

DOI:10.11686/cyxb2025033

http://cyxb.magtech.com.cn

肖娅婷, 唐彧哲, 王潞, 等. 光谱指数助力机器学习提高玉米叶片含水量估测精度. 草业学报, 2025, 34(12): 85-96.

XIAO Ya-ting, TANG Yu-zhe, WANG Lu, *et al.* Use of spectral index-assisted machine learning to improve the accuracy of maize leaf water content estimation. Acta Prataculturae Sinica, 2025, 34(12): 85-96.

光谱指数助力机器学习提高玉米叶片含水量估测精度

肖娅婷^{1,2}, 唐彧哲^{1,2}, 王潞^{1,2}, 白宇飞^{1,2}, 杨海波^{1,2}, 李斐^{1,2*}

(1. 内蒙古农业大学草原与资源环境学院, 内蒙古自治区土壤质量与养分资源重点实验室, 内蒙古呼和浩特 010018; 2. 农业生态安全与绿色发展自治区高等学校重点实验室, 内蒙古呼和浩特 010018)

摘要: 玉米水分状况的快速无损监测对于水分诊断和灌溉管理具有重要意义, 而光谱指数是作物叶片含水量(LWC)无损实时估测的重要指标之一, 然而传统的光谱指数在估测LWC时易受外界环境因素影响, 导致估测精度较低。机器学习(ML)算法在预测作物水分状况方面有显著的优势, 特别是针对精准农业和作物水分监测等应用。因此, 本研究旨在将光谱指数与机器学习结合起来进一步提高LWC的估测精度, 实现玉米水分的高效利用。研究于2023-2024年在内蒙古玉米种植的典型区域进行不同水分梯度的田间试验, 测定玉米叶片3个关键生育时期的高光谱反射率, 分析13种水敏感的光谱指数与玉米LWC的相关关系后, 利用Relieff技术筛选主要光谱特征作为偏最小二乘(PLSR)、随机森林(RF)、高斯过程回归(GPR)3种机器学习算法的输入变量, 构建玉米LWC的估测模型。结果表明, 在13种水情指数中改良的DATT指数(MDATT)预测性能最佳(决定系数 $R^2=0.52$), 但估测精度受生育时期和玉米叶片层位影响较大, 不能有效监测玉米LWC。而将全波段光谱(350~2500 nm)和通过Relieff技术筛选的主要光谱指数分别投入3种机器学习算法中, LWC的估测精度提升了7%~45%。其中以光谱指数作为输入特征的模型整体表现较好, RF和GPR模型在LWC估测中表现较为优越, 可解释玉米LWC 88%~89%的变异。最后利用独立数据集对RF和GPR模型进行验证, RF和GPR模型决定系数 R^2 分别为0.89和0.88, 均方根误差(RMSE)为1.95%和2.04%。总体来看, 光谱指数与RF和GPR算法耦合起到了级联效应, 可以显著提高玉米LWC的估测精度, 研究结果将为玉米LWC的估测提供可靠的方法, 并为玉米水肥一体化管理提供科学有效的依据。

关键词: 玉米叶片; 含水量; 光谱指数; 机器学习

Use of spectral index-assisted machine learning to improve the accuracy of maize leaf water content estimation

XIAO Ya-ting^{1,2}, TANG Yu-zhe^{1,2}, WANG Lu^{1,2}, BAI Yu-fei^{1,2}, YANG Hai-bo^{1,2}, LI Fei^{1,2*}

1. College of Grassland, Resources and Environment, Inner Mongolia Agricultural University, Inner Mongolia Key Laboratory of Soil Quality and Nutrient Resources, Hohhot 010018, China; 2. Key Laboratory of Agricultural Ecological Security and Green Development in Higher Education Institutions of Inner Mongolia Autonomous Region, Hohhot 010018, China

Abstract: Rapid and non-destructive monitoring of the water status of maize (*Zea mays*) is important for water status diagnosis and irrigation management. Spectral indices serve as crucial tools for non-destructive real-time estimation of crop leaf water content (LWC). However, traditional spectral indices are sensitive to external environmental factors, resulting in reduced prediction accuracy when they are used to estimate LWC. Machine learning (ML)

收稿日期: 2025-02-02; 改回日期: 2025-03-10

基金项目: 国家重点研发计划(2023YFD1900404-03)和国家重点研发计划课题(2024YFD1700400)资助。

作者简介: 肖娅婷(1998-), 女, 内蒙古巴彦淖尔人, 在读硕士。E-mail: XiaoYT@emails.imau.edu.cn

* 通信作者 Corresponding author. E-mail: Lifei@imau.edu.cn

algorithms demonstrate distinct advantages in predicting crop water status, particularly when applied in precision agriculture and crop water status monitoring. Therefore the aims of this study were to enhance the accuracy of LWC estimation by integrating spectral indices with ML approaches, with an overall goal to facilitate efficient water resource utilization during maize cultivation. Field experiments with varying water gradients were conducted in typical maize cultivation regions of Inner Mongolia during 2023–2024. The hyperspectral reflectance of maize leaves were measured across three critical growth stages, and then correlation analyses were conducted between maize LWC and 13 water-sensitive spectral indices. To develop LWC estimation models, spectral features selected *via* the ReliefF technique were used as input variables for three ML algorithms-partial least squares regression (PLSR), random forest (RF), and Gaussian process regression (GPR). The results demonstrate that among the 13 hydrological indices, the modified DATT index exhibited optimal predictive performance (coefficient of determination $R^2=0.52$). However, its accuracy was affected by the growth stage and leaf canopy position, limiting its effectiveness for LWC monitoring. Integrating full-spectrum data (350–2500 nm) with ReliefF-selected spectral indices into ML algorithms enhanced the accuracy of LWC estimates by 7%–45%. Models utilizing spectral indices as input features demonstrated superior overall performance, with the RF and GPR models explaining 88%–89% of LWC variability. Independent validations confirmed the robustness of the models, with coefficient of determination R^2 values of 0.89 (RF) and 0.88 (GPR) and root mean square error values of 1.95% and 2.04%. Our results show that the synergistic combination of spectral indices with RF/GPR algorithms had cascading effects, significantly improving the accuracy of LWC estimation. This methodology provides a reliable approach for monitoring maize water status and establishes a scientific foundation for the development of precise integrated water-fertilizer management systems.

