

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.

利用机器学习预测华北地区冬小麦农田氮淋失

李涛宇,许秀春,杨 轩,崔 斌,陈恒爱,赵晓莹,袁 宁,孟凡乔* (中国农业大学资源与环境学院农田土壤污染防治与修复北京市重点实验室,北京 100193)

摘要: 为明确华北地区冬小麦生产中的氮淋失特征及影响因素,整合了 2000~2023 年发表的、在华北地区进行的冬小麦氮淋失试验研究,分别构建了单因素和多因素回归模型,以及基于机器学习的随机森林和 XGBoost 模型,对农田氮淋失进行定量分析.研究表明,氮肥和水分投入量、土壤性质(pH 值、粘粒和砂粒含量)是影响华北地区冬小麦生产中氮淋失的主要因素,秸秆还田能够有效抑制氮淋失.利用随机森林方法,基于全部预测变量和逐步回归筛选后预测变量构建的模型拟合优度分别为 0.628 和 0.708,而经过贝叶斯搜索和早停法优化后构建的 XGBoost 模型拟合优度分别为 0.745 和 0.722.机器学习模型构建过程中充分考虑了多因素对冬小麦生产中氮淋失的影响,采用经验统计筛选的预测指标构建的模型拟合优度高,适用性好,对于华北地区冬小麦农田氮淋失的阻控具有重要技术支撑.

关键词: 冬小麦; 氮淋失; 机器学习; XGBoost; 网格搜索; 贝叶斯搜索

中图分类号: X592 文献标识码: A 文章编号: 1000-6923(2025)01-0343-12

Prediction of nitrogen leaching from winter-wheat production in North China based on random forest and XGBoost. LI Tao-yu, XU Xiu-chun, YANG Xuan, CUI Bin, CHEN Heng-ai, ZHAO Xiao-ying, YUAN Ning, MENG Fan-qiao* (College of Resources and Environmental Sciences, Beijing Key Laboratory of Prevention, Control and Restoration of Farmland Soil Pollution, China Agricultural University, Beijing 100193, China). *China Environmental Science*, 2025,45(1): 343-354

Abstract: Winter wheat is the main cereal crop in Northern China, and excessive nitrogen (N) fertilization and irrigation are employed in recent years to ensure a high grain yield. A high amount of N was lost via leaching, which exacerbated the risk of non-point source pollution and further increased the resource waste. It is highly necessary to clarify the characteristics and influencing factors of N leaching loss during the winter wheat season in the region. The literature on N leaching loss from the winter wheat production in Northern China, published from 2000 to 2023 was screened, and linear, multiple-factor regression models, as well as random forest and XGBoost models were established in the study. The research showed that the N leaching during the winter wheat season was mainly affected by the fertilizer rate and irrigation water, as well as soil properties (pH, clay and sand content), and was effectively inhibited by crop straw incorporation. Multiple variable combinations were constructed based on the results of importance analysis and stepwise regression. Grid search method, Bayesian and the combination of Bayesian and Early stopping were used to optimize the model parameters. The models constructed through Random Forest, based on all influential variables and based on the influential variables screened by stepwise, had the R^2 of 0.628 and 0.708, respectively, and the R^2 for the corresponding models constructed through XGBoost were 0.745 and 0.722, respectively. This indicates that the prediction effects of N leaching based on Random Forest and XGBoost were much better than the linear and multiple-factor regression models. The influence of multiple factors on N leaching was comprehensively considered in the machine learning models, and the effects of prediction were better when choosing the influential variables screened by empirical statistical methods as independent variables. The results of this study can provide technical support for reducing the N leaching in the winter wheat production in Northern China.

Key words: winter wheat; nitrogen leaching; machine learning; XGBoost; grid search; bayesian optimization

华北地区是我国主要的粮食产区,冬小麦是该地区主要的粮食作物,种植面积约 $3.3 \times 10^6 \text{hm}^2$ ^[1].近年来,随着人口增加和生活水平的提高,粮食产量需求持续增长,高量水肥投入成为提高冬小麦产量的主要措施^[2].当前,在华北地区冬小麦生产中,农民常规管理措施情形下,单季灌溉水和氮肥投入总量高

达 350mm 和 300kg/hm^2 ^[3-6],远超冬小麦的合理水肥需求量.过量施用的氮肥一部分在土壤中残留,很大一部分随水分迁移并淋失到作物根区以外、进入地

收稿日期: 2024-05-30

基金项目: 十四五国家重点研发计划项目(2023YFD1701701)

* 责任作者, 教授, mengfq@cau.edu.cn

下水体^[7],造成面源污染等环境问题^[8-9].华北地区冬小麦生产中硝态氮淋失量达 26.3~70.5kg/hm²,占化肥投入量的 11.7%~23.5%左右^[10-12].明确华北地区冬小麦生产中的氮淋失特征并进行定量预测,对于提高农业生产效益和保护环境具有重要意义.

研究表明,华北地区冬小麦季氮淋失受到多种因素的影响,包括氮肥施用^[13]、灌溉制度^[14]、土壤类型^[15]、秸秆还田^[16]等,目前大部分进行的氮淋失预测模型主要关注单个因素,如肥料或者水分对氮淋失的影响^[17].Cui 等^[18]和杨旺鑫^[19]分别建立了氮淋失量与施氮量和降雨量的指数和线性模型,回归决定系数(R^2)分别为 0.50 和 0.37.生产实际中,由于不同地区土壤和环境条件差异较大,农田生产管理措施各异,氮淋失的变化受到众多因素影响,而不是仅局限于肥料或者水分,单因素预测往往效果较差^[17],多因素、非线性预测模型更能反映生产实际中氮淋失的变化规律.

机器学习作为一种新兴的数据处理与统计方法,近年来在环境科学领域展现出其强大的应用潜力^[20].研究表明,广义加线性模型^[21]、人工神经网络^[22]以及随机森林模型^[23]等机器学习方法在环境污染预测中具有较高的准确性(R^2 均大于 0.76),能够有效揭示污染物变化规律及其影响因素,明确缓解环境污染问题的有效措施.机器学习方法,主要是基于概率论和统计学,从大规模数据中挖掘出隐藏的规律和关联,并利用这些规律为指标定量预测提供更加精准的模型^[24].机器学习作为一种回归方法,与统计回归分析具有相似之处,但是与经验统计回归模型相比,能够纳入更多的影响因子,并综合考虑不同影响因子的重要性程度,提高模型的预测精度.Ying 等^[25]和 Liu 等^[26]通过随机森林模型对农田氮淋失进行预测,结果表明其拟合效果要优于常规指数模型,且施用生物炭能够显著降低农田土壤氮淋失量.但综合分析,基于机器学习的农业面源污染研究多集中于大气^[21]、重金属^[23]、水质^[27]等方面,在氮淋失方面相关研究较少.

对于华北地区农田土壤氮淋失进行直接测定的方法在 2000 年以前尚不完善^[28],而对氮淋失的研究大多通过采集土壤样品或在室内进行模拟试验的方法进行^[29].因此,本研究整合了 2000~2023 年发表的、在华北地区进行的冬小麦农田氮淋失研究(表

S1),筛选包括土壤性质、氮肥施用量、秸秆还田量、水分投入量和氮淋失量等数据.通过随机森林、极限梯度提升(XGBoost)等机器学习方法,并结合网格搜索、贝叶斯搜索、早停法等进行参数优化,系统分析华北地区冬小麦生产中环境条件和农田管理措施对氮淋失的影响及其重要程度.根据不同影响因子的重要程度进行逐步筛选,并结合经验统计分析方法获得构建最佳模型的影响因子组合,以提高对氮淋失预测的模型精度,同时分析机器学习方法的适用性,为降低华北地区氮淋失风险和提提高氮肥利用率提供科学依据.

