接受直射光的潜在可能。CDI用于评估单位土地面积上的冠层生物量密度，作为光竞争强度的指标。根据Monteith光截获模型[29]，群体光合产量与冠层所截获的光合有效辐射（PAR）成正比，而叶面积指数（LAI）是决定光截获量的核心结构参数。由于在本研究条件下，直接、无损地测定各小区LAI成本较高，而在品种一致、环境可控的试验中，单株叶片数与叶面积、叶干质量存在显著的正相关关系[30]。因此，本研究采用单位面积上的平均单株叶片数作为CDI的值，用以反映不同处理间冠层密度和光竞争强度的相对差异。当植株感知到来自相邻植株的避荫信号时，会启动一系列适应性形态调整，典型的表现为促进节间伸长，以株高增加的方式竞争光照，同时抑制侧生器官生长，主要表现为分蘖数减少。因此，分蘖数与株高的比值（SRI）能够综合反映植株在特定群体结构下所受的光竞争胁迫程度及其形态可塑性响应策略。

1.5 基于随机森林回归与递归特征消除（RFE）构建产量预测模型

本研究聚焦冬黑麦栽培参数优化，通过分析不同播种行距与播种量对植株形态指标及产量构成要素的影响，构建基于随机森林回归与递归特征消除（RFE）的产量预测模型。RFE技术通过迭代评估特征重要性并剔除冗余变量，筛选出关键产量驱动因子，提升模型对高维农艺性状数据的解析能力，尤其适用于冬黑麦生长中复杂的非线性关系。

随机森林模型作为一种集成学习算法，通过多棵决策树提高模型预测的准确性和稳健性。决策树将输入空间划分为多个区域，并在这些区域内为每个样本分配预测值；其中输入特征$X\in R^n$，输出目标变量为$y\in R$，对于每个样本i，模型通过树状结构选择最佳的特征x_j及其分隔值t_j，从而将数据分为两部分：

$$R_1(j,t_j)=\{X|X_j\leqslant t_j\},R_2(j,t_j)=\{X|X_j>t_j\} \quad (1)$$

每棵树中预测值的输出为每个叶片点区域内的平均值：

$$\hat{y}(X)=\frac{1}{|R_k|}\sum_{i\in R_k}y_i \quad (2)$$

随机森林通过生成多个随机化的决策模型，并将每棵树的预测结果取平均值，从而提高模型的稳健性，假设模型中设置T棵树，每棵树的预测值为$\hat{y}(X)$，则随机森林模型的最终输出为：

$$\hat{y}(X)=\frac{1}{T}\sum_{t=1}^{T}\hat{y}_t(X) \quad (3)$$

特征重要性用于量化随机森林每个输入特征对预测准确性贡献性大小的核心指标：

$$I_j=\sum_{t=1}^{T}\sum_{s\in S(j,t)}\Delta I(s,t) \quad (4)$$

式(4)中，I_j表示输入特征x_j；$S(j,t)$表示在树t中与特征x_j相关的分裂；$\Delta I(s,t)$表示通过特征x_j进行分裂所带来的不纯度减少。

递归特征消除的目的是通过反复评估特征重要性来逐步移除不重要的特征，最终保留最具有影响力的特征。

对于当前的特征子集S_k，使用交叉验证训练随机森林回归模型，并通过计算均方根误差（RMSE）来评估模型性能：

$$(S_k)_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}(S_k))^2} \quad (5)$$

其中S_k是第k轮递归消除中的特征子集；y_i是真实值，$\hat{y}(S_k)$是特征S_k预测值。

在每一轮中，通过模型的特征重要性值I_j来移除贡献最小的特征。随着递归特征消除的进行，不断减少特征数量，并在每轮训练后计算模型的RMSE。最终通过比较所有特征子集的RMSE值，选择最小的特征作为最终的特征集：

$$S^*=\arg\min_{S_k}(S_k)_{\mathrm{RMSE}} \quad (6)$$

1.6 行距-播种量响应曲面反演与边际效应分析

通过三维响应曲面反演技术，系统分析行距与播种量互作效应对产量形成的调控机制。采用行距分辨率0.5 cm、播种量分辨率5 g精细化网格搜索策略，根据田间试验数据，将其他农艺参数固定为最优值、仅改变行距和播种量以量化评估栽培参数组合

的边际效应。

三维响应曲面是超维预测空间在目标变量方向的二维投影，其多元函数可以表示为：

$$\hat{y}=f(x_1,x_2,c) \tag{7}$$

式(7)中，x_1为行距，x_2为播种量，c指其他固定特征向量(株高、穗长、穗质量等)，$\hat{y}$指模型预测产量。行距范围设置为30～50 cm，以0.5 cm为步长，生成41个行距点，精准捕获窄行作物的敏感变化。播种量范围设置为480～720 g，以5 g为步长，生成49个播量点，生成41×49共2 009个预测点。将预测向量转换为41×49矩阵，以产量为z坐标，构建三维响应曲面。通过二阶差分判断极值点，保存曲面最大值，通过此方法，确定最高产量曲面坐标，并将95%最大产量定义为高产标准，同时兼顾高产与操作容错空间。

边际效应通过数值微分法实现，表示某一输入参数单位变化引起的输出变量变化量，主要由一阶偏导产生：

$$\triangle=\frac{\partial\hat{y}}{\partial x_i} \tag{8}$$

式(8)中，$\hat{y}$为模型预测产量，x_i为目标参数。计算行距边际效应时，固定播种量600 g/40 m^2，沿行距方向计算相邻预测产量插值，据此计算行距边际效应；计算播种量边际效应时，固定行距为40 cm，沿播种量方向计算相邻预测产量插值，然后计算播种量边际效应。

2　结果与分析

2.1　冬黑麦苗期生长动态观测与分析

试验结果(表1)显示，在不同播种行距条件下，冬黑麦苗期生长性状表现出显著差异。当行距为30、40、50 cm时，平均株高分别为9.88、9.82和10.21 cm，其中50 cm行距组株高显著高于其他组($P<0.05$)；平均分蘖数分别为4.23、4.07和3.33个，随行距增加而显著递减；平均根长为16.33 cm(40 cm行距)＞15.53 cm(30 cm行距)＞13.83 cm(50 cm行距)。播种量为480、600、720 g/40 m^{2}时，冬黑麦平均株高分别为9.97、9.92和10.16 cm，随着播种量呈小幅递增趋势；但分蘖受到抑制，平均分蘖数由4.8个降至2.5个($P<0.05$)。当播种行距为40 cm、播种量为720 g/40 m^{2}时，群体总干质量达到最大值16.31 g，表现出明显的群体产量优势。