Key words: maize leaves; water content; spectral index; machine learning

青贮玉米(*Zea mays*)作为一种高营养、高能量的饲料作物,具有重要的农业和畜牧业意义。与其他传统饲料(如干草、秸秆等)相比,青贮玉米能够提供更高的营养密度,促进牲畜的快速生长和健康^[1]。然而,青贮玉米对水分的需求较大,尤其是在干旱半干旱的地区,缺水可能导致其生长受阻或早期枯萎,从而影响产量和品质。叶片含水量(leaf water content, LWC)是青贮玉米生长和产量形成的关键因素之一,合理管理水分供给可以显著提高青贮玉米的生长性能和抗逆能力。但在大规模集约化种植管理中,传统获取LWC的方法通常费时费力,且对作物具有一定破坏性,难以提供长期、动态、可靠的数据。因此,能够快速无损检测并获取LWC成了提高青贮玉米品质和产量的重要前提。

在现代农业中,遥感技术通过分析玉米叶片的反射光谱监测其水分状况已成为一种重要手段,并在农业中得到了广泛的应用。早期的光谱建模研究通过从光谱带中2到3个敏感波段构建各种类型的光谱指数。其中,Rouse等^[2]最早利用红光波和近红外波的反射率差异,构建了归一化差值植被指数(normalized difference vegetation index, NDVI),用于表征植被的生长状况。在NDVI的基础上,研究者们进一步开发了多种基于光谱敏感带的水敏感光谱指数,以提高LWC的预测精度。常用的水指数(water index, WI)和归一化植被指数的比值(WI/NDVI)在预测水稻(*Oryza sativa*)、花生(*Arachis hypogaea*)、大豆(*Glycine max*)和小麦(*Triticum aestivum*)的LWC方面表现出较好的效果^[3]。同时,研究也表明LWC与归一化差异水分指数(normalized difference water index, NDWI)、简单比值(simple ratio index, SR)以及短波红外垂直失水指数(shortwave infrared perpendicular water stress index, SPSI)之间存在显著的相关性^[4]。近年来,研究者们持续开发新的水敏感指数用于预测各种作物的水分状况^[5]。虽然水敏感光谱指数已经实现LWC的无损估测,但基于2到3个光谱波段构建的光谱指数通常未能考虑生长环境和生长特性对作物LWC的影响。气候特征和植物生理状况是影响LWC的重要因素^[6],被忽视的信息可能是现有植被指数模型在实际应用中泛化能力差的主要原因。因此,由于考虑了生长环境和生长

特性对LWC的潜在影响,光谱指数研究不能仅依赖于单一的遥感波段数据,而应结合多源数据进行综合分析,以实现更精确的LWC预测,从而提高模型的适用性和鲁棒性。

近年来,随着深度学习和人工智能技术的发展,光谱指数与机器学习结合的研究逐渐增多。通过将传统光谱指数和机器学习算法相结合,可以从遥感数据中提取出更为复杂和细致的特征,从而实现对作物水分状况更精确地预测。研究表明,利用光谱指数和水分敏感波段与偏最小二乘回归(partial least squares regression, PLSR)、随机森林(random forest, RF)、极端随机树(extremely randomized trees, ERT)、K近邻(K nearest neighbors, KNN)结合构建的小麦LWC估测模型,成功提高了水分预测精度^[7]。同时,结合近红外和红边波段衍生的光谱变量的随机森林算法,也有效估算了玉米等效水厚度(equivalent water thickness, EWT)、燃料水分含量(fuel moisture content, FMC)^[8]。而通过竞争性自适应重加权采样(competitive adaptive reweighting sampling, CARS)和随机蛙跳算法(random frog jump algorithm, RFJA)选取特征波段,并运用偏最小二乘回归、支持向量回归(support vector regression, SVR)和BP神经网络回归(backpropagation neural network, BPNN)等方法,建立的马铃薯(*Solanum tuberosum*)LWC估算模型也表现出良好的适用性^[9]。这些方法能够处理复杂的非线性关系,并通过训练模型学习光谱特征与叶片水分含量之间的映射关系。但当使用全频段数据进行建模时,数据冗余和频带干扰问题可能导致模型性能下降,同时显著增加计算成本和处理时间,这对模型精度产生了一定的负面影响^[10]。因此,筛选与植被参数密切相关的敏感波段作为模型输入变量对于提高模型精度和效率至关重要。