1 材料与方法

1.1 研究区域

传统意义上,华北地区包括北京、河北、山西、内蒙古和天津等省市^[1].与华北地区接壤的河南、山东和陕西等省份,当前大田作物主要也是冬小麦-夏玉米一年两熟的种植制度,气温、降水、日照和土壤等环境条件以及农田管理措施与河北、山西等地相似.因此,本研究区域包括北京、河北、山西、内蒙古、天津、河南、山东、陕西等省市地区.

1.2 数据来源

本次研究,主要对 Web of Science 和中国知网(CNKI)等文献平台检索了 2000~2023 年间发表的,且研究区域为华北地区的相关文献.搜索关键词包括“氮淋失”“氮素平衡”“硝态氮”“冬小麦”和“华北地区”等以及相应英文,并将这些关键词进行组合.筛选文献时,研究区域为河南、山东、陕西等省市地区的文献也被考虑在内.文献选择标准如下:1)研究应为在农田进行的田间试验,氮淋失应是通过淋溶盘法、渗漏池法或陶土头配合压力计法的其中一个方法直接测定;2)氮淋失监测时段必须包括一个完整的冬小麦生长周期;3)不同文献,如果数据来源于同一个地点、同一试验且相同年份的,选择最后一个发表年度的文献数据.

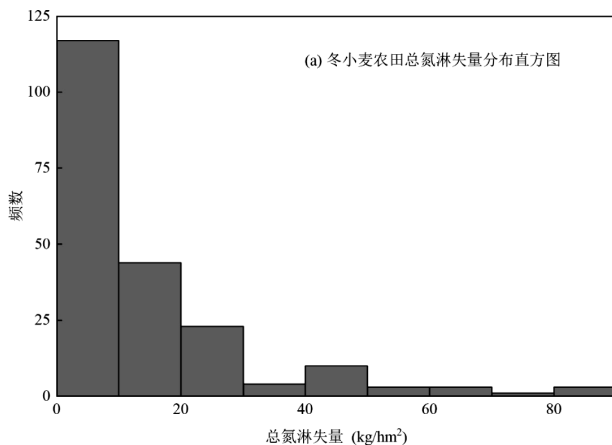
研究收集的数据主要包括土壤性质,包括 pH 值、有机质、全氮、粘粒、砂粒以及土壤类型,采样(采样方法、采样深度),农田管理措施,包括化肥施氮量、还田秸秆量、秸秆还田输入氮量、有机肥输入氮量、总施氮量、水分投入量等,以及总氮淋失量.对于只有图的文献,使用 GetData Graph Digitizer 软

件从图中提取数值.

本研究共选取了 22 篇期刊文献和硕士、博士论文,共获得 208 条数据.选取土壤性质、采样、农田管理措施等共 14 个因素作为预测变量,将总氮淋失量作为目标变量.预测变量中的总氮施用量包括化肥施氮量、秸秆还田输入氮量和有机肥输入氮量,水分投入量包括降雨量和灌溉量.

1.3 数据预处理

1.3.1 缺失值处理 本研究中缺失值主要集中在土壤粘粒、砂粒和总氮淋失量等数据.对于土壤粘粒和砂粒含量缺失的文献,查找与缺失文献试验区域具有相同经、纬度的文献中的相关数据进行补充.



硝态氮是华北地区总氮淋失的主要形式^[30],对于仅提供硝态氮淋失量的文献,其总氮淋失量按照硝态氮淋失占比为 95%的比例进行换算.

1.3.2 对数化处理 研究获得的 208 条数据中,冬小麦季总氮淋失量小于 20kg/hm² 的数据条约占 78% (图 1(a)).总氮淋失量的最小值、最大值和平均值分别为 0.10,88.5 和 13.9kg/hm².由于总氮淋失量分布偏离正态分布,对总氮淋失量进行取对数处理(LN),使数据更接近正态分布(图 1(b)),同时也减少了异常值的影响.对数化处理后的总氮淋失量仅用于机器学习模型的构建,进行其它回归分析时,未对总氮淋失量进行对数化处理.

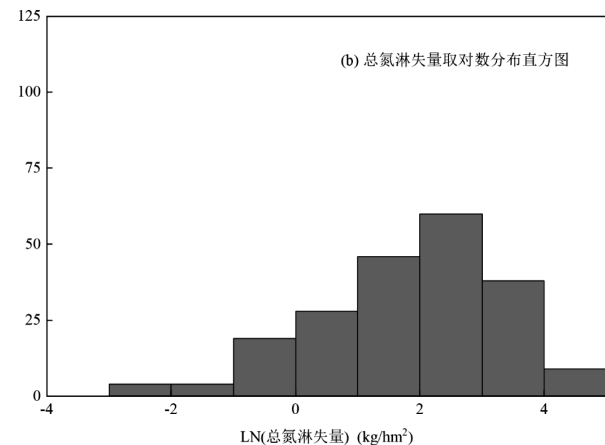


图 1 冬小麦农田总氮淋失量分布直方图和总氮淋失量取对数分布直方图

Fig.1 Distribution histogram of total N leaching loss and logarithmic total N leaching loss during the winter wheat season

1.3.3 标准化处理 为消除不同变量之间量纲的影响,研究中采用了标准化处理方法.标准化后的数据均值为 0,标准差为 1.标准化的过程采用 Z-score 标准化公式^[31]:

$$Z_i = \frac{(X_i - \mu)}{\sigma} \quad (1)$$

式中: Z_i 是标准化后的数据; X_i 为需要标准化的数据; μ 是该组数据的均值; σ 是该组数据的标准差.

1.3.4 独热编码处理 针对本研究中的土壤类型、采样方法等分类变量,本文采用了独热编码进行处理,将这些分类变量按照类别转化为 0 和 1 的编码形式.对于土壤类型,本文建立了三维向量,其中[1,0,0]代表棕壤土,[0,1,0]代表褐土,[0,0,1]代表潮土.对于采样方法,[1,0,0]代表陶土头-张力计法,[0,1,0]代表蒸渗仪法,[0,0,1]代表淋溶盘法.该编码处理方法解

决了因常规编码引起的分类变量在回归时权重发生变化问题,也避免模型对分类变量产生偏好,从而使解释和分析变量时更为准确和可靠.

1.4 相关性分析与回归分析

本研究采用 SPSS 27.0.1 进行 Pearson 相关性分析,一元线性回归分析、二元线性回归分析和多元回归分析,来明确施氮量、水分投入量和土壤性质等因子对总氮淋失量的影响.通过回归方程的 R^2 和 P 值来识别最佳模型.采用 Origin 2021 作图.

1.5 基于机器学习算法的模型建立

本研究采用 Python(Python version 3.10)相关库,包括 pandas(2.2.0)、numpy(1.26.3)、scikit-learn(1.4.0)和 xgboost(2.0.3))进行随机森林和 XGBoost 的模型建立.搜索方法采用 scikit-learn(1.4.0)中的 GridSearchCV 方法进行网格搜索,scikit-optimize

(0.9.0)库中的 BayesSearchCV 方法进行贝叶斯搜索.模型构建中均采用五折交叉检验的方法来评估参数,并采用 scikit-learn(1.4.0)库中的相关方法($r2_score$)计算 R^2 进行评价.