不同行距与播种量处理组合对冬黑麦性状产生了复杂的交互作用。在50 cm行距下，冬黑麦株高表现最优，但受限于单株空间分布，其分蘖数和根长表现相对较差；而30 cm窄行距虽显著促进了分蘖数的增加，但株高表现受到抑制。播种量的增加虽然在一定程度上促进了株高的伸长，但导致单株分蘖数大幅减少，反映出高密度种植对单株发育的竞争性抑制。值得注意的是，40 cm × 600 g/40 m^{2}处理组合下，冬黑麦在单株发育与群体积累之间达到了较好的平衡，群体干物质量达到12.42 g，且各项生长指标较为均衡。

表1　冬黑麦生长不同处理组苗期生长性状对比

Table 1　Comparison of growth traits among different treatment groups of winter rye at seedling stage

行距/cm Spacing	播种量/(g/40 m^2) Seeding rate	株高/cm Plant height	平均分蘖数 Tillers number	根长/cm Root length	叶片数 Leaves number	单株总干质量/g Single plant dry mass	群体总干质量/g Total plant dry mass
30	480	9.88±1.66ab	5.1±1.1a	15.20±2.05a	5.82±0.01a	10.42±0.88c	9.60±1.66ab
	600	9.88±2.19b	4.2±0.9c	15.50±1.98b	4.57±0.12c	11.56±1.05b	9.98±2.19b
	720	9.88±1.54c	3.4±0.7d	15.90±1.45c	3.67±0.08d	9.83±1.21c	10.00±1.54c
40	480	9.50±1.05ab	4.8±0.8a	16.90±1.55a	5.57±0.15a	9.69±0.74c	9.70±0.74ab
	600	9.70±2.22a	4.2±1.0b	16.30±2.11ab	5.21±0.09b	12.42±0.96b	12.42±0.96a
	720	10.20±1.83d	3.2±0.6e	15.80±1.67c	3.05±0.05e	16.31±1.44a	16.31±1.44d
50	480	10.00±1.95a	4.6±1.2ab	14.00±2.34a	5.40±0.11ab	10.85±0.92bc	10.50±0.92a
	600	10.20±1.34c	4.4±0.8de	14.00±1.12c	3.53±0.07d	11.38±1.10b	11.00±1.10c
	720	10.40±2.45c	1.0±0.5e	13.50±1.89c	5.69±0.14a	12.03±1.02b	12.00±1.02c

注：采用One-way ANOVA(单因素方差分析)及Duncan's多重比较，显著性水平设定为$\alpha=0.05$；不同小写字母表示同一性状在不同处理间存在显著差异($P<0.05$)。下同。Note: A one-way ANOVA and Duncan's multiple comparison were employed, with a significance level set at $\alpha=0.05$.Different letters represent the significant difference among different treatments of each trait at 0.05 level.The same below.

较窄播种行距导致植株间资源竞特别是光照竞争加剧，诱导了典型的避荫反应，促进植株的补偿机制，表现为株高降低、分蘖数增加，这种表型可塑性是作物对高密度种植的适应性响应；与此同时，根系

空间的物理限制显著抑制了根系的横向扩展。较宽的播种行距能够提供充足的生长空间，减少株间竞争，单株受光更充分，促进同化物积累。播种量梯度试验表明，随着播种量增加，苗期株高呈小幅递增趋势，但分蘖受到抑制；而高播种量则促进了生物量向地上部的分配。这种密度依赖的生长响应符合经典的自疏法则。40 cm行距×600 g/40 m²播种量处理组合能够保证单株长势较好的同时，群体长势均衡。

2.2 行距与播种量对冬黑麦越冬率的影响

冬黑麦越冬率作为决定产量的关键农艺指标，受到播种行距与播种量的显著调节，且表现出明显的交互效应（表2）。结果显示，30 cm行距配合480 g/40 m²播种量处理的越冬率最高，达到92.50%，显著优于其他处理组（$P<0.05$）。相比之下，40 cm×720 g/40 m²和50 cm×600 g/40 m²组合的越冬率分别为75.38%和75.65%，与最优组之间存在极显著的统计学差异，反映了高密度种植或过宽行距对作物越冬安全性的负面影响。较窄的行距在越冬时减少地表裸露，降低土壤温度昼夜波动，缓冲极端低温冲击但通风较差、可能增加冠层湿度，诱发雪霉病等低温病害；若春季返青后分蘖过多，可能导致水肥竞争加剧。较宽的行距存在双重效应，单株营养空间大，可积累更多的可溶性糖、脯氨酸等物质抵御寒害；同时，更好的通风透光条件能够降低病害发生率；但过宽的行距造成的低覆盖度易导致土壤直接暴露、冻结深度增加（尤其无积雪时），威胁浅层根系和分蘖节，且昼夜冻融交替易引发“冻拔”；裸露地表易受风力侵蚀，加重土壤水分流失。而适中的播种行距，能够平衡群体防护与个体发育的矛盾，但适中行距在高密度下表现最差，显示了行距与密度的交互效应。

2.3 返青至灌浆期冬黑麦株高动态监测

冬黑麦从返青期至灌浆期株高动态变化如图1所示，所有处理组均表现出典型的Logistic生长曲线特征。所有处理组从返青期至灌浆期株高均呈现明显的增长趋势，4月16日，群体平均株高为7.1～73.2 cm；至6月10日灌浆期，最终株高为41.8～82.1 cm。植株在5月10日至5月22日期间生长速度最为显著（日增量最大），这与禾本科作物在抽穗前后的生理发育规律一致。

50 cm行距组表现出最强的生长势，平均株高最高（141.8 cm）；而30 cm窄行距组由于群体内部竞争剧烈，株高显著低于其他处理。这表明株高与播种量呈正相关。在相同行距条件下，720 g/40 m²播种量组的株高始终高于600 g/40 m²和480 g/40 m²处理组。尽管50 cm行距配合720 g/40 m²播种量处理株高最高，但结合生产实际，40 cm行距配合600 g/40 m²播种量在生长中后期表现出更平稳的增长曲线，最终株高稳定在111.1 cm左右，兼顾了植株个体发育与群体稳定性。

为量化不同行距对冠层透光率的影响，基于株高、分蘖数与叶片数等实测数据计算了间接光照评估指数，结果（表3）显示，该处理下的SRI指数最高（0.021 0）显著高于40 cm和50 cm处理（$P<0.05$）且CDI达到10.66。这表明窄行距导致群体过度郁闭，植株间存在强烈的生存胁迫，限制了单株的充分发