青贮玉米作为重要的饲料作物,提高其LWC的估测精度对于田间作物精准灌溉生产管理和作物的生理需水特性研究具有重要意义。尤其是在干旱半干旱地区,水分是植物生长发育的首要限制因素。因此需要精准的滴灌技术实现高效的水分利用,而这需要高精度的LWC实时估测数据作为支撑。虽然光谱指数在作物LWC预测中表现出了较好的优势,但其易受饱和效应的影响,限制了在高含水量条件下的准确性。在实际农业环境中,如何实现快速、实时、低成本的叶片水分预测仍是亟待解决的问题。尤其是在大规模农田监控时,如何将光谱学方法与机器学习模型结合,开发适应复杂环境和生长阶段的高效预测模型,仍是研究难点。本研究旨在选择水敏感光谱指数来评估青贮玉米LWC,并利用特征选择方法优化预测模型的输入变量,将光谱指数和全波段光谱数据作为机器学习模型的输入变量,对比3种机器学习算法的估测精度获得性能最佳的模型,用于估计青贮玉米LWC。以期为干旱及半干旱地区青贮玉米田间水分状况的动态诊断、精准滴灌实施以及灌溉策略的智慧化决策提供技术支撑与科学依据。

1 材料与方法

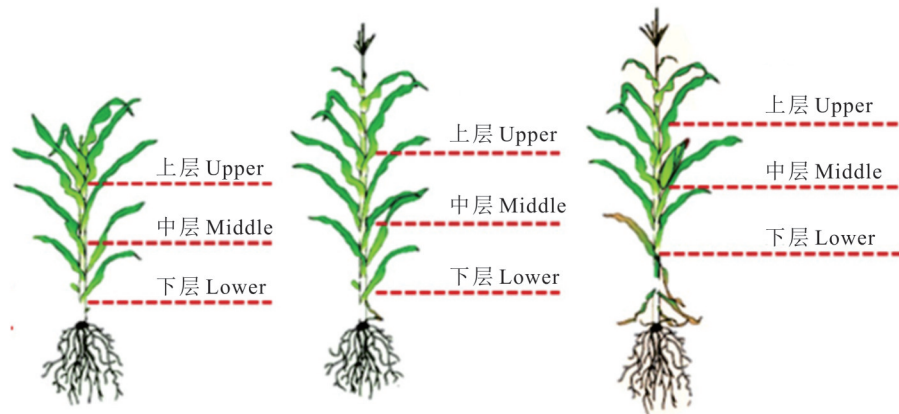
1.1 试验设计

本研究于2023—2024年进行了两个玉米田间试验,试验设计如下:试验1于2023年在巴彦淖尔市乌拉特前旗开展,种植的品种为‘新玉12’,种植区长57 m,宽53 m,每个处理设置3个重复,试验小区随机排列分布。共设有8个不同处理,每个处理的灌溉水量分别为0、18、240、300、360、420、480、540 m³。分别在拔节期、小喇叭口期、大喇叭口期、抽雄期、开花吐丝期和开花吐丝期后第7、14、21、28、38天进行灌溉。试验2于2024年在巴彦淖尔市杭锦后旗,种植品种为‘西蒙3358’,种植区长90 m,宽66 m。每个处理设置3个重复,试验小区随机排列分布。共设置3个不同处理,W₁灌水量维持田间持水量的90%,W₂灌水量维持田间持水量的70%,W₃灌水量维持田间持水量的50%。

1.2 测定项目及数据收集

1.2.1 采样时间和位置 试验1和试验2均在玉米的大喇叭口期(twelfth leaf, V12),抽雄期(tasseling, VT)和开花吐丝期(silking, R1)进行叶片分层测定和取样,大喇叭口期叶片分为3层;抽雄期穗位叶上的3片叶子作为上层,穗位叶及以下两层叶片为中层,以此往下的3片作为下层;开花吐丝期穗位叶两片作为中层,穗位叶上3片叶子作为上层,穗位叶下3片叶作为下层(图1)。

1.2.2 光谱数据的采集 利用美国SPECTRALEVOLUTION公司生产的PSR-3500全光谱手持式地物光谱仪(美国)采集玉米叶片光谱反射率数据,该仪器的光谱为350~2500 nm,输出间隔为1 nm。每个小区选取具



大喇叭口期 Twelfth leaf (V12) 抽雄期 Tasseling (VT) 开花吐丝期 Silking (R1)

图1 春玉米大喇叭口期、抽雄期和开花吐丝期阶段叶片垂直分层

Fig. 1 The vertical division of the spring maize leaf at the V12, VT, and R1 growth stage

有代表性的植株3株,使用主动光源手柄叶片夹夹取叶片的中间部位,并且保证避开叶脉区域,同时叶片保持水平并垂直于叶片夹,所有的光谱测量都是在晴朗天气的上午11:30至下午2:00进行的。每个叶片样本进行了6次重复测量,取平均值作为这个特定玉米叶片的光谱反射率,采集完叶片光谱反射率数据后存储在配套的掌上电脑中。

1.2.3 光谱指数的选择 植被水分指数是基于特定波段组合的反射率比值,这些指数能够有效地捕捉到水分变化带来的反射率差异,一般由参考波段和特征波段组成,能够反映植被在900~2500 nm植被叶片含水量变化及光谱反射率的响应变化。通常是以选取的波段进行相比和归一化形式表达。参考波段通常选择与含水量相关性较弱的光谱波段,例如820和860 nm等波段。特征波段通常以与植被含水量相关性较大的波段为依据,900~1300 nm光谱区和1500~2500 nm光谱区内选取特征波段。因此本研究通过文献统计方法选用了13种对于植被水分比较敏感的光谱指数(表1),以测试其在估算本研究中玉米LWC方面的有效性。