1.5.1 随机森林回归算法 随机森林是一种集成学习方法,在建立回归模型过程中通过构建多个决策树进行预测,每个决策树输出一个连续的数值,然后将决策树预测的结果,通过取平均值或加权平均值的方法整合得到最终结果^[32].通过随机森林进行模拟的过程中需要对参数(决策树的数量、树的深度、节点最小分裂样本数以及决策树考虑的特征数量等)进行优化以获得最优模型,参数优化方法包括网格搜索法和贝叶斯优化法.网格搜索是一种通过在预定义参数空间中穷举组合来确定最优模型参数的方法^[33].对于每一组参数组合,模型都会进行训练和评估,最终选择在指定评估指标表现最好的参数组合.而贝叶斯优化是一种基于贝叶斯定理的参数优化方法,通过在每一步选择具有最大潜在改善的参数组合,逐步缩小搜索空间,从而高效地寻找全局最优解^[33].

在建立随机森林模型的过程中,将预处理后的数据分为 80% 的训练集和 20% 的测试集.采用网格搜索法和贝叶斯搜索法在训练集上搜索最优参数组合,建立随机森林模型.为确保结果的稳定性,本文进行 200 次随机模拟.随后,在测试集上对建立好的模型进行了测试和评估,取均方误差和解释率的平均值趋于稳定的结果,并对所有模拟的模型拟合优度求平均作为最终结果.

1.5.2 XGBoost 回归算法 XGBoost 是一种梯度提升算法,通过迭代训练一系列弱学习器(本研究中为决策树),每次迭代都调整前一轮模型的错误,以逐步提高整体模型的性能^[32].该算法引入正则化项以避免过拟合,并具有高度优化的计算效率.

在 XGBoost 模型的建立过程中,将数据分为 80% 的训练集和 20% 的测试集.利用贝叶斯搜索法对决策树数量、树深度、决策树考虑的特征数量、学习率等 XGBoost 相关超参数进行优化并建立模型.取 200 次模型模拟 R^2 的平均值作为最终结果.为避免模型在拟合过程中出现过拟合的现象,采用了早停法.通过在训练过程中监控验证集的性能,并在性能不再提升时停止模型的训练,从而抑制了过拟

合的现象.

2 结果与分析

2.1 相关性分析和重要性分析

Pearson 相关性分析可用于评估变量之间的线性相关性,而随机森林或 XGBoost 重要性分析则可以考虑到预测变量与目标变量间的相对非线性关系.基于 Pearson 相关性分析和随机森林或 XGBoost 重要性分析,可以对预测变量进行筛选,以确定其对目标变量的影响程度和重要性.

在 Pearson 相关性分析中(图 2),总氮淋失量与农田主要管理措施(总施氮量、水分投入量、化肥施氮量)和土壤 pH 值具有极显著的正相关关系,而与土壤有机质含量则呈极显著的负相关关系.总氮淋失量与其他预测变量的相关性较弱,未达显著水平.需要指出的是总氮淋失量与还田秸秆量和秸秆还田输入氮量呈负相关关系.

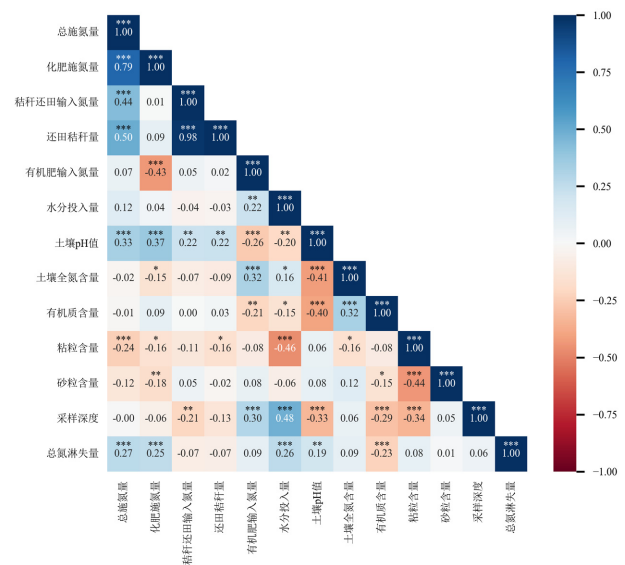


图2 冬小麦农田总氮淋失量与预测变量的相关性分析

Fig.2 Correlation between total N leaching loss and predictive variables during the winter wheat season

通过构建随机森林或 XGBoost 模型,可以量化每个预测变量对模型预测性能的贡献程度,从而识别出对目标变量影响较大的指标.随机森林重要性分析结果表明水分投入量、总施氮量、化肥施氮量、土壤粘粒含量和砂粒含量对总氮淋失量影响最大(图 3),而 XGBoost 重要性分析则表明土壤砂粒含量、采样方法和化肥施氮量是影响总氮淋失量最重要的因素.

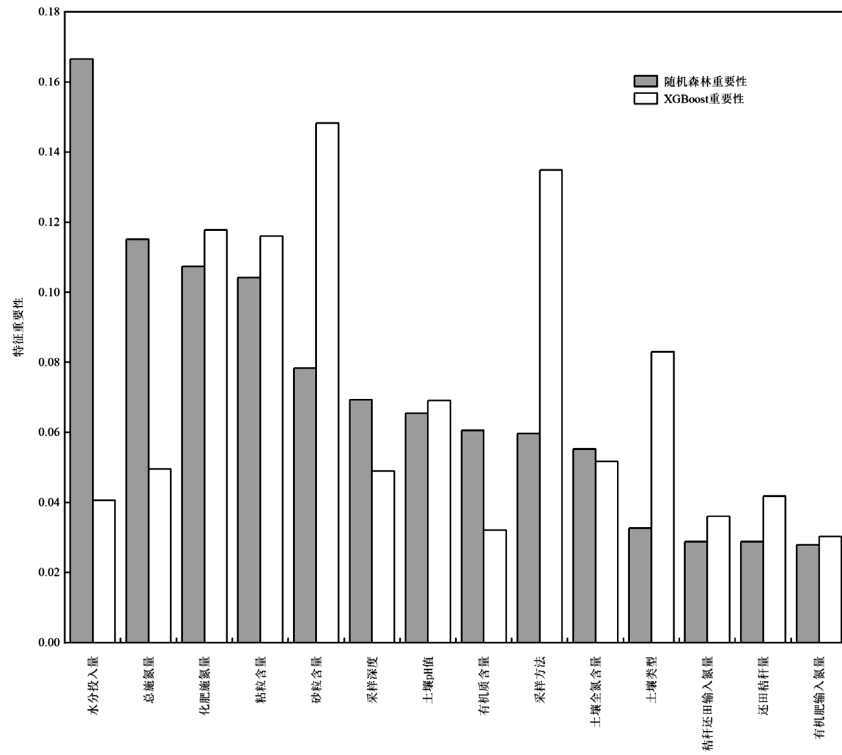


图3 冬小麦农田总氮淋失量各预测变量的重要性分析

Fig.3 Importance analysis of various predictive variables for total N leaching loss during the winter wheat season