表2 不同处理组冬黑麦越冬率

Table 2 Overwintering rate of winter rye in different treatment groups

行距/cm Spacing	播种量/(g/40 m²) Seeding rate	活苗数 Survival plants	死苗数 Dead plants	越冬率/% Overwintering rate
30	480	37.0	3.0	92.50±3.15a
	600	43.3	9.0	82.80±4.58bc
	720	29.0	4.3	87.00±8.44ab
40	480	29.3	6.0	83.01±6.22bc
	600	46.3	9.3	83.23±5.14bc
	720	49.0	16.0	75.38±4.29d
50	480	32.7	5.7	85.22±7.11b
	600	48.7	15.7	75.65±6.85d
	720	42.0	12.0	77.78±5.37cd

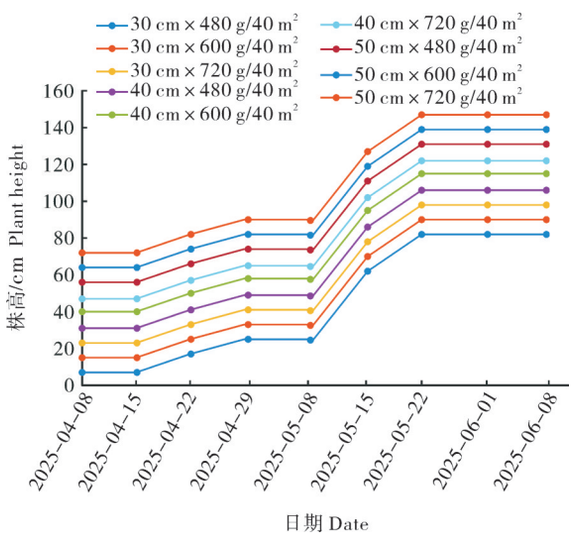


图1 不同处理组冬黑麦返青期至灌浆期株高

Fig.1 Plant height of winter rye from the greening period to the grain filling period in different treatment groups

育，导致其最终株高在所有组别中处于最低水平。50 cm 处理组表现出最高的 GTI(0.370)，说明冠层中下部透光性极佳。然而，由于空间充足，该组植株表现出较强的纵向伸长趋势，虽株高最高，但群体稳定性相对较低。40 cm 行距处理的GTI 与 CDI 间获得了平衡，其 GTI 显著优于窄行距，保证了有效光合面积；同时 SRI 处于适中水平(0.014 5)，说明光竞争胁迫强度适中。

表 3　不同行距处理冬黑麦形态指标与间接光合评估指数

Table 3　Morphological indicators and indirect photosynthetic evaluation index for different row spacing treatments

行距/cm Spacing	平均株高/cm Average plant height	平均分蘖数 Average tillers number	平均叶片数 Average number of leaves	GTI	CDI	SRI
30	88.5±5.4c	2.91±0.35a	10.66±1.12a	0.226c	10.66a	0.021 0a
40	109.1±7.2b	1.94±0.22b	8.82±0.95b	0.299b	8.82b	0.014 5b
50	133.0±8.6a	1.92±0.18b	9.69±1.04ab	0.370a	9.69ab	0.014 0b

综上，40 cm行距配合 600 g/40 m^{2}播种量，能够通过优化群体结构和资源分配效率，实现生长稳定性与产量形成的协同提升。

2.4　行距与播种量互作下冬黑麦产量性状分析

不同处理组冬黑麦产量构成要素的测定结果如表4所示。统计分析表明，行距对穗部形态具有显著的调节作用。从行距梯度来看，宽行距(50 cm)在单穗发育上表现出显著优势，其穗长与百粒质量均达到最大值(分别为 15.24 cm 和 2.58 g)，显著高于 30 cm 窄行距处理(P<0.05)。这表明在较宽行距下，冬黑麦个体受光充足，更有利于光合产物向籽粒转移，促进了籽粒质量的提升。从播种量梯度来看，产量要素对密度的响应遵循平衡与补偿的生物学原则。穗长随播种量增加呈现递减趋势，而单穗种子数量则在 720 g/40 m^2 处理下达到峰值。其中，40 cm 行距配合 600 g/40 m^2 播种量处理在所有组合中表现最为突出，其单穗种子质量达到 23.15 g，显著优于其他处理组(P<0.05)，且标准差最小，显示出极高的生长稳定性；在该组合下，适中的行距(40 cm)提供了较佳的冠层透光率，使单位面积穗数与单穗生产力之间达到了最佳平衡点，有效克服了高密度下的个体竞争负效应。作物的产量与农艺操作工艺存在复杂的非线性关系：在较窄行距(30 cm)下，随播种量增加，穗长无显著变化；而高播种量虽使种子数量达到峰值，但导致个体竞争剧烈，百粒质量降至最低。中等行距(40 cm)配合中等播种量在所有组合中性状表现最佳，其穗质量与种子质量均处于高水平，且播种量的增加对穗节数提供了正向补偿效应。较宽行距(50 cm)配合 480 g/40 m^{2}播种量使穗长与百粒质量达到最大值，随着该行距下播种量的增加，种子数量提升的效果最明显。

表 4　不同处理组产量构成性状

Table 4　Yield traits of different treatment groups

行距/cm Spacing	播种量/(g/40 m^2) Seeding rate	穗长/cm Spike length	穗质量/g Spike mass	单株种子数量 Number of spike nodes	百粒质量/g 100-seeds weight	单穗种子质量/g Single seeds weight per spike
30	480	13.12±0.41c	17.15±0.85d	1 015±42d	1.92±0.11c	19.45±1.05cd
	600	13.38±0.48c	17.52±0.92cd	1 210±65a	1.68±0.08e	21.72±0.95b
	720	13.25±0.38c	16.96±1.15d	1 182±78a	1.74±0.10de	20.15±1.22bc
40	480	13.52±0.35bc	17.45±0.72cd	1 042±48cd	1.95±0.12c	20.92±1.10b
	600	13.84±0.42b	18.62±0.82a	1 086±55b	2.25±0.08b	23.15±0.98a
	720	13.68±0.40bc	17.45±0.98cd	1 076±52bc	1.67±0.09e	19.28±1.25d
50	480	15.24±0.65a	18.42±1.05ab	1 025±51d	2.58±0.12a	21.42±1.15b
	600	14.85±0.58a	18.15±0.88b	1 088±49b	1.58±0.07f	19.92±1.12c
	720	14.46±0.60ab	17.28±0.95d	1 125±62b	1.75±0.10de	19.25±1.08d