表1 本研究选取的光谱指数

Table 1 Spectral indices used in this study

光谱指数 Spectral index	公式 Formula	文献 References
归一化差异水分指数 Normalized difference water index (NDWI)	$(R_{860} - R_{1240}) / (R_{860} + R_{1240})$	[11]
水分指数 Water index (WI)	R_{900} / R_{970}	[12]
水分胁迫指数 Moisture stress index (MSI)	R_{1600} / R_{820}	[13]
简单比值水分指数 Simple ratio water index (SRWI)	R_{860} / R_{1240}	[14]
植被水分指数 Plant water index (PWI)	R_{970} / R_{900}	[15]
生理反射指数 Physiological reflectance index (PRI)	$(R_{531} - R_{570}) / (R_{531} + R_{570})$	[16]
改良的DATT指数 Modified DATT (MDATT)	$(R_{1740} - R_{2370}) / (R_{1740} - R_{1750})$	[17]
水带指数 Water band index (WBI)	R_{950} / R_{900}	[18]
叶水指数 Leaf water index (LWI)	R_{1300} / R_{1450}	[19]
归一化差异水分胁迫指数 Normalized different water stress index (NDWSI)	$(R_{850} - R_{970}) / (R_{850} + R_{970})$	[20]
归一化多波段干旱指数 Normalized multi-band drought index (NMDI)	$R_{860} - (R_{1640} - R_{2130}) / R_{860} + (R_{1640} - R_{2130})$	[21]
全球植被水分指数 Global vegetation moisture index (GVMI)	$(R_{820} + 0.1) - (R_{1600} + 0.02) / (R_{820} + 0.1) + (R_{1600} + 0.02)$	[22]
水分胁迫指数的倒数 Reciprocal of moisture stress index (RMSI)	R_{860} / R_{1650}	[23]

R: 反射率 Reflectivity; 例如 R_{860} 代表在 860 nm 波段的反射率 For example, R_{860} represents reflectance at 860 nm.

1.2.4 特征挑选 输入最少的特征变量得到最好的结果对于模型的构建具有重要意义。Relieff (recursive feature elimination with cross-validation) 是一种特征选择技术,常用于从高维数据中挑选最重要的特征。是一种基于实例的特征选择算法,广泛应用于高维数据的特征选择中,尤其适用于分类任务。它的核心思想是通过评估特征在区分不同类别实例时的重要性,从而选择最具代表性的特征。该方法在分类和回归任务中被广泛应用,尤其是在机器学习领域,用于提高模型的预测能力和减少过拟合^[24]。本研究在将数据输入机器学习算法构建模型前,采用 Relieff 对特征进行重要度排序,将重要值为正值特征投入到算法中。

1.2.5 数据处理及模型的构建与验证 采用 Excel 2019 和 MATLAB 14.0 (the MathWorks, Inc., Natick, MA) 软件进行基于光谱指数和机器学习的计算,使用 Origin 2021 软件进行制图。利用采集的光谱反射率数据计算光谱指数,选用偏最小二乘回归^[25]、随机森林算法^[26]、高斯过程回归 (Gaussian process regression, GPR)^[27] 3 种机器学习模型比较分析,并采用 Matlab 2021 软件建立估测模型。将 2023 年采集的 150 个玉米含水量数据用于模型的训练,2024 年采集的 62 个玉米叶片含水量数据用于模型的验证。使用决定系数 (coefficient of determination, R^2)、均方根误差 (root-mean-square error, RMSE) 和相对误差 (relative error, RE) 比较预测能力的差异和模型的稳定性。决定系数 R^2 越大, RMSE 和 RE 越小,表明模型估测 LWC 的精度和准确性越高。计算公式如下:

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

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

$$RE = \frac{RMSE}{\bar{y}} \times 100\% \quad (3)$$

式中: n 是样本数量; y_i 是实际值; \bar{y}_i 是目标值的平均值; \hat{y}_i 是预测值。

2 结果与分析

2.1 玉米叶片含水量的描述性统计

利用试验 1 中的数据进行描述性统计分析发现不同生育时期和层位之间的玉米 LWC 变异系数差异较大 (表 2), 说明生育时期、层位对玉米 LWC 有重要影响。叶片含水量在 VT 时期达到最大, 平均为 85.07%, 在 R1 生育时期最小, 平均为 73.24%, 随着生育时期的推进玉米 LWC 呈先升高后降低的趋势。在不同层位中, 上层的变异系数最大, 为 12.15%, 从上层到下层玉米 LWC 逐渐降低。从不同生育时期和层位的标准差来看, 除 R1 和上层外, 其他生育时期和层位的标准差和变异系数差异较小。标准差越大, 表示 LWC 变化的幅度越大, 数据的代表性越强, 进而能够提高模型构建的准确性。玉米叶片的光谱反射率随生育时期的变化体现了叶片的结构、叶绿素含量和水分含量等因素的动态变化。在可见光波段随着玉米从 V12 时期到 R1 时期的生长, 反射率呈逐渐增高的趋势 (图 2), 尤其是在 R1 时期时, 叶片黄化, 反射率进一步上升。叶片水分含量对光谱反射率的响应主要在近红外和短波红外波段, 尤其在 1400 和 1900 nm 波段附近具有明显的吸收峰, LWC 越高反射率越低, 表明 LWC 吸收带位于 1350~2500 nm, 水分的减少会影响光谱中的水分吸收特征, 导致短波红外区反射率增加 (表 1 和图 2)。

