



基于增强回归树的海河平原小麦赤霉病预测模型构建与验证

陶晔¹, 齐永志¹, 屈赞², 曹志艳¹, 赵绪生¹, 甄文超³✉

¹ 河北农业大学植物保护学院, 河北保定 071001; ² 河北农业大学现代教育技术中心, 河北保定 071001; ³ 河北农业大学农学院/华北作物改良与调控国家重点实验室/河北省作物生长调控重点实验室, 河北保定 071001

摘要:【背景】自1995年至今,小麦赤霉病(Fusarium head blight, FHB)逐渐在海河平原蔓延,由零星出现演变成连片发生,在流行年份呈现出暴发快、面积大、损失重的特点,小麦赤霉病已由次要病害上升为主要病害之一。准确的预测预报是有效控制小麦赤霉病发生与发展的关键和难点。【目的】根据海河平原小麦赤霉病发生情况的监测分析,构建适宜的小麦赤霉病预测模型,为科学防控赤霉病提供技术支撑。【方法】基于2001—2016年海河平原21个小麦主产县(市)的赤霉病病穗率数据,以及小麦关键生育期内的气象数据,采用逐步回归分析,筛选影响小麦赤霉病发生的关键气象因子,构建基于多元线性回归模型和增强回归树模型的小麦赤霉病发生预测模型。【结果】明确了增强回归树模型的学习效率(lr)为0.005、树的复杂度(tc)为6时,模型的预测偏差最低,残差标准误为0.006311;筛选出8个对海河平原小麦赤霉病发生影响显著的关键气象因子,即MRH₁₅、Rain₃₅、MRH₅₅、SD₁₅、LT₄₅、MWS₃₅、MT₂₅、DRain₁₅,并构建了含有8个预测变量的多元线性回归模型($R^2=0.8158$, 校正 $R^2=0.8018$, $P<2.2\times 10^{-16}$)。同时,应用增强回归树模型评估了上述8个关键气象因子的重要性,分别为69.62%、14.08%、4.89%、4.34%、3.35%、2.02%、1.20%、0.50%;根据重要的预测变量进一步简化预测模型,构建了含有4个预测变量的多元线性回归模型($y=-19.45376+0.11689MRH_{15}+0.17346Rain_{35}+0.04185SD_{15}+0.26592MRH_{55}$, $R^2=0.7575$, 校正 $R^2=0.7468$, $P<2.2\times 10^{-16}$);当预测变量由8个调减至4个时,利用2008、2010、2012年安新、定州、馆陶等地历史数据验证模型预测病穗率的准确度,多元线性回归模型预测准确度由88.43%降至85.90%,增强回归树模型预测准确度由87.72%升至91.23%;利用2001—2016年正定、栾城的历史数据验证模型预测病穗率的准确度,两个模型预测准确度无显著变化,多元线性回归模型预测准确度由87.53%变为87.42%,增强回归树模型预测准确度由89.20%变为89.21%。整体而言,多元线性回归模型预测准确度呈下降趋势,而增强回归树模型预测准确度呈上升趋势。【结论】研究构建了含有4个预测变量的增强回归树模型,其预测准确度达89.21%,病穗率预测值与实际观测值的波动趋势基本一致,表明增强回归树模型在海河平原小麦赤霉病预测预报中具有很好的应用前景。

关键词: 小麦赤霉病; 禾谷镰孢; 预测模型; 增强回归树

Construction and Verification of Fusarium Head Blight Prediction Model in Haihe Plain Based on Boosted Regression Tree

TAO Bu¹, QI YongZhi¹, QU Yun², CAO ZhiYan¹, ZHAO XuSheng¹, ZHEN WenChao³✉

¹College of Plant Protection, Hebei Agricultural University, Baoding 071001, Hebei; ²Modern Educational Technology Center, Hebei Agricultural University, Baoding 071001, Hebei; ³College of Agronomy, Hebei Agricultural University/State Key Laboratory of North China Crop Improvement and Regulation/Key Laboratory of Regulation and Control of Crop Growth of Hebei, Baoding 071001, Hebei

收稿日期: 2021-02-01; 接受日期: 2021-03-02

基金项目: 国家重点研发计划(2017YFD0300906, 2018YFD0300502)、河北省现代农业产业技术体系(HBCT2018010205)

联系方式: 陶晔, Tel: 0312-7526131; E-mail: taobu@hebau.edu.cn. 齐永志, E-mail: qiyongzhi1981@163.com. 陶晔和齐永志为同等贡献作者。通信作者甄文超, Tel: 0312-7528996; E-mail: wenchao@hebau.edu.cn

Abstract: 【Background】 Since 1995, *Fusarium head blight* (FHB) has gradually spread and risen from a secondary disease to a major disease in Haihe Plain, from sporadic occurrence to continuous occurrence, showing the characteristics of rapid outbreak, large area and heavy loss in epidemic years. To realize effective prevention and control of FHB, accurate forecasting technology is an important prerequisite for controlling the occurrence and development of FHB. 【Objective】 According to the occurrence of FHB in Haihe Plain, the prediction model of FHB suitable for Haihe Plain was established to provide technical supports for scientific prevention and control of FHB. 【Method】 Based on the data about spike rate of FHB and meteorological factors of key growth stage of wheat in 21 counties of Haihe Plain from 2001 to 2016, the key meteorological factors which have significant influences on the FHB occurrence in Haihe Plain were screened by stepwise regression analysis, and the prediction models of FHB occurrence based on multiple linear regression model and boosted regression tree model were constructed, respectively. 【Result】 When the learning efficiency (lr) of the boosted regression tree model was 0.005 and the complexity (tc) of the tree was 6, the prediction deviation of the model was the lowest, and the residual standard error was 0.006311. Eight key meteorological factors, including MRH_{15} , $Rain_{35}$, MRH_{55} , SD_{15} , LT_{65} , MWS_{55} , MT_{25} and $DRain_{15}$, which had a significant impact on the occurrence of FHB in Haihe Plain, were screened out, and a multiple linear regression model with eight predictive variables was established ($R^2=0.8158$, corrected $R^2=0.8018$, $P<2.2\times 10^{-16}$). Meanwhile, the importance of each key meteorological factor was evaluated by using the boosted regression tree model, with the values of 69.62%, 14.08%, 4.89%, 4.34%, 3.35%, 2.02%, 1.20% and 0.50%, respectively. According to the key predictive variables, the prediction model was further simplified, and a multiple linear regression model with four predictive variables was constructed ($y=-19.45376+0.11689MRH_{15}+0.17346Rain_{35}+0.04185SD_{15}+0.26592MRH_{55}$, $R^2=0.7575$, corrected $R^2=0.7468$, $P<2.2\times 10^{-16}$). When the prediction variables was reduced from 8 to 4, the prediction accuracy of the multiple linear regression model decreased from 88.43% to 85.90%, but the prediction accuracy on the disease spike rate of the boosted regression tree model increased from 87.72% to 91.23%, which was verified by using the historical data of Anxin, Dingzhou and Guantao, etc in 2008, 2010 and 2012. The prediction accuracy on the disease spike rate of the multiple linear regression model and the boosted regression tree model changed from 87.53% to 87.42% and from 89.20% to 89.21%, respectively, but there was no significant difference between the multiple linear regression model and the boosted regression tree model, when they were verified with the historical data of Zhengding and Luancheng from 2001 to 2016. In a word, the prediction accuracy of multiple linear regression model showed a downward trend, while the prediction accuracy of boosted regression tree model showed an upward trend. 【Conclusion】 In this study, the boosted regression tree model with four predictive variables was constructed, with the prediction accuracy of 89.21%. At the same time, the disease spike rate predicted by the boosted regression tree model was basically consistent with the observed fluctuation trend, indicating that the boosted regression tree model had a good application prospect in the prediction of FHB in Haihe Plain.