2.2 常规回归模型建立

2.2.1 单因素和双因素回归分析 基于 Pearson 相关性分析和随机森林重要性分析结果,构建总氮淋

失量与施氮量 and 水分投入量的单因素(图 4)和双因素(图 5)回归模型,从而量化化肥施氮量、总施氮量以及水分投入量对总氮淋失量的影响。

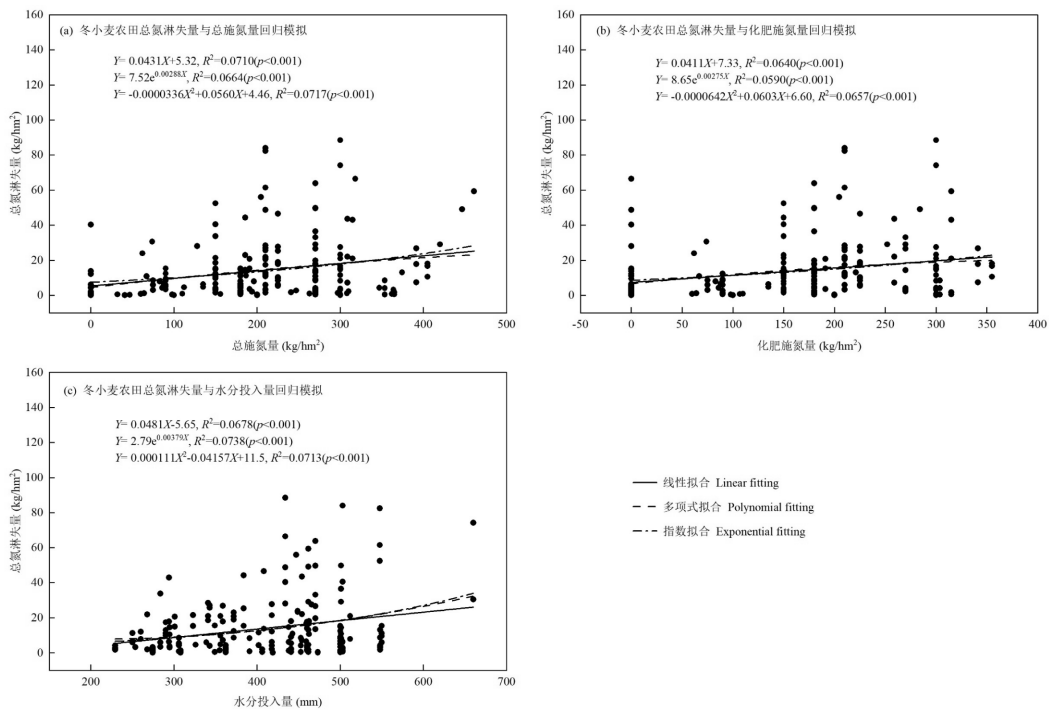


图4 冬小麦农田总氮淋失量与化肥施氮量、总施氮量和水分投入量的回归模拟

Fig.4 Regression simulation of total N leaching loss with fertilizer N rate, total N rate and water input in winter wheat fields

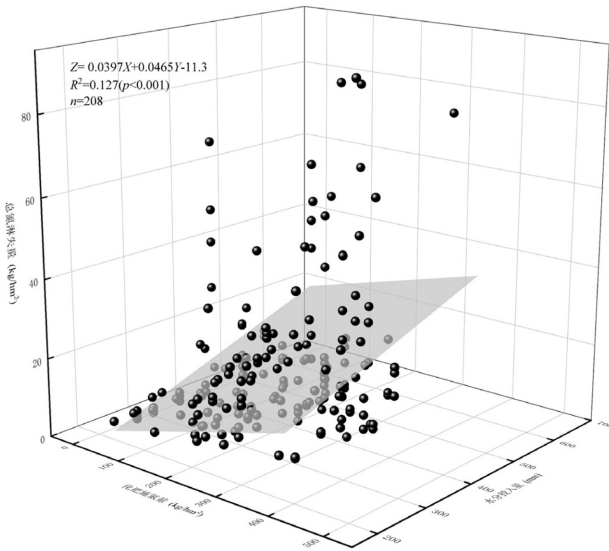


图 5 冬小麦农田总氮淋失量与化肥施氮量和水投入量的双因素回归模拟

Fig.5 Two factor regression simulation of total N leaching loss with fertilizer N rate and water input during the winter wheat season

线性回归表明,总氮淋失量随着氮肥和水投入量的增加呈增加的趋势(图 4)。通过总施氮量(图 4(a))和化肥施氮量(图 4(b))对总氮淋失量进行

预测的各模型拟合优度相似,且通过总施氮量进行预测的模型 R^2 (线性模型、指数模型、多项式模型分别为 0.071、0.066、0.072)略高于化肥施氮量,但解释度都小于 10%。水分投入量预测总氮淋失量的 3 种模型(线性模型、指数模型、多项式模型)拟合优度相差不大(图 4(c)),解释度也都仅为 7%左右。

在双因素回归分析中,化肥施氮量和水投入量对总氮淋失量的解释度为 12.7%,回归方程为 $Y_{\text{总氮淋失量}} = 0.0397X_{\text{化肥施氮量}} + 0.0465X_{\text{水分投入量}} - 11.3$, $P < 0.001$,在拟合优度上相比单因素回归模型有了一定的提升。化肥施氮量和水投入量标准化后的回归系数分别为 0.244 和 0.252,水分投入量对总氮淋失量的影响大于化肥施氮量。

2.2.2 多元回归分析 单因素和双因素回归分析的 R^2 值相对较低($P < 0.01$),表明单一因素对总氮淋失量模型的贡献仍有限,无法充分解释总氮淋失量的变化。因此,为更准确地预测和解释总氮淋失量,本文构建了所有预测变量与总氮淋失量的多元回归模型(表 1)。

表 1 冬小麦农田总氮淋失量的多元回归模型参数

Table 1 Parameters of multiple regression model for total N leaching loss in winter wheat fields

参数	系数预测值	标准差	t 值	P 值
截距 Intercept	-113.721	35.454	-3.208	0.002
化肥施氮量 Fertilizer N rate	0.066	0.012	5.341	<0.001
秸秆还田输入氮量 Straw incorporation N rate	-0.022	0.025	-0.871	0.385
有机肥输入氮量 Organic fertilizer N rate	0.059	0.025	2.338	0.020
水分投入量 Water input	0.085	0.015	5.874	<0.001
土壤 pH 值 Soil pH	6.051	3.364	1.799	0.074
土壤全氮含量 Soil total N	12.583	6.015	2.092	0.038
有机质含量 Soil organic matter	-0.340	0.431	-0.789	0.431
粘粒含量 Clay content	0.892	0.165	5.408	<0.001
砂粒含量 Sand content	0.314	0.087	3.605	<0.001
采样深度 Sampling depth	-0.010	0.032	-0.322	0.747
还田秸秆量 Straw incorporation	—	—	—	—
总施氮量 Total N rate	—	—	—	—
F 值 F statistic: 10.311				P 值 P-value: 6.43×10^{-14}
拟合优度 Multiple R^2 : 0.344				修正拟合优度 Adjusted R^2 : 0.310

多元回归模型对总氮淋失量的预测和解释的拟合优度($R^2=0.310, P < 0.001$)较单因素和双因素回归模型高,且达到统计学显著水平。构建的模型方程为 $Y_{\text{总氮淋失量}} = 0.066X_{\text{化肥施氮量}} - 0.022X_{\text{秸秆还田输入氮量}} + 0.059X_{\text{有机肥输入氮量}} + 0.085X_{\text{水分投入量}} + 6.051X_{\text{土壤 pH 值}} + 12.583X_{\text{土壤全氮含量}} - 0.340X_{\text{有机质含量}} + 0.892X_{\text{粘粒含量}} + 0.314X_{\text{砂粒含量}} - 0.010X_{\text{采样深度}} -$

113.721 。由于总施氮量(包括化肥施氮量、秸秆还田输入氮量和有机肥输入氮量)与各施氮量之间、还田秸秆量与秸秆还田输入氮量之间存在共线性,在多元回归模型的拟合中去除了总施氮量和还田秸秆量两个预测变量。回归结果表明,总氮淋失量与秸秆还田输入氮量、土壤有机质含量以及采样

深度呈负相关关系,而与其他变量均呈正相关关系.然而,秸秆还田输入氮量、土壤 pH 值、土壤有机质含量和采样深度的回归系数没有通过显著性检验($P>0.05$).而逐步回归通过反复添加或删除预测变量,可以有效地筛选出对模型贡献显著的预测变量,提高模型的可靠性.因此,采用逐步回归方法进一步对预测变量进行筛选以建立更优的回归方程.