通过对产量构成要素的综合分析发现，冬黑麦对农艺措施的响应遵循平衡与补偿的生物学原则，如表4所示，行距由 30 cm增加至50 cm，单穗穗质量与百粒质量分别提升8.1%和33.7%。然而行距与播种量的交互作用展现了复杂的非线性关系。600 g/40 m^2×40 cm行距处理组单穗种子质量达到峰值

(23.15 g),显著优于其他处理组,且综合性状稳定,标准差最小。这表明适中的行距提供了较佳的冠层透光率,使单位面积穗数和单穗生产力之间达到平衡,克服高密度下的个体竞争负效应。

综上所述,40 cm 行距配合 600 g/40 m²播种量为本试验条件下的最优农艺组合。该工艺通过优化冠层结构,在单株发育与群体积累之间达到了最佳协调,其单穗种子质量显著优于其他组合且性状最为稳定。这一结果表明,适中的行距配置能有效缓解高密度种植下的个体竞争负效应,是实现冬黑麦抗逆稳产与高产协同提升的关键。

2.5 融合递归特征消除与随机森林回归模型的建立

基于前期田间试验,在冬黑麦完全成熟期,系统采集不同处理组的植株样本,并采用标准化试验流程对关键农艺性状,包括穗长、穗质量、穗节数、株高、根长、有效分蘖数、叶片数以及单株种子数进行量化表征,共获得540组有效数据,应用于RFE-RF模型构建(图2)。在数据处理环节,采用防御性编程框架,通过try-catch结构确保流程稳健性,运用standardizeMissing函数和中位数填充处理缺失值,然后自动生成包含样本数、特征数和目标变量统计报告。在构建特征工程环节,将农学指标转化为数学特征,采用穗长/穗质量和株/根长等比值特征解释生物量分配规律;采用穗长×穗质量等乘积特征,表征穗部综合发育状况;采用对数变换和平方根变换处理非线性数据,定义safe_divide函数确保计算安全。

在模型优化环节,采用两阶段特征选择策略。首先,基于OOB重要性进行预筛选,保留前50%特征;然后,通过RFE算法进行迭代优化,实现特征动态选择。在模型训练过程中通过网格搜索优化叶子大小(1/3/5)和决策树数量(200/300),并采用改进的交叉验证方法(k_{Fold})和RMSE来评估指标。

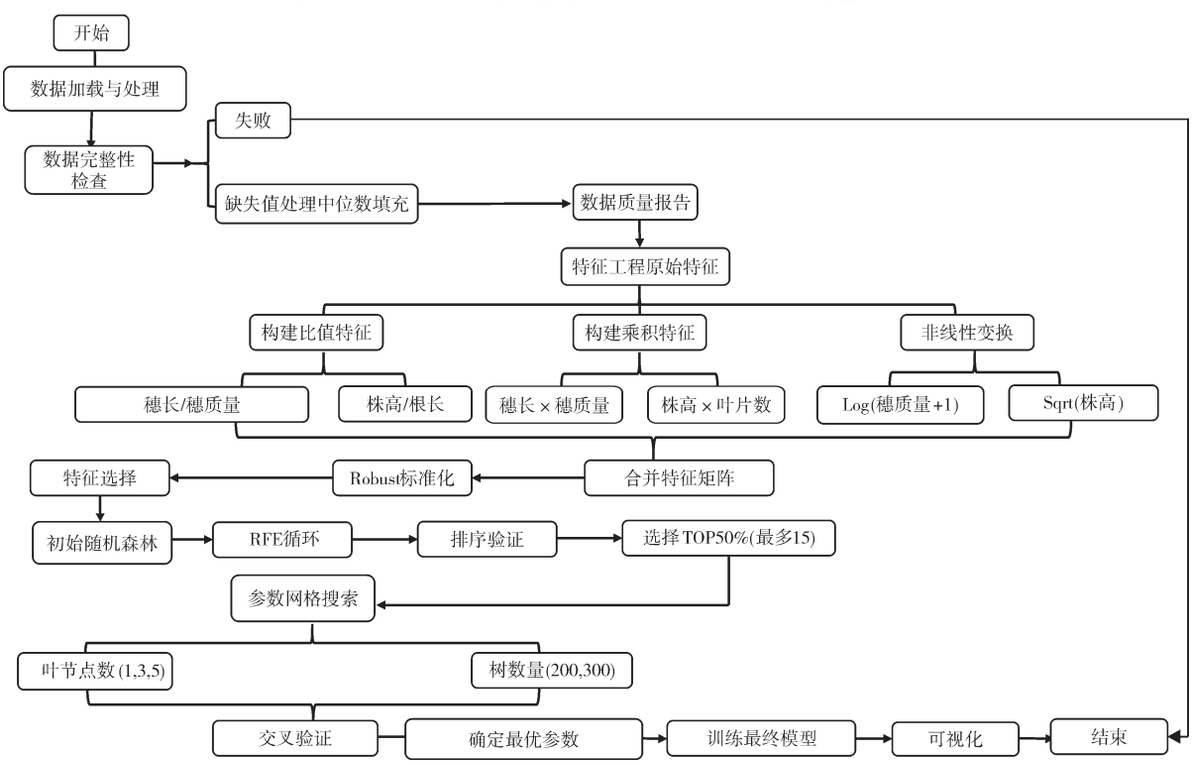


图2 算法结构流程图

Fig.2 Algorithm structure flowchart

2.6 模型参数评估

基于前期田间试验,以单株种子数为预测目标,构建包含播种行距、播种量、穗长、穗质量、穗节数、株高、根长、有效分蘖数、叶片数等原始变量及6个衍生特征的多维特征集。通过随机森林递归特征消除(RF-RFE)算法筛选出8个关键特征,其重要性分布如图3所示。结果显示,log(穗质量+1)特征贡献度最高,这既体现了穗质量作为籽粒产量核心指标的生物学意义,同时,通过数学变换有效缓解了顶端籽粒灌浆不均导致的右偏分布问题;播种行距和播种量分别通过调控群体光能利用率和基本苗数来影响产量构成,其标准化处理增强了特征的可解释性;穗

长×穗质量乘积特征同时捕获穗形态和籽粒充实度，而高重要性分数表明了模型更关注穗部整体潜力而非单一性状；株高×叶片数表征冠层生物量，反映植株光合能力与抗倒伏的平衡，协同影响群体透光性和籽粒充实度，揭示冬黑麦株型结构与产量形成机制；而穗长对产量的解释存在局限性，因此，包含在其他特征中；原始穗质量与 log（穗质量+1）高度相关，且对数变化后信息更有效。本研究提出的 RF-RFE 模型通过特征重要性解析了冬黑麦产量形成的农学机制，可为作物表型组学研究提供可解释的预测方法。