2.2 基于光谱指数的玉米叶片含水量估测

对不同灌溉处理下玉米 LWC 与 13 种水分敏感光谱指数的相关性进行了分析 (表 3)。从整体生育时期来看, 只有改良的 DATT 指数 (modified DATT, MDATT) 和水带指数 (water band index, WBI) 的决定系数超过了 0.5, 其他指数与 LWC 的相关性都相对较低, 预测能力不太理想。随着生育时期的推进, 不同光谱指数对各个生育时期 LWC 的估测能力存在明显差异, 对 R1 时期进行估测的 13 个指标评估中, 除 PRI 指数外其他光谱指数都取得了不错的预测效果, 而 VT 和 V12 时期的相关性不明显, 只有少数几个光谱指数在 V12 时期决定系数在 0.5 以上。从不同叶片层位来看, 层位同样对光谱指数与 LWC 的相关性具有一定的影响。相比于各生育时期, 光谱

表2 不同生育时期及层位玉米叶片含水量描述性统计

Table 2 Descriptive statistics of maize leaf water content at different growth stages and canopy layers (%)

生育时期及层位 Growth stage and canopy layer	最大值 Maximum	最小值 Minimum	平均值 Mean	标准差 Standard deviation	变异系数 Coefficient of variation
大喇叭口期 Twelfth leaf (V12)	84.50	70.50	77.72	3.68	5.64
抽雄期 Tasseling (VT)	89.94	80.91	85.07	2.10	2.58
开花吐丝期 Silking (R1)	83.66	61.45	73.24	7.05	10.21
上层叶片 Upper layer	89.71	61.45	74.58	8.70	12.15
中层叶片 Middle layer	86.63	71.62	78.47	4.19	5.17
下层叶片 Lower layer	89.94	74.52	82.99	3.27	4.62
全生育时期 All	89.94	61.45	78.66	6.84	9.53

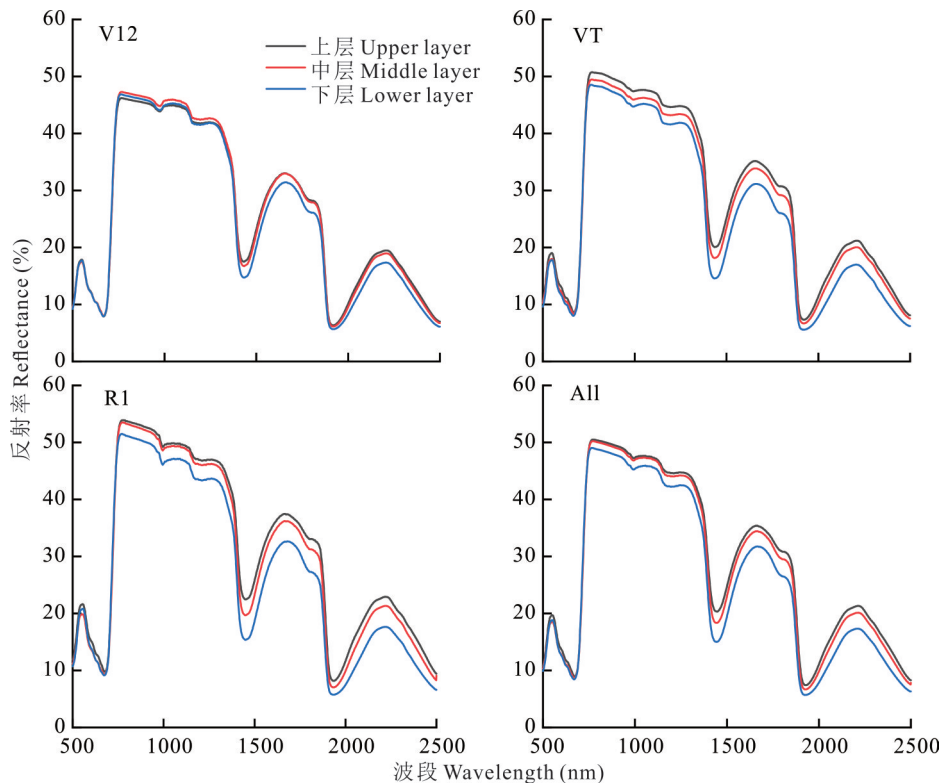


图2 不同生育时期和层位下叶片高光谱反射率

Fig. 2 Hyperspectral reflectance of leaves at different growth stages and layers

指数对不同层位 LWC 的估测能力整体较弱,只有生理反射指数(physiological reflectance index, PRI)指数的上、中层决定系数在 0.5 以上。为进一步比较光谱指数估测玉米 LWC 的性能,将相关性最好的 MDATT 指数与 LWC 在不同生育时期不同层位(图 3)下进行模型的构建。从模型中可以看出各个生育时期的 MDATT 和 LWC 的决定系数在 R1 时期对于 LWC 的估测能力最优。结果表明这 13 种光谱指数受到生育时期和层位的影响明显,对 LWC 的敏感性较低,导致其无法有效捕捉玉米 LWC 的变化。

2.3 基于机器学习算法的玉米叶片含水量估测

基于光谱指数与玉米 LWC 相关性分析结果,本研究发现传统的光谱指数预测方法难以处理复杂的非线性问题,且预测精度较低。而输入有效特征结合机器学习的方法在处理这一问题有很好的效率。因此,本研究采用 ReliefF 特征选择技术优化模型输入参数的数量,筛选重要值为正数的 MDATT、GVMI、WI、WBI、NDWI、