逐步回归剔除了化肥施氮量、秸秆还田输入氮量、

有机肥输入氮量、有机质含量和采样深度 5 个预测变量,模型方程为 $Y_{\text{总氮淋失量}}=0.064X_{\text{总施氮量}}-0.001X_{\text{还田秸秆量}}+0.087X_{\text{水分投入量}}+0.893X_{\text{粘粒含量}}+0.305X_{\text{砂粒含量}}+7.472X_{\text{土壤 pH 值}}+11.170X_{\text{土壤全氮含量}}-129.773$.回归方程的检验显示,模型 F 统计量的 $P<0.001$.逐步回归所选取指标的回归系数都通过了显著性检验($P<0.05$).标准化回归系数表明土壤粘粒含量(0.489)、水分投入量(0.470)、总施氮量(0.395)等是影响冬小麦农田总氮淋失的重要因素.

表 2 冬小麦农田总氮淋失量的逐步回归模型参数

Table 2 Parameters of stepwise regression model for total N leaching loss in winter wheat fields

参数	系数预测值	标准化系数	标准差	t 值	P 值
截距 Intercept	-129.773		23.956	-5.417	<0.001
总施氮量 Total N rate	0.064	0.395	0.012	5.270	<0.001
还田秸秆量 Straw incorporation	-0.001	-0.203	0.000	-3.019	0.003
水分投入量 Water input	0.087	0.470	0.013	6.642	<0.001
粘粒含量 Clay content	0.893	0.489	0.151	5.926	<0.001
砂粒含量 Sand content	0.305	0.272	0.082	3.708	<0.001
土壤 pH 值 Soil pH	7.472	0.201	2.644	2.826	0.005
土壤全氮含量 Soil total N	11.170	0.132	5.525	2.022	0.045
F 值 F statistic: 14.3				P 值 P -value: 4.32×10^{-15}	
拟合优度 Multiple R^2 : 0.335				修正拟合优度 Adjusted R^2 : 0.311	

2.3 随机森林和 XGBoost 回归

研究采用网格搜索法和贝叶斯搜索法对预处理后的数据建立随机森林和 XGBoost 模型,经过多次实验优化后,随机森林和 XGBoost 模型超参数搜索范围如表 3.随机森林模型网格搜索法和贝叶斯搜

索法的决策树最大个数分别为 150 和 200,两种搜索法中树的最大深度和最小节点数均为 20 和 1. XGBoost 模型贝叶斯搜索法决策树个数、树的最大深度分别为 200 和 10.本研究中,进行 200 次模拟后,模型的平均拟合优度趋于稳定.

表 3 超参数范围

Table 3 Hyperparameter range

模型	超参数	参数范围	
		网格搜索法	贝叶斯搜索法
随机森林 Random Forest	决策树的数量($n_{\text{estimators}}$)	50, 100, 150	50~200
	树的最大深度(max_depth)	10, 15, 20	1~20
	节点最小分裂样本数($min_samples_split$)	2, 4, 8	2~10
	叶子节点最小样本数($min_samples_leaf$)	1, 2, 4	1~10
	决策树考虑的特征比例($max_features$)	0.3, 0.4, 0.5	0.1~1.0
XGBoost eXtreme Gradient Boosting	决策树的数量($n_{\text{estimators}}$)	—	50~200
	树的最大深度(max_depth)	—	3~10
	学习率($learning_rate$)	—	0.01~0.4
	叶子节点中最小样本权重和(min_child_weight)	—	3~9
	样本训练比例($subsample$)	—	0.3~0.8
	决策树考虑的特征比例($colsample_bytree$)	—	0.1~1.0
	γ	—	0.01~0.1
α	—	1.0~1.5	
λ	—	0~0.1	

结果表明,基于两种搜参方法的随机森林模型表现相似(表 4),通过对全部指标(组合 1)和逐步回归筛选后的指标(组合 10)进行模拟的模型拟合优度均高于多元回归模型.根据重要性指标剔除预测变量进行模拟时,当删除秸秆还田输入氮量(比较组合 3 和组合 4)和土壤粘粒、砂粒含量(比较组合 8 和组合 9)时,模型拟合优度显著下降.

对于全部指标和通过逐步回归筛选出的指标,基于贝叶斯搜参法构建的 XGBoost 模型的 R^2 分别为 0.702(组合 1)和 0.710(组合 10).虽然 XGBoost 模型在提高拟合程度方面表现出色,但在训练过程中仍然存

在严重的过拟合现象.为了进一步抑制模型的过拟合并提高训练效率,本文在贝叶斯搜参的基础上应用了早停法.在 XGBoost 回归中,研究首先对除决策树数量外的参数进行贝叶斯搜参,然后基于这些参数继续进行模型训练,当连续 20 次 5 折交叉验证性能没有提升时,停止模型的训练并在测试集上进行验证,这一方法有效地抑制了过拟合现象(全部指标 $R^2=0.745$,逐步回归筛选出指标 $R^2=0.722$).当删除秸秆还田输入氮量(比较组合 3 和组合 4)时,模型拟合优度显著下降.但是与随机森林不同,删除土壤粘粒、砂粒含量(比较组合 8 和组合 9)时,模型拟合优度变化不显著.

表 4 不同模型性能评价对比

Table 4 Evaluation results of different models

预测变量组合	基于不同优化方法下不同模型 R^2				
	随机森林 网格搜索法	随机森林 贝叶斯搜参	XGBoost 贝叶斯搜参	XGboost 贝叶斯搜参 +早停法	多因素线性 回归
(1)全部预测变量 All predicted variables	0.627	0.628	0.702	0.745	0.344
(2)水分投入量,化肥施氮量,砂粒含量,粘粒含量,土壤 pH 值,有机质含量,土壤全氮含量, 土壤类型,秸秆还田输入氮量,有机肥输入氮量 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH, Soil organic matter, Soil total N, Soil type, Straw incorporation N rate, Organic fertilizer N rate	0.609	0.607	0.681	0.671	0.343
(3)水分投入量,化肥施氮量,砂粒含量,粘粒含量,土壤 pH 值,有机质含量,土壤全氮含量, 土壤类型,秸秆还田输入氮量 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH, Soil organic matter, Soil total N, Soil type, Straw incorporation N rate	0.625	0.624	0.692	0.699	0.325
(4)水分投入量,化肥施氮量,砂粒含量,粘粒含量,土壤 pH 值,有机质含量,土壤全氮含量, 土壤类型 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH, Soil organic matter, Soil total N, Soil type	0.540	0.542	0.563	0.567	0.324
(5)水分投入量,化肥施氮量,砂粒含量,黏粒含量,土壤 pH 值,有机质含量,土壤全氮含量 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH, Soil organic matter, Soil total N	0.543	0.546	0.589	0.585	0.324
(6)水分投入量,化肥施氮量,砂粒含量,粘粒含量,土壤 pH 值,有机质含量 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH, Soil organic matter	0.537	0.542	0.575	0.524	0.292
(7)水分投入量,化肥施氮量,砂粒含量,黏粒含量,土壤 pH 值 Water input, Fertilizer N rate, Sand content, Clay content, Soil pH	0.530	0.531	0.560	0.553	0.289
(8)水分投入量,化肥施氮量,砂粒含量,黏粒含量 Water input, Fertilizer N rate, Sand content, Clay content	0.523	0.524	0.549	0.519	0.279
(9)水分投入量,化肥施氮量 Water input, Fertilizer N rate	0.310	0.311	0.607	0.610	0.127
(10)总施氮量,还田秸秆量,水分投入量,砂粒含量,粘粒含量,土壤 pH 值,土壤全氮含量 Total N rate, Straw incorporation, Water input, Sand content, Clay content, Soil pH, Soil total N	0.705	0.708	0.710	0.722	0.335

注:多因素线性回归中预测变量不包括土壤类型.