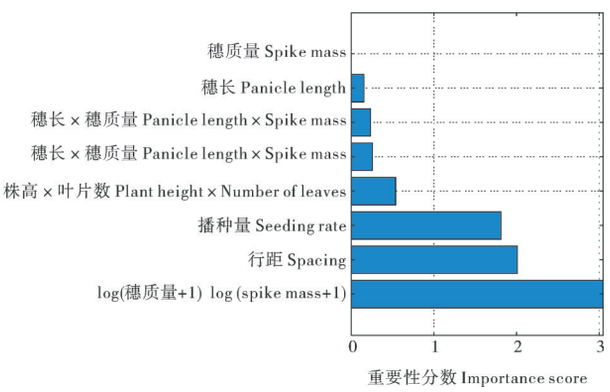


图 3　特征重要性分数图

Fig.3　Feature importance score plot

基于 RF-RFE 模型，超参数优化结果（表 5）显示，采用叶节点数为 1 和 200 棵决策树的组合的 RMSE 值最大，为 6.177，能够实现最优预测。冬黑麦产量形成过程涉及复杂的农艺性状互作，特别是穗部性状之间存在显著的非线性关系，需要精细的模型进行捕捉，当叶节点数增大到 3 或 5 时，模型性能明显下降，RMSE 增加约 4～8 百分点，表明冬黑麦产量预测受多因素协同影响。同时，当决策树数量超过 200 时，呈现边际效应递减规律，说明模型复杂度达到饱和状态。

表 5　超参数优化

Table 5　Hyperparameter optimization

叶节点大小 Leaf node size	树数量 Number of trees	RMSE	排序 Rank
1	200	6.177	1
1	300	6.177	2
3	200	6.196	6
3	300	6.157	3
5	200	6.165	5
5	300	6.159	4

以每穗粒数为例，RF-RFE 模型训练集（图 4A）与测试集（图 4B）预测结果均集中分布在理想线左右，误差集中在 10% 以内，说明模型具有良好的预测效果。从模型性能来看，训练集 RMSE 值为 2.28，相对误差控制在 5% 以内，决定系数 R^2=0.951，证明特征工程发挥了明显作用，特别是对穗质量进行对数变换，显著提高了模型解释力。模型的 OOB-RMSE 为 4.332，与训练误差比值控制在 1.9 倍以内，MAE 约为 1.448（图 5），证明模型具备良好的泛化能力，未出现严重的过拟合现象。

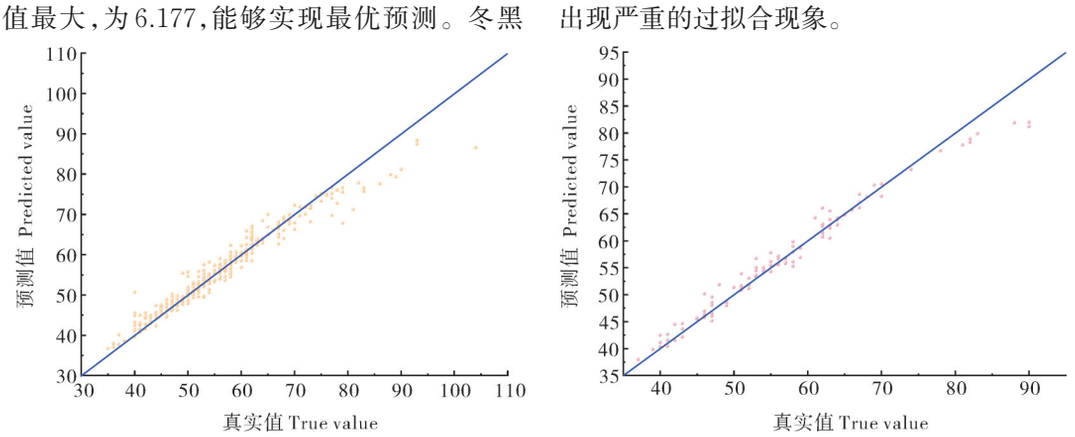


图 4　RF-RFE 模型训练集（A）与测试集（B）的冬黑麦每穗粒数预测结果

Fig.4　Prediction results of grains per spike of winter rye training set（A） and test set（B） of RF-RFE model

2.7　多模型对比分析

为验证 RF-RFE 模型的适用性，选取 XGBoost、LightGBM、支持向量机（SVM）以及多元线性回归（MLR）作为对比模型，使用相同的训练集与测试集进行预测，各模型的性能对比结果（表 6）显示，本研究构建的 RF-RFE 模型在预测精度、稳定性和泛化能力上均表现出显著的优势；在预测精度方面，RF-RFE 模型的决定系数最高，为 0.951，比 XGBoost、LightGBE、SVM（RBF 核）、多元线性回归模型分别提高了 1.9、3.3、5.0 和 10.6 百分点；同时其 RMSE 和

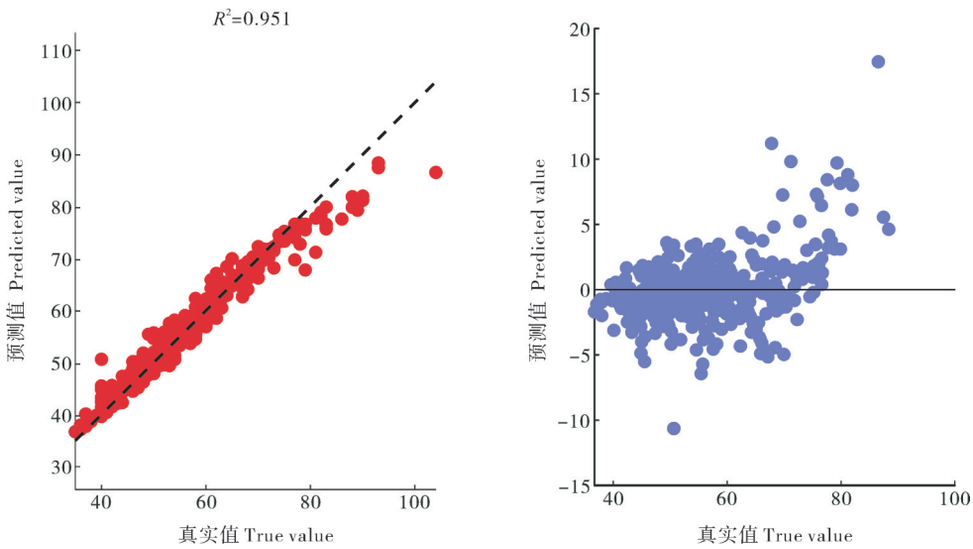


图5 RF-RFE模型决定系数(A)与残差(B)分布图

Fig.5 RF-RFE model determination coefficient(A) and residual distribution(B)