Key words: *Fusarium head blight* (FHB); *Fusarium graminearum*; prediction model; boosted regression tree (BRT)

0 引言

【研究意义】小麦赤霉病 (*Fusarium head blight*, FHB) 是小麦生产上发生面积最广、危害程度最大的麦类病害之一^[1], 该病是以禾谷镰孢 (*Fusarium graminearum*) 为主要致病菌的真菌性病害^[2]。据报道, 自 1990 年以来美国小麦种植面积因赤霉病流行不断压缩, 2018 年小麦种植面积减少 1 200 万公顷^[3], 2016 年小麦赤霉病造成加拿大萨斯克彻温省经济损失约为 10 亿美元^[4]。同时病菌产生的脱氧雪腐镰刀菌烯醇 (DON 毒素) 和玉米赤霉烯酮 (ZEN 毒素) 等危害人畜健康, 对小麦品质和产量造成严重影响^[5-6]。近年来, 因气候条件变化、耕作制度改变, 我国小麦赤霉病发生呈日趋严重的趋势, 由长江中下游麦区逐渐向北扩展, 淮河流域地区成为重发区, 在黄

淮北片麦区和北部冬麦区也成为常发病害^[7], 2010 年以来重发频率在 50% 以上, 2015、2016、2018 年发生面积均超过 550 万公顷^[8]。自 1995 年以来, 小麦赤霉病逐渐在海河平原 (也称河北平原) 蔓延, 已由零星出现逐渐演变成连片发生, 并由次要病害上升为主要病害之一, 年均发生面积达 26.7 万公顷以上^[7]。小麦赤霉病在流行年份具有短期内暴发快、面积大、损失重的特性, 因此, 明确海河平原影响赤霉病发生的关键气象因子, 建立适宜该区域的病害预测模型, 提供准确的预测预报信息, 对有效防控病害蔓延具有重要意义。【前人研究进展】DE WOLF 等^[9-10]以小麦开花前 7 d 和开花后 10 d 的气象因子作为预测变量, 用逻辑回归建立了小麦赤霉病测报模型; 在此基础上, SHAH 等以品种抗性、玉米残茬以及前期研究获得的 4 个气象因子作为变量,

通过 R 语言，建立了基于 Leaps and Bounds 算法的 Logistic 回归模型^[11]和增强回归树（boosted regression tree, BRT）模型^[12]，结果表明，增强回归树模型误判率低于 Logistic 回归模型；HOOKER 等^[13]以抽穗前 4—7 d 降雨天数和温度为预测变量，建立了含有指数项的模型，预测小麦 DON 毒素含量；DEL PONTE 等^[14]以空中孢子捕捉量和感病组织为基础建立测报模型；ROSSI 等^[15]以菌源量、小麦关键生育期为基础，综合考虑日产孢率、孢子分散率、侵染机率和小麦生育期等因素，预测小麦赤霉病发生风险；MUSA 等^[16]建立了基于 web 的瑞士小麦赤霉病预警系统 FusaProg，预测小麦赤霉病发生、DON 毒素含量并指导杀菌剂科学使用。国内专家学者从气候预测、菌量预测、气候菌量相结合预测等方面展开了研究，建立了长期预测、中期预测与短期预测模型，同时，借助神经网络^[17-18]、支持向量机^[19]、无人机高光谱图像^[20]等技术，不断提高了预测预报的准确度。一般情况下，预测模型存在可移植性差、跨地区应用准确度下降等问题。【本研究切入点】增强回归树是以分类回归树（classification and regression tree, CART）算法为基础的一种自学方法，通过自我学习和随机选择生成多重回归树，提高模型稳定性和预测精度。YOU 等利用该模型明确了环境变量与品种对牧草病害发生的影响，取得了较好的效果^[21]，为评估小麦赤霉病主要影响因子重要性提供了一种新的思路。【拟解决的关键问题】根据影响小麦赤霉病流行的关键生育期，选择温度、湿度、降雨、日照、风速等气象因子为预测变量，筛选出重要预测变量，并分析其对病害发生的影响，以期提升模型预测准确度，为小麦赤霉病发生预测预报提供参考，同时也可建立该病害综合、高效防控体系提供技术支撑。

1 材料与方法

试验数据采集于 2014—2016 年完成，数据处理、模型构建及检验于 2017—2019 年完成。

1.1 病害数据和气象信息采集

收集整理 2001—2013 年海河平原小麦主产区安新、望都、定州、新乐、正定、无极、栾城、辛集、平山、行唐、灵寿、阜城、武邑、景县、临西、宁晋、磁县、馆陶、曲周、永年、大名共 21 个县（市）定点监测小麦赤霉病发病情况基本数据，来源于河北省植保植检总站。

2014—2016 年在上述 21 县（市）田间调查小麦赤霉病发生情况，每县选择 10 个调查点，每个调查点随机取样 500 穗，并计算病穗率。依据国家《小麦赤霉病测报技术规范》GB/T15796—2011 将小麦赤霉病划分为不发生、轻度流行、中度流行、重度流行 4 个等级：不发生（0 级，病穗率<0.1%，对小麦生产未造成减产）、轻度流行（1 级，0.1%≤病穗率<5%，对小麦生产造成局部减产）、中度流行（2 级，5%≤病穗率<10%，对小麦生产造成部分减产）、重度流行（3 级，病穗率≥10%，对小麦生产造成明显减产）。

小麦生育期内气象资料：来源于河北省气象局 2001—2016 年 21 个县（市）逐日最高温度、最低温度、平均温度、日照数据、平均风速、平均相对湿度、总降雨量等。

1.2 数据处理

河北省南部和北部冬小麦生育进程存在一定的时间差^[22]，南部麦区比中部麦区的播种期、越冬期晚 3—5 d，其他生育期早 3—7 d；北部麦区比中部麦区的播种期、越冬期早 5—7 d，其他生育期晚 5—7 d（表 1）。按照河北中部麦区常年小麦生育进程分别计算 21 个县（市）小麦抽穗期初始日期，每 5 d 编为一组，分别以当地的小麦抽穗期初始日为起点，小麦抽穗期初始日向前选择 6 组，即抽穗期初始日向前 26—30、21—25、16—20、11—15、6—10 和 1—5 d，抽穗期初始日向后选择 2 组，即抽穗期初始日向后 1—5、6—10 d。以中部麦区的正定、栾城等地最高温度为例，选择小麦抽穗期初始日（5 月 1 日）为起点，HT_{.65}、HT_{.55}、HT_{.45}、HT_{.35}、HT_{.25}、HT_{.15} 分别代表小麦抽穗