3 讨论

3.1 环境条件和农田管理措施对氮淋失的影响

Pearson 相关性分析和随机森林重要性分析结果表明,施氮量和水分投入量是影响总氮淋失量最为重要的因素.丁燕等^[34]通过 6a 田间试验,发现在

小麦-玉米种植体系中,随着施氮量和水分投入量的增加,总氮淋失量呈指数增加。张玉铭等^[35]的试验表明,随着施氮量和灌溉量的增加,土壤中硝态氮累积峰面积增大,且出现位置加深,高施氮处理硝态氮累积峰的位置相对于低施氮处理从 80cm 处移动至 100cm 处。表明施氮量和灌溉量的增加不仅增加氮淋失量,而且会增加氮淋失深度,进而影响地下水安全^[36]。除施氮量外,氮肥种类或施肥方式等对氮淋失也有一定的影响。研究表明,与传统农业相比,有机无机肥配施可以提高作物氮素利用率,较单施化肥显著降低总氮淋失量 40%左右^[11]。施用有机肥提高了土壤中反硝化菌群的活性^[37],促进硝态氮向有机氮的转化,抑制硝态氮淋失,从而降低对环境的污染风险。

在 Pearson 相关性分析中,秸秆还田输入氮量和总氮淋失量呈负相关,与多元回归中得出的结果一致,且在随机森林和 XGBoost 模型中对秸秆还田输入氮量进行剔除时,均出现拟合优度显著下降的现象,这表明了秸秆还田在对氮淋失的预测中可能占据重要地位,这是在以前的研究中大多得到忽视的。秸秆还田对氮淋失的影响主要是由于:一改善表层土壤团聚体性能^[38],增加土壤胶体的吸附能力;二还田秸秆的覆盖作用防止水分蒸发、提高土壤保水能力^[38];三秸秆还田为土壤微生物提供了碳源,从而促进微生物对氮的利用和固定^[39-40],因此,秸秆还田能够显著降低氮淋失量。

氮淋失还受到土壤性质等因素的影响。在 pearson 相关性分析中总氮淋失量与土壤 pH 值呈显著正相关关系,且在随机森林重要性分析中砂粒、粘粒含量等土壤性质的重要性也仅次于施氮量和水分投入量。多元回归分析的 R^2 较一元和二元回归分析的 R^2 显著增加,以及在随机森林模型和 XGBoost 模型模拟过程中减少预测变量后的 R^2 显著降低,表明对于整个华北地区冬小麦氮淋失量的预测,必须考虑农田管理措施和土壤性质等多种因素的共同影响。

3.2 机器学习与传统统计模型在氮淋失预测中的应用

传统的经验统计模型(一元或多元回归分析)主要是基于变量之间的关系来进行建模和预测的一种统计方法,通过将数据拟合到已知的统计理论(例

如 t 分布或 F 分布等)或统计模型(例如线性回归模型或指数回归模型等)来检验假设^[41],并通过模型解释度 R^2 和 F 检验(P 值)来判别模型的可靠性,从而说明预测变量对目标变量的影响。机器学习则是在定义的数据集上进行学习,在给定的“输入”和“输出”变量之间构建算法,从大量的数据中捕捉到传统经验模型难以表达的规律,从而对目标变量进行预测。机器学习过程中更注重预测精度 (R^2)^[42],而不是估计变量之间的相关性^[41]。

传统的经验统计模型计算简单,只需要少量参数即可获得预测模型^[43],但是只通过单一预测变量或两个预测变量对总氮淋失量进行预测的解释度不高。本研究通过多元回归分析构建的模型,相较一元和二元回归分析的 R^2 虽有所提升,但解释度依然不高,主要是多元回归模型也无法很好地拟合数据的非线性关系^[44]。机器学习仅通过农田管理措施(化肥施氮量和水分投入量)建立的 XGBoost 预测模型就表现出较高的拟合度($R^2=0.610$),预测精度显著提升。并且相较于一元回归模型、二元回归模型和多元回归模型,机器学习在将多个因素作为预测变量的同时,量化每个预测变量对模型预测性能的贡献程度,从而识别出对目标变量具有重要影响的指标,通过构建多次模拟获得最佳的预测模型。本研究结合逐步回归筛选的指标构建的随机森林和 XGBoost 模型在氮淋失预测中 R^2 值(最大为 0.722)显著优于常规回归模型。表明机器学习模型在如华北地区的较大区域内进行预测比传统统计模型更具推广性。

单纯依赖机器学习模型进行预测和解释华北地区冬小麦农田氮淋失也存在一定局限性。机器学习模型通常以“黑盒”的形式呈现,无法像传统统计模型一样明确各个预测变量对目标变量的具体影响^[45],对预测变量的添加与是否存在一定的盲目性,使得模型的解释变得困难。另外,机器学习模型的性能极大依赖于数据的质量^[46],如果数据集中存在噪声数据或异常值,模型可能表现出较低的鲁棒性。部分机器学习算法(尤其是深度学习)相较于传统的经验模型需要大量的计算资源,对训练和预测成本具有较高的要求^[47]。

随机森林算法基于引导聚集算法(Bagging)^[32],通过对多个决策树的投票来进行预测,因此每个指标的重要性均被平等地衡量。与之相比,XGBoost 采

用梯度提升算法(Boosting)^[32],其特点是每一轮迭代都会根据上一轮的模型结果来调整.这意味着对于每棵决策树的建立,特征分裂的机会并不均等^[48],使得 XGBoost 更加关注对目标函数影响最大的特征,降低 XGBoost 在重要性分析中的解释率.在机器学习过程中,由于模型复杂度过高也会存在过拟合现象,而 XGBoost 模型中早停法能够在模型拟合优度不再提升时,及时停止训练,有效地抑制了过拟合现象.

本研究中的模型是基于华北地区的数据集训练构建的,其适用范围仅限于本研究所定义的华北区域,若要在其他地区应用,则需进一步进行验证.此外,机器学习构建模型过程中仅根据重要性分析的结果筛选变量存在一定的片面性,本试验结合逐步回归筛选预测变量可提高预测模型的拟合优度.在 XGBoost 构建模型过程中去除土壤粘粒和砂粒含量,模型的拟合优度没有如随机森林模型一样显著下降(比较组合 8 和组合 9),与土壤的粘粒含量和砂粒含量是影响氮淋失的重要因素的事实不符^[49].考虑到 R^2 虽然是常用的衡量模型拟合程度的指标,但其只能提供模型对预测变量的解释程度.因此,机器学习模型在预测农田氮淋失时还需要结合相关领域知识、统计方法或其他机器学习算法的结果对这些指标进行筛选,以确保选取的预测变量具有更高的可解释性和预测能力^[50].

4 结语

华北地区冬小麦农田氮淋失受农田管理措施和土壤性质等因素共同影响.施氮量和水分投入量对冬小麦农田氮淋失的影响最为重要,秸秆还田能够有效抑制氮淋失的发生.随机森林在多元指标拟合方面表现出色,通过对全部指标和逐步回归筛选指标建立的模型的 R^2 分别为 0.628 和 0.708,但在建立化肥施氮量和水分投入量双因素预测模型时, R^2 仅有 0.311.研究结果表明将随机森林模型的构建与经验统计分析方法相结合,确定了构建模型的最佳因子组合,提高了模型的预测精度和可靠性.经过贝叶斯搜索和早停法优化后的 XGBoost 模型在全部预测变量上拟合程度最优($R^2=0.745$),对双因素预测变量的 R^2 也提高至 0.610.机器学习模型在预测农田氮淋失时还需要结合相关领域知识、统计方法或其

他机器学习算法的结果对这些指标进行筛选,以提高模型的可行性和适用性.