表6 不同模型对比结果

Table 6 Comparison results of different models

模型 Model	决定系数 R^2	均方根误差 RMSE	平均绝对误差 MAE	排名 Rank
RF-RFE	0.951	2.28	1.45	1
XGBoost	0.932	2.65	1.68	2
LightGBE	0.928	2.71	1.72	3
SVM(RBF)	0.901	3.12	2.01	4
多元线性回归 MLR	0.845	3.89	2.54	5

MAE均为最低值,表明RF-RFE模型预测结果也最为可靠。因此,本研究构建的RF-RFE模型在预测精度、稳定性和泛化能力上均表现出显著的优势。

本研究的特征集包含原始变量及衍生变量,且变量间存在复杂的非线性关系,RF-RFE模型集成了随机森林处理非线性关系的优势与RFE的特征选择能力。与梯度提升树模型(XGBoost/LightGBM)相比,虽然算法均为优秀的集成算法,但RF-RFE通过两阶段的特征选择,更易剔除冗余特征、降低噪声干扰,从而获得更优的泛化能力,而XGBoost和LightGBM虽然构建单棵树方面更为高效,但在本研究的数据集上可能对某些异常值或噪声更为敏感,RF-RFE模型指标更为优异。SVM在处理高维数据时能依赖于核函数与参数的选择,即使使用径向基核函数,SVM的预测精度仍低于RF-RFE模型,基于决策树集成的RF-RFE模型具有更强的适应性和稳健性。MLR模型决定系数为0.845,显著低于其他机器学习模型,表明冬黑麦产量与表型性状间存在明显的非线性关系,传统线性假设无法有效描述其内在规律。RF-RFE模型在获得最高的精度的同时,其OOB-RMSE与训练误差比值约为1.9,也在合理的范围,未出现明显的过拟合现象。

综合比较发现,RF-RFE模型凭借其独特的集成学习框架与递归特征消除策略,在本研究涉及的冬黑麦高维、非线性农艺数据集中表现出了最佳的预测性能与稳健性。

2.8 行距-播种量响应曲面反演与边际效应

由三维响应曲面(图6)分析可知,高产区间为行距35～49 cm、播种量535～680 g/40 m^2,优化方案可选择行距(40±2) cm、播种量(600±50) g/40 m^2。

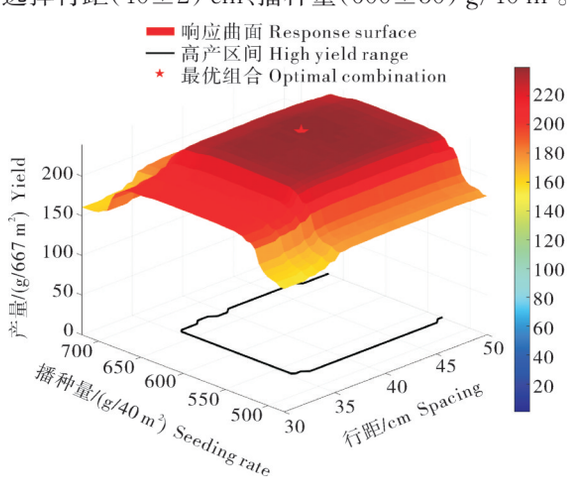


图6 行距-播种量三维响应曲面

Fig6 Three-dimensional response surface of row spacing-seeding rate

由行距边际效应分析结果(图7A)可见,行距≤40 cm时,斜率为正,属于增产区域;行距为40～45 cm时,斜率趋近于0,产量没有明显增加,属于效益

平台区；当行距≥45 cm时，斜率为负值，属于减产区域；当行距≥45 cm时，每增加1 cm行距，相应地减产1.23 kg。由播种量边际效应（图7B）可见，当播种量为650 g/40 m²时，出现效益拐点，超过此阈值虽仍增产但边际收益递减。因此，建议生产实践中将行距控制在40～45 cm、播种量控制在600～650 g/40 m²，既可保证高产又留出操作容错空间。

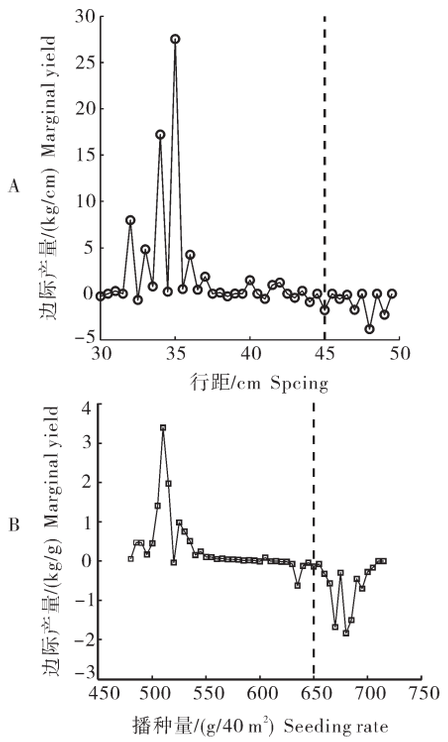


图7　行距（A）与播种量（B）边际效应

Fig 7　Marginal response diagram of row spacing（A） and seeding rate（B）

3　讨　论

本研究通过行距梯度与播种量梯度处理双因素田间试验，解析了冬黑麦行距与播种量的互作效应。结果表明，40 cm行距×600 g/40 m²播种量处理组合在农艺形状上表现出显著优势，该处理干物质积累达16.31 g、越冬率高、株高动态稳定性最优且产量构成要素协调。这一处理组合通过优化冠层透光率、平衡“源-库”关系及提升抗逆性，实现了冬黑麦高产稳产的协同调控。