表 1 河北省中部麦区常年小麦生育进程
Table 1 Growth process of wheat in middle wheat region in Hebei Province^[22]

生育期 Growth stage	播种期 Sowing stage	越冬期 Overwintering stage	返青期 Turning green stage	起身期 Rising stage	拔节期 Jointing stage	抽穗期 Heading stage	成熟期 Maturity stage
时间 Time	10 月 5 日至 12 日 October 5th to 12th	11 月底至 12 月初 From the end of November to the beginning of December	2 月底至 3 月初 From the end of February to the beginning of March	3 月下旬 Later March	4 月初 Early April	4 月底至 5 月初 From the end of April to the beginning of May	6 月 10 日至 15 日 June 10th to 15th

期初始日之前 26—30、21—25、16—20、11—15、6—10 和 1—5 d 的最高温度平均值, HT_{15} 、 HT_{25} 分别代表小麦抽穗期初始日之后 1—5、6—10 d 的最高温度平均值。主要包括 9 个气象因子: 最高温度 (HT)、最低温度 (LT)、平均温度 (MT)、平均风速 (MWS)、平均相对湿度 (MRH)、相对湿度 $\geq 85\%$ 天数 (RH85)、降雨天数 (DRain)、总降雨量 (Rain)、总日照时数 (SD)。

1.3 预测模型构建及参数选择

1.3.1 预测模型构建 研究以不同气象因子为预测变量、小麦赤霉病病穗率为响应变量, 采用增强回归树模型建模, 模型拟合使用 R 语言 (3.6.1 版本, R 核心开发组, 2019) gbm 包和 Elith 的函数包^[23]。增强回归树结合了提升 (boosting) 和分类回归树 (CART) 两种技术, 通过组合大量相对简单的决策树的方式以优化模型的预测性能, 模型可写成 M 棵分类回归树相加的形式。

$$f_M(X) = \sum_{m=1}^M T_m(X, \gamma_m) \quad (1)$$

式中, $T_m(X, \gamma_m)$ 为第 m 棵分类回归树, X 为气象因子预测变量, γ_m 为其参数, 是该决策树分裂点和叶子结点的赋值, 求解 γ_m 的过程即为单棵决策树的学习过程。

1.3.2 模型参数选择 在模型运行过程中, 需要优化迭代次数 (the number of trees, nt)、树的复杂度 (tree complexity, tc)、学习效率 (learning rate, lr)、抽样比率 (bag fraction, bf)、函数损失形式 (distribution)、交叉验证折数 (cv.folds) 等参数^[12]。树的复杂度即为单棵决策树的叶节点数量, 它是模型拟合环境因子间交互作用的阶数。增强回归树中所有决策树的叶节点数量相同, 训练过程中叶节点达到一定数量时则停止生长, 不需要剪枝^[23-24]。学习效率决定了模型达到最优所需训练的时间, lr 值过小, 则收敛速度慢、训练时间越长; lr 值过大, 抽样时容易产生噪音, 导致函数平滑性降低、稳定性差^[25]。通常情况下, 迭代次数 (nt) 要达到 1 000 以上模型才趋于稳定, 树的复杂度 (tc) 1—16, 学习效率 (lr) 0.001—0.1, 抽样比率 (bf) 为 0.75, 函数损失形式为 “gaussian”。由于 tc 和 lr 的取值影响模型的预测准确度, 随机选择 70% 训练集数据用于构建模型, 剩余 30% 的数据用于计算模型的预测偏差, 根据模型预测偏差大小选择最优的 tc 值和 lr 值。

随着决策树数量的增加, 增强回归树模型的拟合

效果会越来越好, 但决策树数量过大会出现过拟合, 导致预测精度降低。本研究以 10 倍交叉验证法 (10-fold cross-validation) 确定最优决策树的数量。

1.3.3 预测因子相对重要性计算 在分类回归树模型中, FRIEDMAN^[26]提出了用 $I_j^2(T)$ 作为第 j 个预测因子 X_j 的相关性的度量, 该度量基于选择 X_j 变量进行决策树的节点分裂时平方误差加权改进, 该度量比其对应的单个分类树更加可靠。

$$I_j^2 = \frac{1}{M} \sum_{m=1}^M I_j^2(T_m) \quad (2)$$

式中, 预测因子 X_j 的相对重要性的平方 I_j^2 即为平方误差加权改进在模型中 M 棵分类回归树上的平均^[27]。通常情况下预测因子的相对重要性以百分数形式表示, 所有预测因子的相对重要性之和为 100。

1.4 模型预测准确度检验

采用最大误差参照法计算预测准确度^[28]:

$$R = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{|F_i - A_i|}{M_i} \right) \quad (3)$$

其中, R 为模型预测准确度, F_i 为模型预测病害流行等级, A_i 为实际病害流行等级, M_i 为第 i 次预测的最大参照误差。

2 结果

2.1 模型预测变量确定

通过对预测集正态性判断、误差独立性判断、线性判断、同方差性判断以及多元回归模型综合判断, 剔除离群值和强影响点, 筛选出影响海河平原小麦赤霉病发生的 8 个关键气象因子, 即 LT_{65} 、 MWS_{55} 、 MRH_{55} 、 $Rain_{35}$ 、 MT_{25} 、 SD_{15} 、 MRH_{15} 、 $DRain_{15}$ 。同时, 构建了小麦赤霉病多元线性回归预测模型 (multiple linear regression model, MLR I): $y = -13.2427 + 0.3145LT_{65} - 0.9824MWS_{55} + 0.1209MRH_{55} + 0.1377Rain_{35} - 0.4184MT_{25} + 0.0814SD_{15} + 0.28024MRH_{15} - 0.8832DRain_{15}$, 该模型 $R^2 = 0.8158$, 矫正 $R^2 = 0.8018$, $P < 2.2 \times 10^{-16}$ 。

其中, y 为小麦赤霉病病穗率, LT_{65} 为抽穗期初始日之前 26—30 d 最低温度, MWS_{55} 为抽穗期初始日之前 21—25 d 平均风速, MRH_{55} 为抽穗期初始日之前 21—25 d 相对平均湿度, $Rain_{35}$ 为抽穗期初始日之前 11—15 d 总降雨量, MT_{25} 为抽穗期初始日之前 6—10 d 平均温度, SD_{15} 为抽穗期初始日之后 1—5 d 总日照时数, MRH_{15} 为抽穗期初始日之后 1—5 d

相对平均湿度， $DRain_{15}$ 为抽穗期初始日之后 1—5 d 降雨天数。

2.2 预测模型拟合的关键影响因素

根据增强回归树 (BRT) 模型拟合曲线 (图 1) 可知, 在不同学习效率 (lr) 和树的复杂度 (tc) 下, 当 lr 为 0.1 和 0.05 时, 模型的最小预测偏差与其他学

习效率相比偏差相对较大, 在不同树的复杂度下, 模型会较早的发生过度拟合。当 lr 为 0.001 时, 模型迭代次数一般在 2 000 左右达到最小预测偏差, 当 lr 为 0.01 时, 模型迭代次数一般在 500—800 范围内达到最小预测偏差。当 lr 为 0.005 时, 模型迭代次数在 900—1 800 范围内达到最小预测偏差。