表 3 关键生育时期和不同层位叶片含水量与光谱指数的决定系数

Table 3 Coefficients of determination between leaf water content and spectral index at key growth stages and different canopy layers

光谱指数 Spectral index	大喇叭口期 V12	抽雄期 VT	开花吐丝期 R1	上层 Upper	中层 Middle	下层 Lower	全生育时期 All
水分指数 WI	0.57	0.22	0.65	0.17	0.08	0.12	0.36
归一化差异水分指数 NDWI	0.23	0.07	0.42	0.28	0.28	0.22	0.19
水分胁迫指数 MSI	0.15	0.03	0.75	0.07	0.13	0.07	0.06
简单比值水分指数 SRWI	0.57	0.22	0.65	0.17	0.08	0.12	0.36
植被水分指数 PWI	0.06	0.08	0.75	0.01	0.05	0.01	0.22
生理反射指数 PRI	0.06	0.05	0.15	0.67	0.51	0.38	0.29
改良的 DATT 指数 MDATT	0.64	0.18	0.76	0.38	0.20	0.32	0.52
水分胁迫指数的倒数 RMSI	0.58	0.06	0.72	0.02	0.02	0.01	0.18
水带指数 WBI	0.24	0.21	0.54	0.43	0.39	0.41	0.51
叶水指数 LWI	0.63	0.11	0.79	0.09	0.01	0.05	0.36
归一化差异水分胁迫指数 NDWSI	0.12	0.09	0.55	0.20	0.20	0.21	0.37
归一化多波段干旱指数 NMDI	0.11	0.04	0.60	0.01	0.02	0.01	0.13
全球植被水分指数 GVMI	0.60	0.06	0.75	0.04	0.04	0.01	0.15

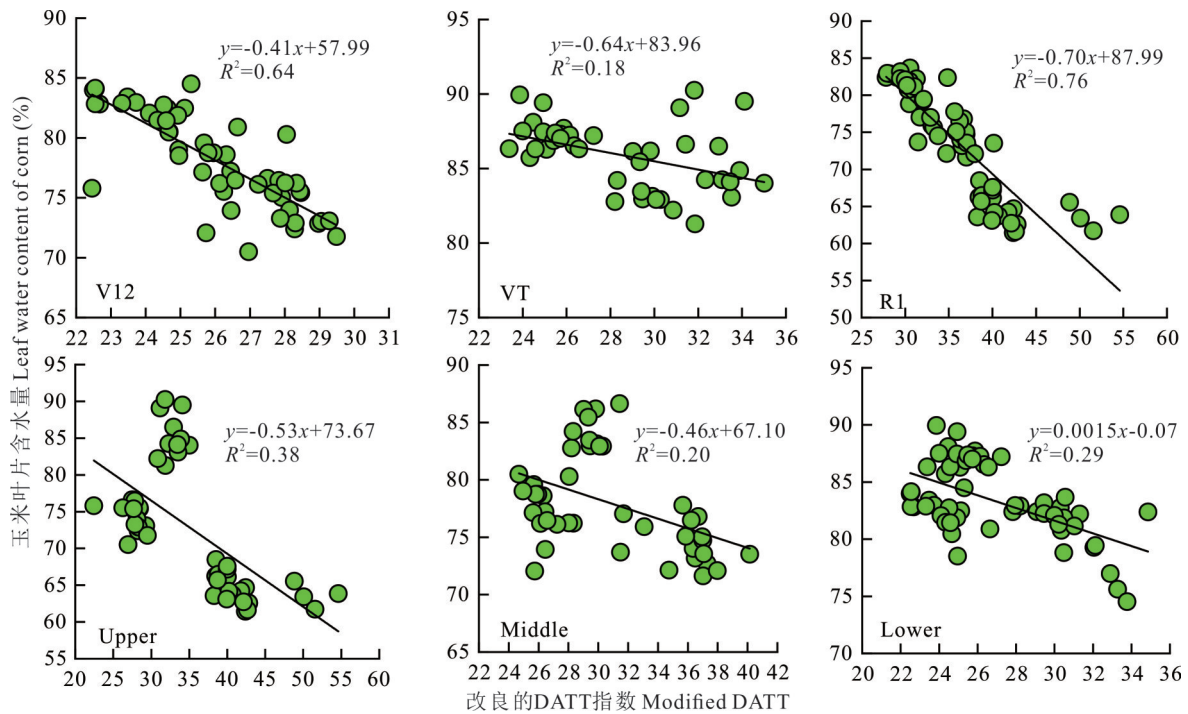


图 3 改良的 DATT 指数在不同生育时期及层位下与玉米叶片含水量的线性相关关系

Fig. 3 Linear correlation between modified DATT index and maize leaf water content at different growth stages and layers

LWI、NDWSI、NMDI、SRWI 这 9 种光谱指数 (图 4) 与全光谱波段 (350~2500 nm) 作为输入特征分别投入到偏最小二乘回归 (PLSR)、随机森林算法 (RF)、高斯过程回归 (GPR) 中构建玉米 LWC 监测模型 (图 5)。从模型表现来看, PLSR、RF 和 GPR 这 3 种方法对玉米 LWC 的预测均表现出较高的敏感性。光谱指数中表现最好的 MDATT 的决定系数为 0.52, 而基于机器学习算法的建模方法使 LWC 估测精度显著提升, 其中 PLSR、GPR 和 RF 的决定系数在 MDATT 的基础上分别提高 25%、44% 和 45% (图 5)。3 种建模方法中, RF 的建模效果最佳,