基于田间试验，以单株种子数为预测目标，构建了包含播种行距、播种量、穗长、穗质量、穗节数、株高、根长、有效分蘖数、叶片数等原始变量及6个衍生特征的多维特征集，构建融合递归特征消除的随机森林模型（RF-RFE），通过特征工程筛选出包括穗质量对数变换、行距-播种量交互项等8个关键特征变量，模型表现出优异的预测性能（R^2=0.951，RMSE=2.28）和泛化能力（OOB-RMSE=4.33）。三维响应曲面分析确定了高产参数区间为行距35～49 cm、播种量535～680 g；边际效应分析显示，行距超过45 cm时每增加1 cm将导致减产1.23 kg，播种量高于650 g/40 m²时呈现边际收益递减规律。结果表明，通过优化群体结构和资源分配效率，可实现冬黑麦高产稳产的协同调控，为寒区粮饲作物的智能化管理提供可行的技术范式。

尽管本研究为冬黑麦的精准栽培提供了可行的决策方案，但在理论及应用上仍有待深入。本研究构建的模型主要基于单一试验点的农艺性状数据构建，在不同生态区及气候波动下的稳健性仍有待验证。未来将整合多试点、多年份的田间数据，并纳入时序遥感影像、高精度气象数据以及土壤墒情与养分数据，使用时空图神经网络、注意力机制与循环神经网络融合构建多尺度、多源数据融合的动态预测，以捕捉环境因子与作物生长的时空动态互作关系，从而建立一套适应不同区域、具有泛化能力的寒区冬黑麦产量与品质协同预测算法。

参考文献 References

[1] CHIKMAWATI T，MIFTAHUDIN M，SKOVMAND B，et al. Amplified fragment length polymorphism-based genetic diversity among cultivated and weedy rye（*Secale cereale* L.） accessions [J]. Genetic resources and crop evolution，2012，59（8）：1743-1752.

[2] KOISTINEN V M，HANHINEVA K. Microbial and endogenous metabolic conversions of rye phytochemicals [J/OL]. Molecular nutrition & food research，2017，61（7）：1600627 [2025-11-23]. https://doi.org/10.1002/mnfr.201600627.

[3] BAZHENOV M，LITVINOV D，KARLOV G，et al. Evaluation of phosphate rock as the only source of phosphorus for the growth of tall and semi-dwarf durum wheat and rye plants using digital phenotyping [J/OL]. PeerJ，2023，11：e15972 [2025-11-23]. https://doi.org/10.7717/peerj.15972.

[4] 王毓，李奇铮，马小东，等. 播种期及收获期对饲用小黑麦生产性能和营养品质的影响 [J]. 草地学报，2024，32（2）：637-645. WANG Y，LI Q Z，MA X D，et al. Effects of different sowing and cutting periods study on the production performance and nutritional value of *Triticosecale* wittmack in rainfed area of Ningxia [J]. Acta agrestia sinica，2024，32（2）：637-645（in Chinese with English abstract）.

[5] 孔广超，曹连莆，艾尼瓦尔，等. 高产优质小黑麦新品种：新小黑麦3号 [J]. 麦类作物学报，2005，25（2）：149. KONG G C，CAO L P，AIN，et al. New *Triticale* No.3：a new *Triticale* variety with high yield and good quality [J]. Acta trital crops，

2005,25(2):149 (in Chinese).

[6] 王汇涵,张泽,康孝岩,等.基于Sentinel-2A的棉花种植面积提取及产量预测[J].农业工程学报,2022,38(9):205-214. WANG H H,ZHANG Z,KANG X Y,et al.Cotton planting area extraction and yield prediction based on Sentinel-2A[J]. Transactions of the CSAE,2022,38(9):205-214 (in Chinese with English abstract).

[7] 王莺,巩垠熙.遥感光谱技术在农作物估产中的应用研究进展[J].中国农学通报,2019,35(3):69-75. WANG Y,GONG Y X. Spectral remote sensing technology applied in crop yield estimation:research progress[J].Chinese agricultural science bulletin,2019,35(3):69-75 (in Chinese with English abstract).

[8] CLAUSS K,OTTINGER M,LEINENKUGEL P,et al.Estimating rice production in the Mekong Delta,Vietnam,utilizing time series of Sentinel-1 SAR data[J].International journal of applied earth observation and geoinformation, 2018, 73: 574-585.

[9] BATTUDE M, AL BITAR A, MORIN D, et al. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data[J].Remote sensing of environment,2016,184:668-681.

[10] TAŞAN S,CEMEK B,TAŞAN M,et al.Estimation of eggplant yield with machine learning methods using spectral vegetation indices [J/OL]. Computers and electronics in agriculture, 2022, 202: 107367 [2025-11-23]. https://doi. org/10.1016/j.compag.2022.107367.

[11] 兰铭,费帅鹏,禹小龙,等.多光谱与热红外数据融合在冬小麦产量估测中的应用[J].麦类作物学报,2021,41(12):1564-1572. LAN M,FEI S P,YU X L,et al.Application of multispectral and thermal infrared data fusion in estimation of winter wheat yield[J].Journal of triticeae crops,2021,41(12):1564-1572 (in Chinese with English abstract).

[12] 甄佳宁,蒋侠朋,赵德梅,等.利用Sentinel-2影像超分辨率重建的红树林冠层氮含量反演[J].遥感学报,2022,26(6):1206-1219. ZHEN J N,JIANG X P,ZHAO D M,et al.Retrieving canopy nitrogen content of mangrove forests from Sentinel-2 super-resolution reconstruction data [J]. National remote sensing bulletin, 2022, 26(6):1206-1219 (in Chinese with English abstract).

[13] 郭晗,陆洲,徐飞飞,等.基于全局敏感性分析与机器学习的冬小麦叶面积指数估算[J].浙江农业学报,2022,34(9):2020-2031. GUO H,LU Z,XU F F,et al.Leaf area index estimation of winter wheat based on global sensitivity analysis and machine learning[J].Acta agriculturae Zhejiangensis,2022,34(9):2020-2031 (in Chinese with English abstract).

[14] PANEK-CHWASTYK E,PARADOWSKI K,RUTKOWSKA B,et al.Advancing early-stage plant phosphorus assessment for winter rye *via* hyperspectral data:a model-based approach harnessing feedforward neural networks[J/OL].European journal of agronomy, 2025, 169: 127667[2025-11-23]. https://doi.org/10.1016/j.eja.2025.127667.

[15] HANSEN V,ENGEDAL T,RASMUSEN J,et al.Quantifying cover crop-derived above- and belowground carbon inputs including rhizodeposition [J/OL]. Agriculture, ecosystems & environment, 2025, 389: 109668 [2025-11-23]. https://doi.org/10.1016/j.agee.2025.109668.

[16] 李涛宇,许秀春,杨轩,等.利用机器学习预测华北地区冬小麦农田氮淋失[J].中国环境科学,2025,45(1):343-354. LI T Y,XU X C,YANG X,et al.Prediction of nitrogen leaching from winter-wheat production in North China based on random forest and XGBoost [J]. China environmental science, 2025,45(1):343-354 (in Chinese with English abstract).