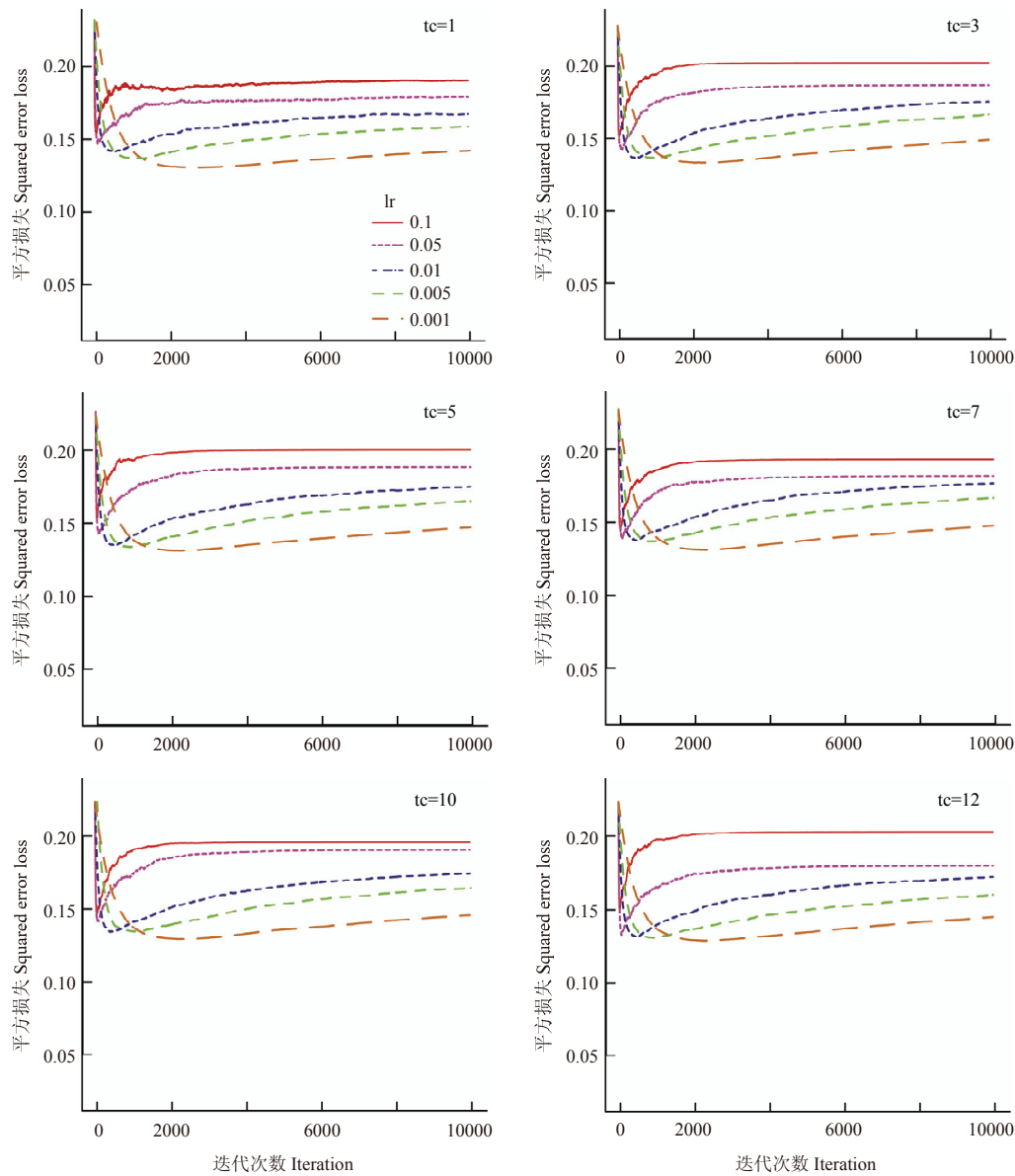


图 1 增强回归树模型的预测误差与决策树数量的关系

Fig. 1 Relationship between the prediction error of BRT model and the number of decision trees

设置lr为0.01、0.005, 由不同tc的残差标准误(residual standard error, 图2) 可知, 在lr为0.01和0.005的学习效率条件下, 当tc为6时增强回归树模型的残差标准误

分别为 0.01004 和 0.006311, 随着 tc 值的增加, 增强回归树模型的预测偏差相对变化不大。综合考虑不同 lr 和 tc 下模型预测偏差, 选择模型的 lr 为 0.005, tc 为 6。

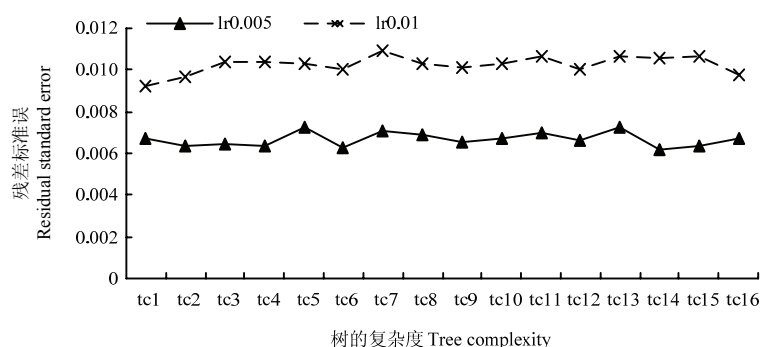


图2 不同树的复杂度下增强回归树模型的残差标准误

Fig. 2 Residual standard error of BRT model under the different tree complexities

2.3 预测变量重要性及对病害发生预测的影响

设置 lr 为 0.005, 在不同 tc 下预测了变量的重要性 (图 3), 随着 tc 增大, 各预测变量重要性排名未发生太大变化, MRH_{15} 、 $Rain_{35}$ 是相对重要的两个预测变量; 其次是 MRH_{55} 、 SD_{15} 、 LT_{65} 、 MT_{25} 、 MWS_{55} 、 $DRain_{15}$ 。

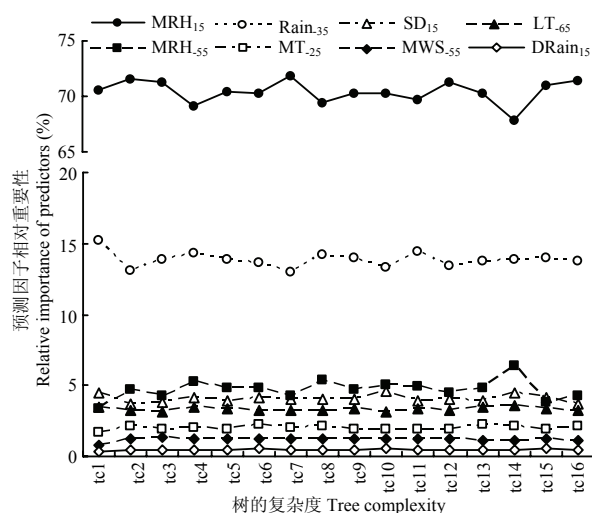


图3 不同树的复杂度下预测变量相对重要性

Fig. 3 Relative importance of the predictive variables under the different levels of tree complexity