决定系数为 0.94 和 0.97,其次是 GPR,决定系数为 0.93 和 0.96,而 PLSR 的建模效果最差,决定系数仅为 0.59 和 0.77。比较两种特征变量的表现,筛选后的光谱指数作为输入特征的模型效果明显优于使用全波段光谱的模型。结果表明,RF 和 GPR 模型结合光谱指数在估算玉米 LWC 方面具有优越的预测性能,同时也证实了机器学习方法与有效因子的结合可以很好地解决农田尺度中非线性问题。

2.4 基于机器学习算法的玉米叶片含水量估测模型评价与验证

为了评估 3 种估测模型在不同输入变量下的表现,对基于 PLSR、RF、GPR 模型在不同输入特征下的玉米 LWC 估测能力进行了验证与评价(试验 2,图 6)。结果显示,当使用全波段光谱数据作为输入变量时 PLSR 估测模型的验证结果均不能令人满意,验证斜率偏移较大,导致该模型无法准确预测实际的玉米 LWC。相比之下,光谱指数作为输入变量在 GPR 和 RF 中均取得了较好的预测效果,决定系数明显提升,且 RMSE 和 RE 也相对较低。总体而言,RF 和 GPR 模型的性能都较为优越,尤其是在以光谱指数为特征输入变量的模型中,RF 模型验证集决定系数达到 0.89, RMSE 和 RE 分别为 1.95% 和 2.01%, GPR 模型验证集决定系数达到 0.88, RMSE 和 RE 分别为 2.04% 和 2.09%。这些结果表明,RF 和 GPR 模型能够有效解决估测模型中的非线性问题,且具有较强的适应性和高精度,能够有效应用于实际生产中。

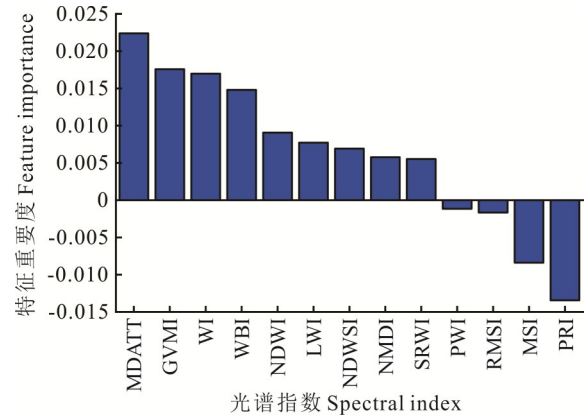


图 4 基于 ReliefF 特征重要度排序

Fig. 4 Sorting based on ReliefF feature importance

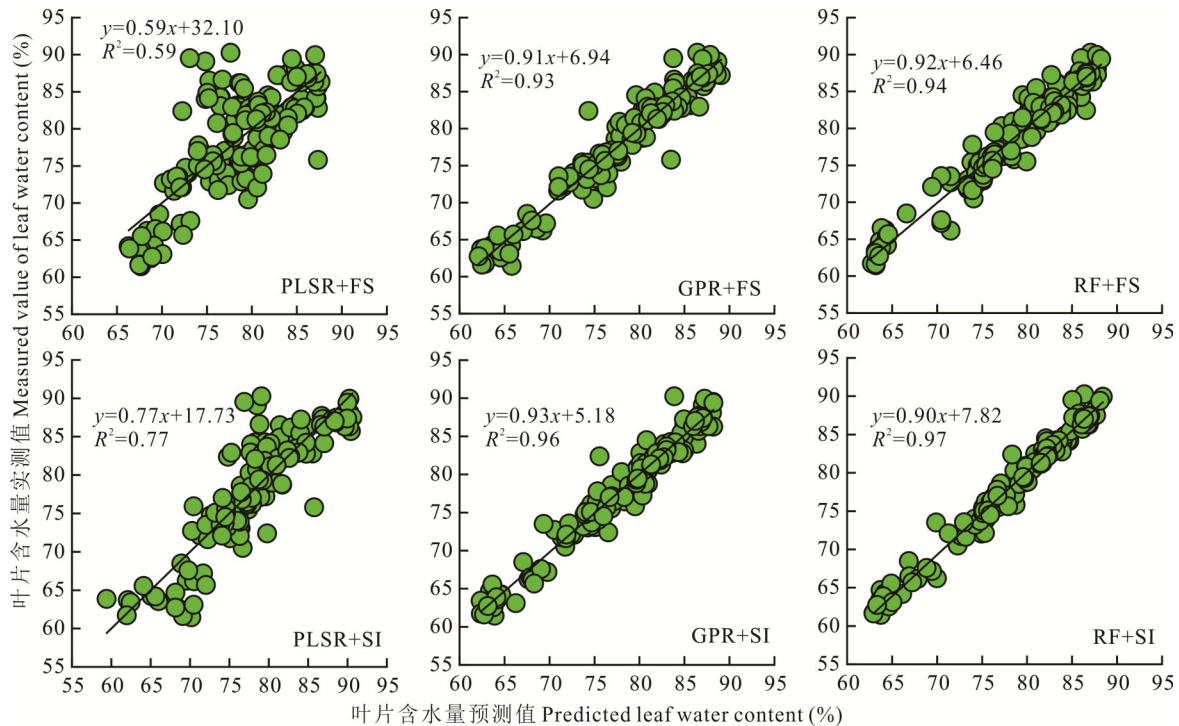


图 5 基于全光谱数据集和光谱指数的玉米叶片含水量估测模型构建

Fig. 5 Construction of maize leaf moisture estimation model based on full-spectrum dataset and spectral index

PLSR: 偏最小二乘回归 Partial least squares regression; GPR: 高斯过程回归 Gaussian process regression; RF: 随机森林 Random forest; FS: 全波段: Full-spectrum; SI: 光谱指数 Spectral index. 下同 The same below.