参考文献:

- [1] 中华人民共和国国家统计局.中国统计年鉴 [EB/OL]. 2023-9-27 [2023-9-27]. <http://www.stats.gov.cn/sj/ndsj/2023/indexch.htm>.
- [2] Lu Y, Wang E L, Zhao Z G, et al. Optimizing irrigation to reduce N leaching and maintain high crop productivity through the manipulation of soil water storage under summer monsoon climate [J]. *Field Crops Research*, 2021,265:1-11.
- [3] 金 梁.基于 SPWS 模型的华北平原农田水氮利用效率及环境效应分析 [D]. 北京:中国农业大学, 2007.
Jin L. Analysis of water and nitrogen use efficiency and environmental effect based on the SPWS model in North China Plain [D]. Beijing: China Agricultural University, 2007.
- [4] Huang M X, Liang T, Wang L Q, et al. No-tillage and fertilization management on crop yields and nitrate leaching in North China Plain [J]. *Ecology and Evolution*, 2015,5(6):1143-1155.
- [5] 刘宏元,张爱平,杨世琦,等.山东省冬小麦-夏玉米轮作体系土壤氮素盈余指标体系的构建与评价——以德州市为例 [J]. *农业环境科学学报*, 2019,38(6):1321-1329.
Liu H Y, Zhang A P, Yang S Q, et al. Construction and evaluation of a soil nitrogen surplus index system for the wheat maize rotation system in Shandong Province, China: A case study of Dezhou City [J]. *Journal of Agro-Environment Science*, 2019,38(6):1321-1329.
- [6] 骆晓声,寇长林,王小非,等.施氮量对潮土区冬小麦-夏玉米轮作农田氮磷淋溶的影响 [J]. *中国生态农业学报(中英文)*, 2021,29(1):29-37.
Luo X S, Kou C L, Wang X F, et al. Effects of nitrogen application on nitrogen and phosphorus leaching in fluvo-aquic soil on a winter wheat-summer maize rotation farmland [J]. *Chinese Journal of Eco-Agriculture*, 2021,29(1):29-37.
- [7] Hou X K, Zhou F, Leip A, et al. Spatial patterns of nitrogen runoff from Chinese paddy fields [J]. *Agriculture, Ecosystems & Environment*, 2016,231:246-254.
- [8] 李晓欣,张菲菲,马洪斌,等.华北平原地区农田硝态盐淋失研究进展 [J]. *华北农学报*, 2011,26(S2):131-139.
Li X X, Zhang F F, Ma H B, et al. Nitrate leaching in North China Plain: A review [J]. *Acta Agriculturae Boreali-Sinica*, 2011,26(S2):131-139.
- [9] 刘宏斌,李志宏,张云贵,等.北京平原农区地下水硝态氮污染状况及其影响因素研究 [J]. *土壤学报*, 2006,43(3):405-413.
Liu H B, Li Z H, Zhang Y G, et al. Nitrate contamination of groundwater and its affecting factors in rural areas of Beijing Plain [J]. *Acta Pedologica Sinica*, 2006,43(3):405-413.
- [10] Zheng W K, Wan Y S, Li Y C, et al. Developing water and nitrogen budgets of a wheat-maize rotation system using auto-weighting lysimeters: Effects of blended application of controlled-release and un-coated urea [J]. *Environmental Pollution*, 2020,263:114383.
- [11] Fan J L, Xiao J, Liu D Y, et al. Effect of application of dairy manure, effluent and inorganic fertilizer on nitrogen leaching in clayey fluvo-aquic soil: A lysimeter study [J]. *Science of the Total*

- Environment, 2017,592:206-214.
- [12] 陈淑峰.华北平原高产粮区地下水硝态氮时空变异与综合调控途径研究——以山东桓台县为例 [D]. 北京:中国农业大学, 2007.
Chen S F. A Study on Spatio-temporal Variability of Groundwater Nitrate Concentration and Comprehensive Improve-control Measures in High-yielding Region in North China Plain——A Case study on Huantai County of Shandong Province [D]. Beijing: China Agricultural University, 2007.
- [13] 刘明鹏,徐开未,肖 华,等.氮肥施用对四川紫色土矿质态氮淋失特征及春玉米产量的影响 [J]. 农业资源与环境学报, 2022,39(1): 88-98.
Liu M P, Xu K W, Xiao H, et al. Effects of nitrogen application on the characteristics of mineral nitrogen leaching in purplish soil and spring maize yield in Sichuan [J]. Journal of Agricultural Resources and Environment, 2022,39(1):88-98.
- [14] 乔毅博,吴鹏年,王艳丽,等.黄淮海平原微喷灌下冬小麦农田水分渗漏及氮素淋失模拟分析 [J]. 农业环境科学学报, 2023,42(7):1540-1553.
Qiao Y B, Wu P N, Wang Y L, et al. Simulation analysis on water leakage and nitrogen leaching of winter wheat farmland under micro sprinkler irrigation in Huang-Huai-Hai Plain [J]. Journal of Agro-Environment Science, 2023,42(7):1540-1553.
- [15] 王洪媛,李俊改,樊秉乾,等.中国北方主要农区农田氮磷淋溶特征与时空规律 [J]. 中国生态农业学报(中英文), 2021,29(1):11-18.
Wang H Y, Li J G, Fan B Q, et al. Nitrogen and phosphorus leaching characteristics and temporal and spatial distribution patterns in northern China farmlands [J]. Chinese Journal of Eco-Agriculture, 2021,29(1):11-18.
- [16] 吴三鼎,董 强,党廷辉.减量施氮及秸秆深埋对春玉米地土壤电导率和硝态氮淋溶的影响 [J]. 水土保持学报, 2018,32(6):46-51.
Wu S D, Dong Q, Dang T H. Effects of reduced nitrogen application and deep burial of straw on soil electrical conductivity and nitrate nitrogen leaching in spring maize field [J]. Journal of Soil and Water Conservation, 2018,32(6):46-51.
- [17] 孟凡乔,王 坤,肖广敏,等.华北平原潮土区粮田氮淋失阻控措施及效果分析 [J]. 中国生态农业学报(中英文), 2021,29(1):141-153.
Meng F Q, Wang K, Xiao G M, et al. Nitrogen leaching mitigation in fluvo-aquic soil in the North China Plain [J]. Chinese Journal of Eco-Agriculture, 2021,29(1):141-153.
- [18] Cui Z L, Zhang X Y, Chen X P, et al. Pursuing sustainable productivity with millions of smallholder farmers [J]. Nature, 2018,555(7696):363-366.
- [19] 杨旺鑫.我国农田氮磷损失影响因素及损失量初步估算 [D]. 南京:南京农业大学, 2015.
Yang W X. Influencing factors and estimation of nitrogen and phosphorus losses from farmlands in China [D]. Nanjing: Nanjing Agricultural University, 2015.
- [20] 黄明祥,张健钦,杨 毅,等.1989~2022 年生态环境中人工智能应用的研究综述——基于 Citespace 的知识图谱分析 [J]. 环境保护科学, 2024,50(2):8-16.
Huang M X, Zhang J Q, Yang Y, et al. Artificial intelligence applications in ecological environments from 1989 to 2022——Knowledge graph analysis based on Citespace [J]. Environmental Protection Science, 2024,50(2):8-16.
- [21] 李 焕,苏 慧,张 婷,等.基于机器学习的汾渭平原 PM_{2.5} 和 O₃ 变化特征及影响因素 [J]. 环境化学, 2024,43(5):1-14.
Li H, Su H, Zhang T, et al. Variation characteristics and influencing factors of PM_{2.5} and O₃ based on machine learning in Fenwei Plain [J]. Environmental Chemistry, 2024,43(5):1-14.
- [22] Liu M Y, Zheng H W, Cai M G, et al. Ocean stratification impacts on dissolved polycyclic aromatic hydrocarbons (PAHs): From global observation to deep learning [J]. Environmental Science & Technology, 2023,57(46):18339-18349.
- [23] Palansooriya K N, Li J, Dissanayake P D, et al. Prediction of Soil Heavy Metal Immobilization by Biochar Using Machine Learning [J]. Environmental Science & Technology, 2022,56(7):4187-4198.
- [24] Genuer R, Poggi J M, Tuleau-Malot C. Variable selection using random forests [J]. Pattern Recognition Letters, 2010,31(14):2225-2236.
- [25] Ying H, Xue Y F, Yan K, et al. Safeguarding food supply and groundwater safety for maize production in China [J]. Environmental Science & Technology, 2020,54(16):9939-9948.
- [26] Liu Q, Liu B, Zhang Y, et al. Biochar application as a tool to decrease soil nitrogen losses (NH₃ volatilization, N₂O emissions, and N leaching) from croplands: Options and mitigation strength in a global perspective [J]. Global Change Biology, 2019,25(6):2077-2093.
- [27] 程婉清,袁定波,熊 鹏,等.基于多种机器学习算法的水质指数预测模型构建与评估 [J]. 环境科学学报, 2023,43(11):144-152.
Cheng W Q, Yuan D B, Xiong P, et al. Construction and evaluation of city water quality index prediction model based on multiple machine learning algorithms [J]. Acta Scientiae Circumstantiae, 2023,43(11): 144-152.
- [28] 张国梁,章 申.农田氮素淋失研究进展 [J]. 土壤, 1998,(6):291-297.
Zhang G L, Zhang S. Research progress on nitrogen leaching from farmland [J]. Soil, 1998,(6):291-297.
- [29] 黄满湘,章 申,张国梁.应用大型原状土柱渗漏计测定冬小麦-夏玉米轮作期硝态氮淋失 [J]. 环境科学学报, 2003,(1):11-16.
Huang M X, Zhang S, Zhang G L. Soil core lysimeter study of nitrate leaching from a winter wheat-summer maize rotation [J]. Acta Scientiae Circumstantiae, 2003,(1):11-16.
- [30] 刘 健.三种质地土壤氮素淋溶规律研究 [D]. 北京:北京林业大学, 2010.
Liu J. Study on nitrogen leaching regulations on three textures of soil [D]. Beijing: Beijing Forestry University, 2010.
- [31] Shalabi L A, Shaaban Z, Kasasbeh B. Data mining: A preprocessing engine [J]. Journal of Computer Science, 2006,2(9):735-739.
- [32] Zounemat-kermani M, Batelaan O, Fadaee M, et al. Ensemble machine learning paradigms in hydrology: A review [J]. Journal of Hydrology, 2021,598:126266.
- [33] 李海霞,宋丹蕾,孔佳宁,等.传统机器学习模型的超参数优化技术评估 [J]. 计算机科学, 2024,51(8):242-255.
Li H X, Song D L, Kong J N, et al. Evaluation of hyperparameter optimization techniques for traditional machine learning models [J]. Computer Science, 2024,51(8):242-255.
- [34] 丁 燕,杨宪龙,同延安,等.小麦-玉米轮作体系农田氮素淋失特征