当 $tc=6$ 时, 预测变量 MRH_{15} 、 $Rain_{35}$ 、 MRH_{55} 、 SD_{15} 、 LT_{65} 、 MWS_{55} 、 MT_{25} 、 $DRain_{15}$ 的重要性由高到低依次为 69.62%、14.08%、4.89%、4.34%、3.35%、2.02%、1.20%、0.50%。

设置 lr 为 0.005、 tc 为 6、抽样比率 (bf) 为 0.75、

函数损失形式为 “gaussian”, 交叉验证折数为 10 次、 $n.trees$ 为 5 000, 确定拟合最终的拟合模型, 模型各预测变量的反应曲线见图 4。

(1) 平均相对湿度对小麦赤霉病发生风险的影响 预测变量 MRH_{15} 对赤霉病发生风险的重要性最高, 为 69.62%。当其小于 46% 时, 其变化对赤霉病发生的影响较小; 当其在 46%—67% 时, 随其增加, 赤霉病发生风险迅速上升; 当其高于 67% 时, 其对赤霉病发生的促进作用趋于平稳。预测变量 MRH_{55} 对赤霉病发生风险的重要性居第 3 位, 为 4.89%。其对赤霉病发生风险是非线性关系, 当其在 35%—48% 时, 其对赤霉病发生风险的作用效果为先升后降; 当其在 48%—58% 时, 随其增加, 赤霉病发生的风险迅速上升; 当其高于 58% 时, 其对赤霉病发生的促进作用趋于平稳。

(2) 总降雨量对小麦赤霉病发生风险的影响 预测变量 $Rain_{35}$ 对赤霉病发生风险的重要性居第 2 位, 为 14.08%。赤霉病的发生风险随其增加而增加, 当其高于 16 mm 时, 其对赤霉病发生风险的促进作用趋于平稳。

(3) 日照时数对小麦赤霉病发生风险的影响 预测变量 SD_{15} 对赤霉病发生风险的重要性居第 4 位, 为 4.34%。模型显示当该时期日照时数在 25—49 h 时, 赤霉病发生风险迅速降低, 当其大于 50 h 时, 日照时数对赤霉病发生风险的抑制作用趋于平稳。

2.4 模型的优化

根据增强回归树模型筛选的 4 个重要的预测变量, 即 MRH_{15} 、 $Rain_{35}$ 、 MRH_{55} 、 SD_{15} , 简化多元线性回归模型 (MLR II): $y = -19.45376 + 0.11689MRH_{15} + 0.17346Rain_{35} + 0.04185SD_{15} + 0.26592MRH_{55}$, 该模型

$R^2=0.7575$, 矫正 $R^2=0.7468$, $P<2.2\times 10^{-16}$ 。

2.5 模型的验证

以 2008、2010、2012 年“同年份多点”的部分地区历史数据为测试集, 对多元线性回归模型、增强回归树模型预测结果进行验证, 观测值与预测值对比发

现, 预测结果与实际观测值基本相符, 但个别地区预测结果略有出入(图 5)。预测变量由 8 个简化为 4 个时, 多元线性回归模型预测准确度由 88.43%降至 85.90%, 增强回归树模型预测准确度由 87.72%升至 91.23%。