[17] JOHNSON M D,HSIEH W W,CANNON A J,et al.Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods[J].Agricultural and forest meteorology,2016,218/219:74-84.

[18] 刘峻明,和晓彤,王鹏新,等.长时间序列气象数据结合随机森林法早期预测冬小麦产量[J].农业工程学报,2019,35(6):158-166. LIU J M,HE X T,WANG P X,et al.Early prediction of winter wheat yield with long time series meteorological data and random forest method[J].Transactions of the CSAE, 2019,35(6):158-166 (in Chinese with English abstract).

[19] HUBER F,YUSHCHENKO A,STRATMANN B,et al.Extreme gradient boosting for yield estimation compared with deep learning approaches[J/OL].Computers and electronics in agriculture,2022,202:107346[2025-11-23].https://doi.org/10.1016/J.COMPAG.2022.107346.

[20] NYÉKI A,KEREPESI C,DARÓCZY B,et al.Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods[J].Precision agriculture,2021,22(5):1397-1415.

[21] CAI Y P,GUAN K Y,LOBELL D,et al.Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches[J].Agricultural and forest meteorology,2019,274:144-159.

[22] CĂLIN A D,COROIU A M,MUREŞAN H B.Analysis of preprocessing techniques for missing data in the prediction of sunflower yield in response to the effects of climate change[J/OL]. Applied sciences, 2023, 13(13): 7415 [2025-11-23]. https://doi.org/10.3390/app13137415.

[23] BANERJEE S, NANDI T, SATI V P, et al. Integrating remote sensing, landscape metrics, and random forest algorithm to analyze crop patterns, factors, diversity, and fragmentation in a kharif agricultural landscape[J/OL].Land, 2025, 14(6): 1203[2025-11-23].https://doi.org/10.3390/land14061203.

[24] MENG X Y,ZHANG S,WANG G Q,et al.Decoding agricultural drought resilience:a triple-validated random forest framework integrating multi-source remote sensing for high-resolution monitoring in the North China plain[J/OL].Remote sensing, 2025, 17 (8): 1404 [2025-11-23]. https://doi. org/10.3390/RS17081404.

[25] LUO D,QIAO X Q.Regional agricultural drought vulnerability prediction based on interpretable random forest[J/OL].Environmental monitoring and assessment, 2024, 196(11): 1123 [2025-11-23].https://doi.org/10.1007/S10661-024-13203-Y.

[26] DHANOA M S,SANDERSON R,LISTER S J,et al.Decision tree learning with random forest models using agricultural and

ecological field data incorporating multi-factor studies and covariate structure [J/OL]. CABI reviews, 2024: cabireviews, 2024.0030 [2025-11-23]. https://doi.org/10.1079/CABIREVIEWS.2024.0030.

[27] 王俊,吴振伟,姜海,等.基于随机森林及遥感植被指数的无人农场水稻产量预测研究[J].智能化农业装备学报(中英文),2025,6(2):97-104. WANG J, WU Z W, JIANG H, et al. Rice yield prediction of unmanned farm based on random forest and remote sensing vegetation index[J]. Journal of intelligent agricultural mechanization, 2025, 6(2): 97-104 (in Chinese with English abstract).

[28] 徐浩,宋华鲁,张海波,等.基于随机森林的作物模型光温产量潜力模拟优化方法[J].湖北农业科学,2024,63(8):132-139. XU H, SONG H L, ZHANG H B, et al. The crop light temperature yield potential simulation optimization method based on random forest[J]. Hubei agricultural sciences, 2024, 63(8): 132-139 (in Chinese with English abstract).

[29] MADDONNI G A, OTEGUI M E. Leaf area, light interception, and crop development in maize[J]. Field crops research, 1996, 48(1): 81-87.

[30] POORTER H, NIKLAS K J, REICH P B, et al. Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control [J]. New phytologist, 2012, 193(1): 30-50.

Establishment of a model for predicting yield of winter rye with a hybrid random forest-recursive feature elimination

LÜ Linyou[1,2], LI Jiahui[1], ZHAO Yan[2], MA Yan[1], HE Ziyuan[1]

1. *School of Environmental Science and Engineering, Liaoning Technical University, Fuxin 123000, China;*
2. *Liaoning Province Institute of Sand Land Management and Utilization, Fuxin 123000, China*

Abstract　A two-factor field gradient experiment was conducted to solve the problems of traditional models being difficult to analyze the nonlinear effects of interaction between row spacing and sowing rate. The biological traits and yield components at the stages of key growth in winter rye were systematically measured. The patterns of interaction between population dynamics and yield components during these stages were analyzed to construct a recursive gradient fusion-based random forest model for predicting the yield of winter rye. The results of field experiment showed that the combination of a row spacing of 40 cm and a sowing rate of 535-680 g/40 m^2 can optimize the light transmittance of canopy, balance the "source sink" relationship, and improve the resistance to stress, achieving coordinated regulation of high and stable yield of winter rye. Samples of plant were systematically collected from different treatment groups at the stage of full maturity in winter rye. A standardized experimental procedure was used to quantitatively characterize key agronomic traits including spike length, spike mass, number of spike nodes, plant height, root length, effective tiller count, leaf count, and seeds per plant. 540 sets of valid data were obtained. A multidimensional feature set encompassing sowing row spacing, seeding rate, spike length, spike mass, number of spike nodes, plant height, root length, effective tiller count, leaf count, and six derived features was constructed using seeds per plant as the prediction target. A hybrid random forest-recursive feature elimination (RF-RFE) with excellent predictive performance (R^2=0.951, RMSE=2.28) and generalization ability (OOB-RMSE=4.33) was constructed. The model for predicting yield was used to invert and obtain a three-dimensional response surface for row spacing and seeding rate to identify the high-yield parameters as row spacing ranging from 35 to 49 cm and seeding rate between 535 and 680 g/40 m^2. The results of marginal effect showed that a 1 cm increase in row spacing beyond 45 cm led to a yield decrease of 1.23 kg, and the marginal benefit decreased as the seeding rate exceeded 650 g/40 m^2. Therefore, it is recommended to control the row spacing within 40-45 cm and the seeding rate between 600-650 g/40 m^2 in the practice of production, which can ensure high yield and leave room for the tolerance of operational error.

Keywords　winter rye; random forest; recursive gradient boosting; yield prediction; marginal effects; three-dimensional response surface

(责任编辑:张志钰)