- 及氮素表观平衡 [J]. 环境科学学报, 2015,35(6):1914-1921.
- Ding Y, Yang X L, Tong Y A, et al. Characteristics of N leaching and apparent N budget in cultivated lands under a Winter wheat–summer maize rotation system [J]. *Acta Scientiae Circumstantiae*, 2015,35(6): 1914-1921.
- [35] 张玉铭,张佳宝,胡春胜,等.水肥耦合对华北高产农区小麦-玉米产量和土壤硝态氮淋失风险的影响 [J]. 中国生态农业学报, 2011, 19(3):532-539.
- Zhang Y M, Zhang J B, Hu C S, et al. Effect of fertilization and irrigation on wheat–maize yield and soil nitrate nitrogen leaching in high agricultural yield region in North China Plain [J]. *Chinese Journal of Eco-Agriculture*, 2011,19(3):532-539.
- [36] Huang T, Ju X T, Yang H. Nitrate leaching in a winter wheat–summer maize rotation on a calcareous soil as affected by nitrogen and straw management [J]. *Scientific Reports*, 2017,7:42247.
- [37] Kramer S B, Reganold J P, Glover J D, et al. Reduced nitrate leaching and enhanced denitrifier activity and efficiency in organically fertilized soils [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006,103(12):4522-4527.
- [38] Blanco-lanco-canqui H, Lal R. Soil structure and organic carbon relationships following 10years of wheat straw management in no-till [J]. *Soil & Tillage Research*, 2007,95(1/2):240-254.
- [39] Meng F Q, Dungait J A J, Xu X L, et al. Coupled incorporation of maize (*Zea mays* L.) straw with nitrogen fertilizer increased soil organic carbon in Fluvic Cambisol [J]. *Geoderma*, 2017,304:19-27.
- [40] Qiu S J, Ju X T, Ingwersen J, et al. Role of carbon substrates added in the transformation of surplus nitrate to organic nitrogen in a calcareous soil [J]. *Pedosphere*, 2013,23(2):205-212.
- [41] Ryu L, Han K. Machine Learning vs. Statistical Model for Prediction Modelling: Application in Medical Imaging Research [J]. *Journal of the Korean Society of Radiology*, 2022,83(6):1219-1228.
- [42] Ley C, Martin R K, Pareek A, et al. Machine learning and conventional statistics: making sense of the differences [J]. *Knee Surgery Sports Traumatology Arthroscopy*, 2022,30(3):753-757.
- [43] 王桂良.中国三大粮食作物农田活性氮损失与氮肥利用率的定量分析 [D]. 北京:中国农业大学, 2014.
- Wang G L. Quantitative analysis of reactive nitrogen losses and nitrogen use efficiency of three major grain crops in China [D]. Beijing: China Agricultural University, 2014.
- [44] 侯佟泽.基于多元回归的过程监测研究及其工业应用 [D]. 北京:北京化工大学, 2024.
- Hou T Z. Research on process monitoring based on multivariate regression and its industrial application [D]. Beijing: Beijing University of Chemical Technology, 2024.
- [45] Xu T, Valocchi A J, Choi J, et al. Use of machine learning methods to reduce predictive error of groundwater models [J]. *Groundwater*, 2014, 52(3):448-460.
- [46] Hu Y, Chapman A, Wen G, et al. What Can Knowledge Bring to Machine Learning?—A Survey of Low-shot Learning for Structured Data [J]. *Acm Transactions on Intelligent Systems and Technology*, 2022,13(3):41-45.
- [47] Khan T, Tian W, Zhou G, et al. Machine learning (ML)-centric resource management in cloud computing: A review and future directions [J]. *Journal of Network and Computer Applications*, 2022, 204:103405.
- [48] Friedman J H. Greedy function approximation: A gradient boosting machine [J]. *Annals of Statistics*, 2001,29(5):1189-1232.
- [49] Zhou J B, Xi J G, Chen Z J, et al. Leaching and transformation of nitrogen fertilizers in soil after application of N with irrigation: A soil column method [J]. *Pedosphere*, 2006,16(2):245-252.
- [50] Salvo C D. Improving results of existing groundwater numerical models using machine learning techniques: A review [J]. *Water*, 2022, 14(15):2307.
- 作者简介:** 李涛宇(2003-),男,江苏扬州人,本科生,研究方向为土壤氮循环.发表论文1篇.yztaoyuli@163.com.