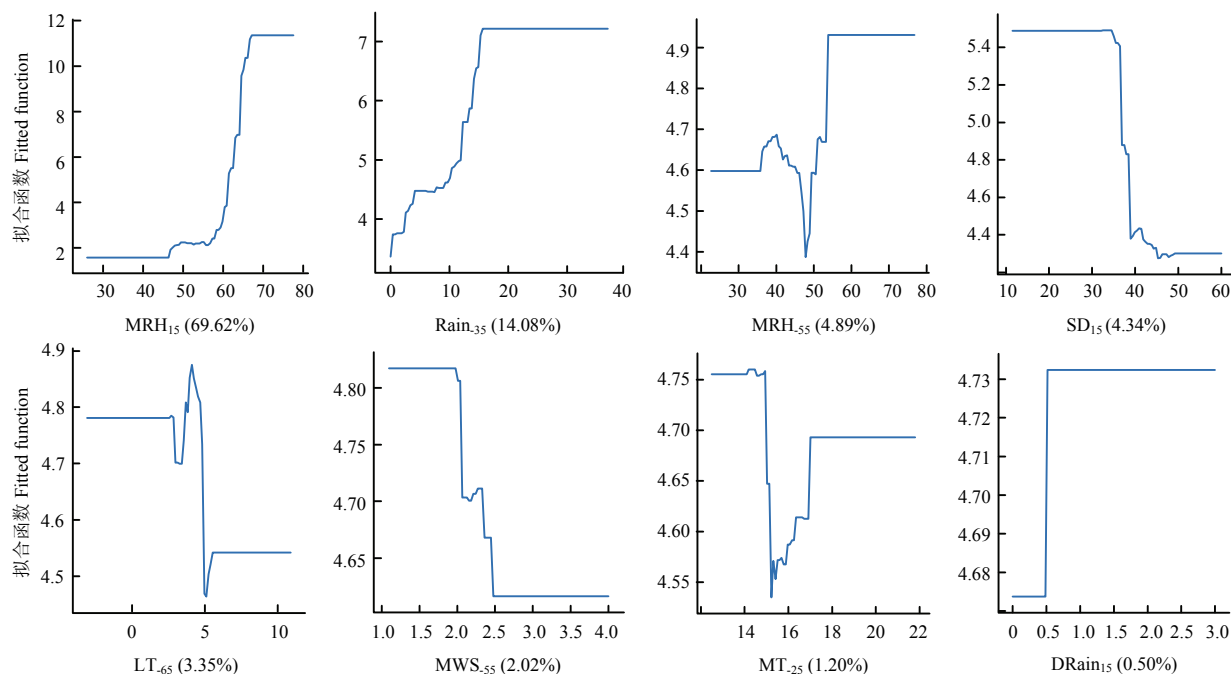


图 4 预测变量反应曲线

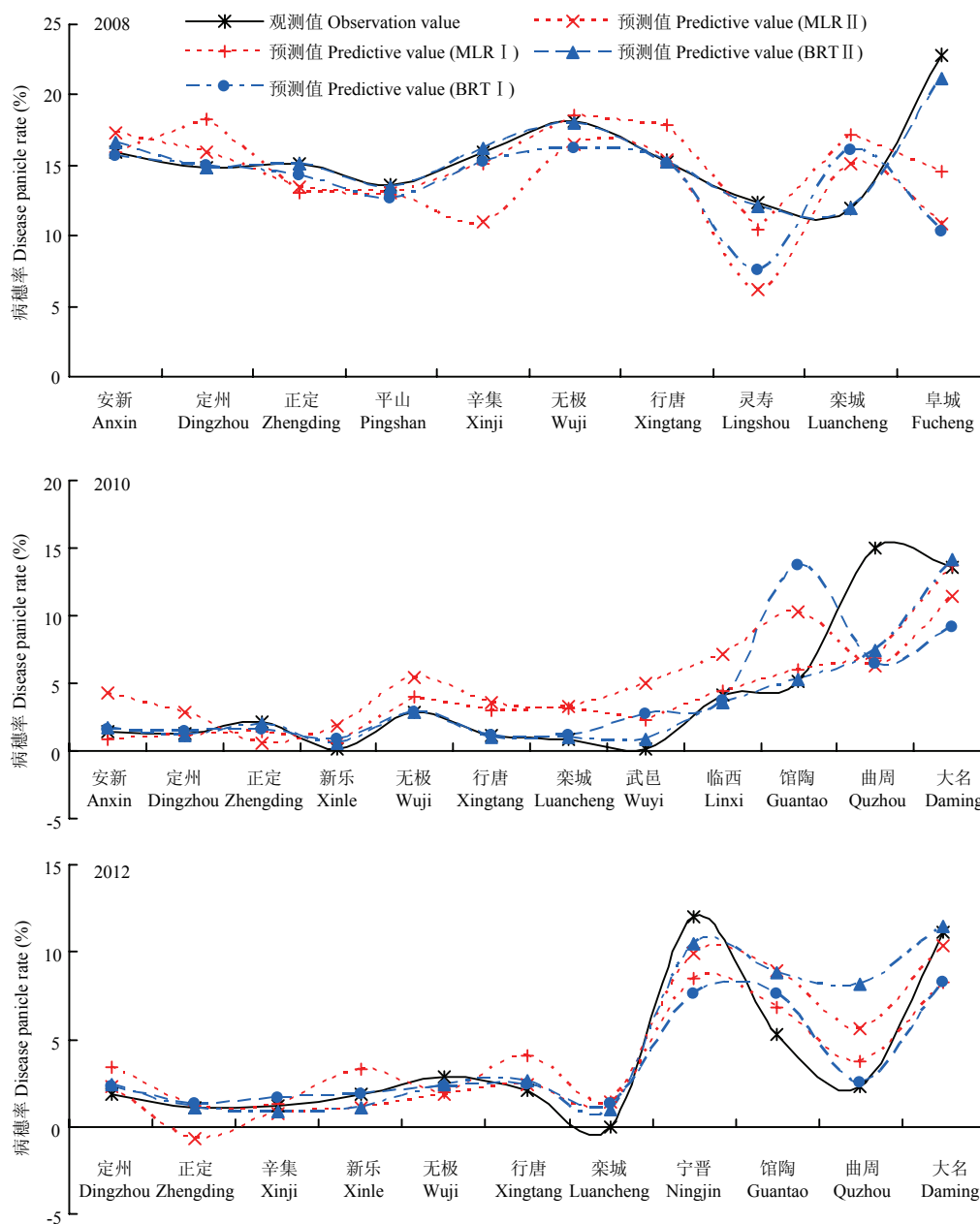
Fig. 4 Response curve of the predictive variables

以 2001—2016 年“多年份定点”的正定、栾城两地小麦赤霉病发生的历史数据为测试集, 对多元线性回归模型、增强回归树模型的预测结果进行验证, 预测结果与实际观测值曲线基本一致(图 6)。预测变量由 8 个简化为 4 个时, 两个模型预测准确度无显著变化, 多元线性回归模型预测准确度由 87.53%变为 87.42%, 增强回归树模型预测准确度由 89.20%变为 89.21%。

3 讨论

小麦赤霉病是典型的气候性病害^[1], 其预测模型受品种抗性、气象因素、田间菌源量等多因素影响, 具有典型的地域特异性。GIROUX 等评价了 9 种不同模型在加拿大魁北克地区预测小麦赤霉病发生或 DON 毒素含量的效果, 美国的两个模型(DE WOLF 等开发^[9-10])和阿根廷的模型(MOSCHINI 等开发^[29-30])预测效果优于其他模型^[31], 表明预测模型

应用存在一定的地域特异性。SHAH 等^[12]利用增强回归树模型预测了严重度大于 10% 的小麦赤霉病发生概率, 测试数据误识率与 logistic 回归模型相比下降 31%, 但该模型未能准确反映出病穗率和病害发生等级; LANDSCHOOT 等^[32]基于比例优势模型, 明确了比利时小麦赤霉病病情指数和 DON 毒素含量的关键参数及参数值, 并开发了 web 系统, 预测效果良好, 但该系统无法在其他地区应用; XU 等^[33]开发了基于 logistic 回归的欧洲小麦赤霉病 DON 毒素含量预测模型, 利用重采样和全子集回归分析表明, 通过气象因子预测 DON 毒素含量的“最佳”模型存在不唯一性; 基于气象因子的 DON 毒素含量预测模型具有可移植性或可替代性, 其通用性较差, 预测结果易出现假阴性。本实验室前期研究^[34]对安徽桐城小麦赤霉病预测模型中的气象因子进行了物候期数据的本地化修正, 构建了河北南部麦区小麦赤霉病预测模型, 历史数据



MLR I、MLR II 分别为含 8 个、4 个预测变量的多元线性回归模型；BRT I、BRT II 分别为含 8 个、4 个预测变量的增强回归树模型 MLR I and MLR II represent the multiple linear regression model with 8 and 4 predictive variables, respectively. BRT I and BRT II represent the boosted regression tree model with 8 and 4 predictive variables, respectively. 图 6 同 The same as Fig. 6

图 5 同年份多监测点的多元线性回归、增强回归树模型小麦赤霉病预测值与观测值对比

Fig. 5 Comparison between the observed and predicted values of multivariate linear regression and BRT models of FHB in multiple monitoring sites in the same year

验证模型准确度为 70.00%，与安徽桐城市小麦赤霉病病穗率的预测结果相比，其准确度相对较低。

本研究综合借鉴本领域专家的研究经验^[9,12,35-40]，以小麦物候期的抽穗期初始日为起点，每 5 d 编为一组，抽穗期初始日向前选择编 6 组，向后选择编 2 组，

以各组内最高温度、最低温度等 9 个气象因子为自变量，赤霉病病穗率为因变量，通过逐步回归分析，筛选出显著的变量，得到最优子集^[41]。观测样本中异常值对预测结果产生影响，利用 R 语言 car 包中的 outlierTest() 函数查找离群点、Cooke 距离判断强影响点，通过消

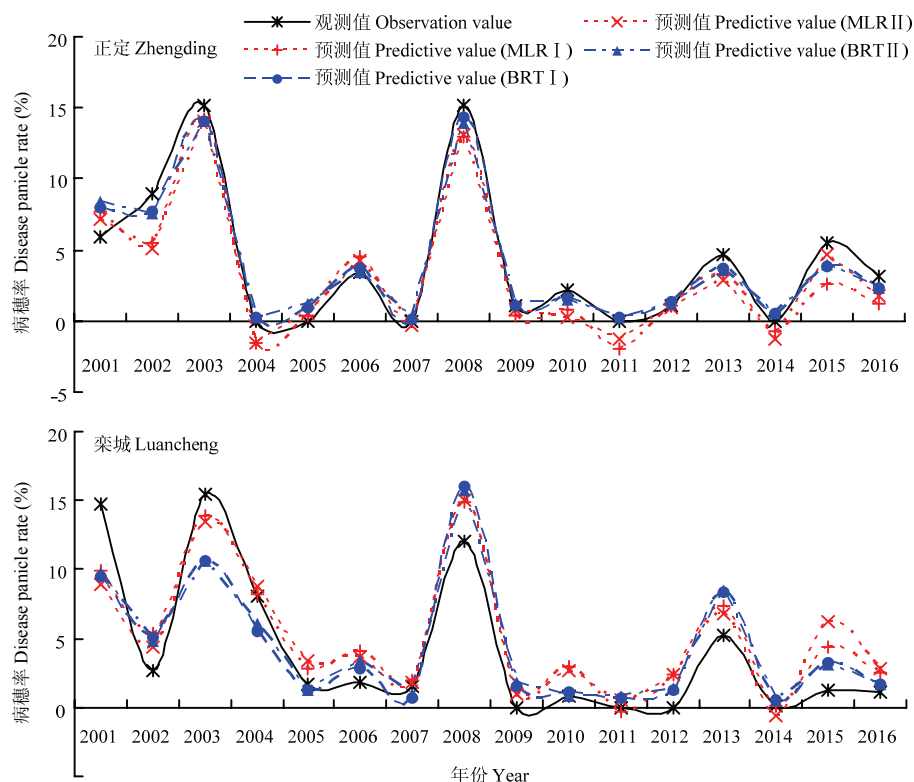


图 6 2001—2016 年定位监测点的多元线性回归、增强回归树模型小麦赤霉病病穗率观测值与预测值对比

Fig. 6 Comparison between the observed and predicted values of multivariate linear regression and BRT models on the diseased panicle rate of FHB in the location monitoring site from 2001 to 2016

除预测数据集中的离群点和强影响点,提升模型预测准确度。同时,通过对预测集正态性判断、误差独立性判断、线性判断、同方差性判断以及多元回归模型综合判断,证实了气象因子与病穗率的多元线性回归假设。SHAH 等^[12]研究表明,增强回归树模型预测效果优于 logistic 回归模型;本研究以“同年份多点”和“多年份定点”历史数据验证了多元线性回归模型和增强回归树模型预测结果,当预测变量由 8 个调减至 4 个时,多元线性回归模型预测准确度呈下降趋势,而增强回归树模型预测准确度呈上升趋势。利用历史数据验证模型预测效果时,2008 年阜城地区和 2010 年曲周地区病穗率的预测值与实际观测值偏差较大,其原因可能与本研究仅选择气象因子有关,赤霉病的发生与发展还与小麦品种抗性、田间管理措施等因素有关。在今后的研究中,还需进一步考虑小麦品种抗性、田间菌源量等因素,以期进一步提高预测准确度。

4 结论

研究构建了基于增强回归树的海河平原小麦赤霉

病病穗率预测模型,该模型含有 4 个预测变量。经两地 16 年历史数据验证,模型预测准确度为 89.21%,病穗率预测值与实际观测值的波动趋势基本一致。研究结果不仅为海河平原小麦赤霉病预测预报提供技术支持,也为小麦赤霉病预测模型优化和改进提供了参考。

参考文献 References

- [1] SAVARY S, WILLOCQUET L, PETHYBRIDGE S J, ESKER P, MCROBERTS N, NELSON A. The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*, 2019, 3(3): 430-439.
- [2] 陆维忠,程顺和,王裕中.小麦赤霉病研究.北京:科学出版社,2001.
LU W Z, CHENG S H, WANG Y Z. Study on Wheat Scab. Beijing: Science Press, 2001. (in Chinese)
- [3] GHIMIRE B, SAPKOTA S, BAHRI B A, MARTINEZ-ESPINOZA A D, BUCK J W, MERGOU M. Fusarium head blight and rust diseases in soft red winter wheat in the southeast United States: State

- of the art, challenges and future perspective for breeding. *Frontiers in Plant Science*, 2020, 11: 1080.
- [4] RUAN Y F, ZHANG W T, KNOX R E, BERRAIES S, CAMPBELL H L, RAGUPATHY R, BOYLE K, POLLEY B, HENRIQUEZ M A, BURT A, KUMAR S, CUTHBERT R D, FBOERT P R, BUERSTIMAYR H, DEPAUW R M. Characterization of the genetic architecture for Fusarium head blight resistance in durum wheat: The complex association of resistance, flowering time, and height genes. *Frontiers in Plant Science*, 2020, 11: 592064.
- [5] CHEN Y, KISTLER H C, MA Z. *Fusarium graminearum* trichothecene mycotoxins: Biosynthesis, regulation, and management. *Annual Review of Phytopathology*, 2019, 57: 15-39.
- [6] 胡文静, 张春梅, 吴迪, 陆成彬, 董亚超, 程晓明, 张勇, 高德荣. 长江中下游小麦抗赤霉病品种的筛选与部分农艺性状分析. *中国农业科学*, 2020, 53(21): 4313-4321.
- HU W J, ZHANG C M, WU D, LU C B, DONG Y C, CHENG X M, ZHANG Y, GAO D R. Screening for resistance to Fusarium head blight and agronomic traits of wheat germplasms from Yangtze river region. *Scientia Agricultura Sinica*, 2020, 53(21): 4313-4321. (in Chinese)
- [7] 张爱民, 阳文龙, 李欣, 孙家柱. 小麦抗赤霉病研究现状与展望. *遗传*, 2018, 40(10): 858-873.
- ZHANG A M, YANG W L, LI X, SUN J Z. Current status and perspective on research against Fusarium head blight in wheat. *Hereditas*, 2018, 40(10): 858-873. (in Chinese)
- [8] 黄冲, 姜玉英, 李春广. 1987年—2018年我国小麦主要病虫害发生危害及演变分析. *植物保护*, 2020, 46(6): 186-193.
- HUANG C, JIANG Y Y, LI C G. Occurrence, yield loss and dynamics of wheat diseases and insect pests in China from 1987 to 2018. *Plant Protection*, 2020, 46(6): 186-193. (in Chinese)
- [9] DE WOLF E D, MADDEN L V, LIPPS P E. Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. *Phytopathology*, 2003, 93(4): 428-435.
- [10] DE WOLF E D, ISARD S A. Disease cycle approach to plant disease prediction. *Annual Review of Phytopathology*, 2007, 45: 203-220.
- [11] SHAH D A, MOLINEROS J E, PAUL P A, WILLYERD K T, MADDEN L V, DE WOLF E D. Predicting Fusarium head blight epidemics with weather-driven pre- and post-anthesis logistic regression models. *Phytopathology*, 2013, 103(9): 906-919.
- [12] SHAH D A, DE WOLF E D, PAUL P A, MADDEN L V. Predicting Fusarium head blight epidemics with boosted regression trees. *Phytopathology*, 2014, 104(7): 702-714.
- [13] HOOKER D C, SCHAAFSMA A W, TAMBURIC-ILINCIC L. Using weather variables pre- and post-heading to predict deoxynivalenol content in winter wheat. *Plant Disease*, 2002, 86(6): 611-619.
- [14] DEL PONTE E M, FERNANDES J M C, PAVAN W. A risk infection simulation model for Fusarium head blight of wheat. *Fitopatologia Brasileira*, 2005, 30(6): 634-642.
- [15] ROSSI V, GIOSUE S, PATTORI E, SPANNAF, DEL VECCHIO A. A model estimating the risk of Fusarium head blight on wheat. *Bulletin OEPP/EPPO Bulletin*, 2003, 33(3): 421-425.
- [16] MUSA T, HECKER A, VOGELGSANG S, FORRER H R. Forecasting of Fusarium head blight and deoxynivalenol content in winter wheat with FusaProg. *Bulletin OEPP/EPPO Bulletin*, 2007, 37(2): 283-289.
- [17] ZHANG D Y, WANG D Y, GU C Y, JIN N, ZHAO H T, CHEN G, LIANG H Y, LIANG D. Using neural network to identify the severity of wheat Fusarium head blight in the field environment. *Remote Sensing*, 2019, 11(20): 2375.
- [18] 刘志红, 张雷, 燕亚菲, 周洁, 张正健, 张涵斌. BP 神经网络在小麦赤霉病气象预测中的应用. *云南农业大学学报*, 2010, 25(5): 680-685.
- LIU Z H, ZHANG L, YAN Y F, ZHOU J, ZHANG Z J, ZHANG H B. Application research on wheat scab forecasting based on BP neural network. *Journal of Yunnan Agricultural University*, 2010, 25(5): 680-685. (in Chinese)
- [19] HUANG L S, WU Z C, HUANG W J, MA H Q, ZHAO J L. Identification of Fusarium head blight in winter wheat ears based on Fisher's linear discriminant analysis and a support vector machine. *Applied Sciences*, 2019, 9(18): 3894.
- [20] LIU L Y, DONG Y Y, HUANG W J, DU X P, MA H Q. Monitoring wheat Fusarium head blight using unmanned aerial vehicle hyperspectral imagery. *Remote Sensing*, 2020, 12(22): 3811.
- [21] YOU M P, RENSING K, RENTON M, BARBETTI M J. Critical factors driving aphanomyces damping-off and root disease in clover revealed and explained using linear and generalized linear models and boosted regression trees. *Plant Pathology*, 2018, 67(6): 1374-1387.
- [22] 李月华, 杨利华. 河北省冬小麦高产节水节肥栽培技术. 北京: 中国农业出版社, 2017.
- LI Y H, YANG L H. Cultivation techniques of high yield, water saving and fertilizer saving in winter wheat in Hebei Province. Beijing: China Agriculture Press, 2017. (in Chinese)
- [23] ELITH J, LEATHWICK J R, HADTIE T. A working guide to boosted regression trees. *Journal of Animal Ecology*, 2008, 77(4): 802-813.
- [24] DEATH G. Boosted trees for ecological modeling and prediction. *Ecology*, 2007, 88(1): 243-251.

- [25] ZHANG Y, CHEN H Y H, REICH P B. Forest productivity increases with evenness, species richness and trait variation: A global meta-analysis. *Journal of Ecology*, 2012, 100(3): 742-749.
- [26] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
- [27] FRIEDMAN J H, MEULMAN J J. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 2003, 22(9): 1365-1381.
- [28] 肖悦岩. 预测预报准确度评价方法的研究. *植保技术与推广*, 1997, 17(4): 3-6.
- XIAO Y Y. The studies on the reliability evaluation of pest prediction and forecasting. *Plant Protection Technology and Extension*, 1997, 17(4): 3-6. (in Chinese)
- [29] MOSCHINI R C, FORTUGNO C. Predicting wheat head blight incidence using models based on meteorological factors in Pergamino, Argentina. *European Journal of Plant Pathology*, 1996, 102(3): 211-218.
- [30] MOSCHINI R C, PIOLI R, CARMONA M, SACCHI O. Empirical predictions of wheat head blight in the Northern Argentinean Pampas Region. *Crop Science*, 2001, 41(5): 1541-1545.
- [31] GIROUX M E, BOURGEOIS G, DION Y, RIOUX S, PAGEAU D, ZOGHLAMI S, PARENT C, VACHON E, VANASSE A. Evaluation of forecasting models for *Fusarium* head blight of wheat under growing conditions of Quebec, Canada. *Plant Disease*, 2016, 100(6): 1192-1201.
- [32] LANDSCHOOT S, WAEGEMAN W, AUDENAERT K, VAN DAMME P, VANDEPITTE J, DE BAETS B, HAESAERT G. A field-specific web tool for the prediction of *Fusarium* head blight and deoxynivalenol content in Belgium. *Computers and Electronics in Agriculture*, 2013, 93: 140-148.
- [33] XU X M, MADDEN L V, EDWARDS S G, DOOHAN F M, MORETTI A, HORNOK L, NICHOLSON P, RITIENI A. Developing logistic models to relate the accumulation of DON associated with *Fusarium* head blight to climatic conditions in Europe. *European Journal of Plant Pathology*, 2013, 137(4): 689-706.
- [34] 张雪娇. 冀中南小麦赤霉病发生趋势气象预测模型及信息发布平台构建[D]. 保定: 河北农业大学, 2017.
- ZHANG X J. The construction of meteorological forecasting model of wheat head blight occurring trend and the information release platform in the mid-south of Hebei Province[D]. Baoding: Hebei Agricultural University, 2017. (in Chinese)
- [35] SHAH D A, PAUL P A, DE WOLF E D, MADDEN L V. Predicting plant disease epidemics from functionally represented weather series. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, 2019, 374(1775): 20180273.
- [36] GUO X W. Development of models to predict *Fusarium* head blight disease and deoxynivalenol in wheat, and genetic causes for chemotype diversity and shifting of *Fusarium graminearum* in Manitoba. Canada: University of Manitoba, 2008.
- [37] KRISS A B, PAUL P A, MADDEN L V. Characterizing heterogeneity of disease incidence in a spatial hierarchy: A case study from a decade of observations of *Fusarium* head blight of wheat. *Phytopathology*, 2012, 102(9): 867-877.
- [38] 范仰东, 张正书, 廖薇薇. 小麦赤霉病流行程度中长期预报探讨. *植物保护*, 1985, 11(4): 4-6.
- FAN Y D, ZHANG Z S, LIAO W W. Discussion on medium and long term forecast of wheat scab epidemic degree. *Plant Protection*, 1985, 11(4): 4-6. (in Chinese)
- [39] 周崇和. 小麦赤霉病流行因子的主成分分析及预测模型探讨. *植物保护学报*, 1990, 17(4): 317-321.
- ZHOU C H. Principal component in epidemic factors of wheat scab and a forecasting model. *Acta Phytophylacica Sinica*, 1990, 17(4): 317-321. (in Chinese)
- [40] 左豫虎, 郑莲枝, 张匀华, 刘惕若, 彭驰, 张运权, 史翠萍, 许群洲, 潘春彦. 黑龙江省春小麦赤霉病流行的预测方法. *植物保护学报*, 1995, 22(4): 297-302.
- ZUO Y H, ZHENG L Z, ZHANG J H, LIU T R, PENG C, ZHANG Y Q, SHI C P, XU Q Z, PAN C Y. Studies on the epidemic method of spring wheat scab in Heilongjiang Province. *Acta Phytophylacica Sinica*, 1995, 22(4): 297-302. (in Chinese)
- [41] 游士兵, 严研. 逐步回归分析法及其应用. *统计与决策*, 2017(14): 31-35.
- YOU S B, YAN Y. Stepwise regression analysis and its application. *Statistics and Decision*, 2017(14): 31-35. (in Chinese)

(责任编辑 岳梅)