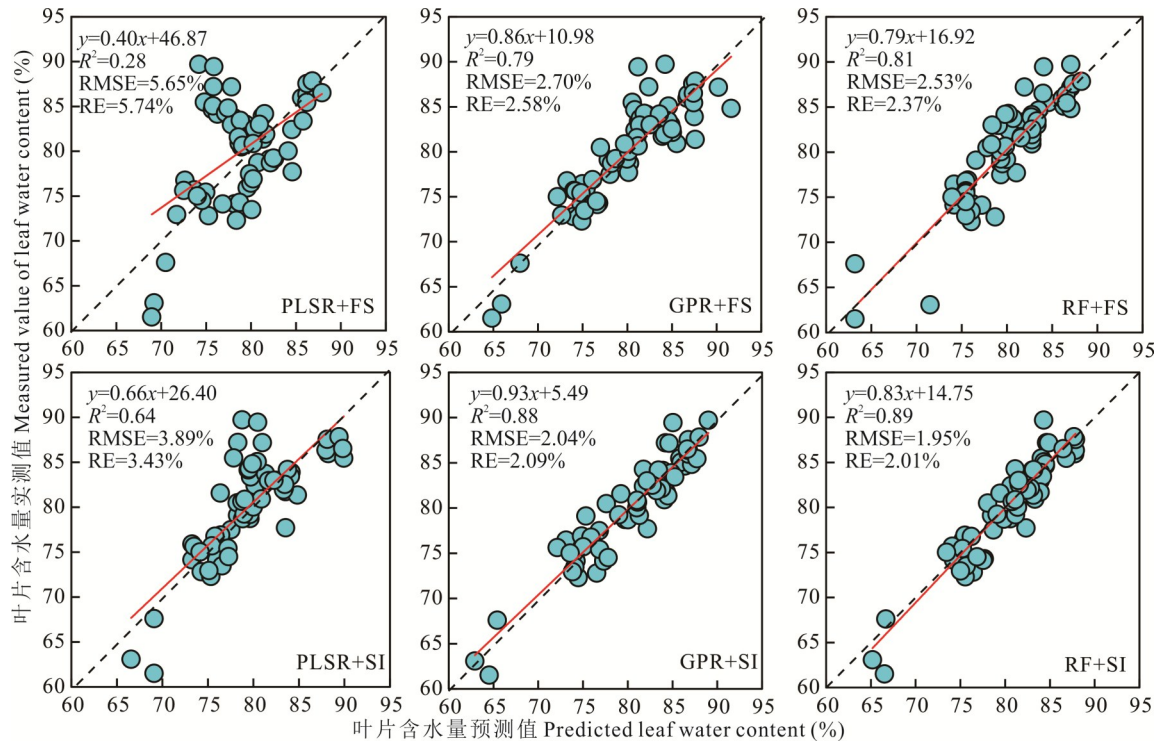


图 6 基于原始光谱数据集和光谱指数的玉米叶片含水量估测模型的验证

Fig. 6 Validation of the maize leaf moisture estimation model based on full-spectrum dataset and spectral index

RMSE: 均方根误差 Root mean squared error; RE: 相对误差 Relative error.

3 讨论

3.1 波段反射率与叶片水分状态的关系

水分的亏缺直接影响青贮玉米的生长发育,作为表征植物水分代谢动态的关键生理指标,LWC的精准监测,可为解析水分利用效率的基因型差异及构建水分限制条件下的作物生产力预测模型提供量化依据^[28]。高光谱反射技术通过捕获 350~2500 nm 波段[涵盖可见光(visible, VIS): 400~700 nm;近红外(near infrared, NIR): 700~1300 nm;短波红外(short-wave infrared, SWIR): 1300~2500 nm]的连续光谱信息,能够非破坏性地反演作物水分状态。并且,近红外和短波红外区域的冠层反射率受到冠层内部叶片结构和含水量的强烈影响。因此,它成功地用于追踪与植物水分状况相关的植物参数的变化,并监测由作物水分间接引起的生物量和叶面积的变化^[29]。在本研究中,玉米叶片的光谱反射率随生育时期的推进逐渐升高,原因可能是在营养生长期玉米叶片的叶绿素浓度增高导致可见光波段的吸收增强,反射率较低,近红外波段的反射率逐渐增高。而在短波红外区域,已有文献表明,作物叶片中的水量与整个光谱中的叶片反射率在统计上具有显著相关性^[30],具体来说,叶细胞中水分子的变化强烈影响近红外区域的太阳辐射反射率,这部分光谱常用于量化叶片水分状况^[31]。因此,在短波红外区域,叶片含水量越高,光谱反射率越低,与描述性统计结果一致。

3.2 光谱指数在玉米叶片含水量预测中的稳定性

通过分析玉米叶片含水量(LWC)与光谱响应特征的内在联系,可为作物水分生理诊断及精准灌溉决策提供理论依据。本研究系统评估了 13 种水分敏感光谱指数与玉米 LWC 的定量关联特征,发现基于全生育期构建的 LWC 光谱监测模型预测性能明显受限,且不同生育阶段谱效关系的异质性特征明显。原因可能是在生长过程中植被生长状况和生育阶段等因素都会导致作物对光谱响应的差异^[32],不同光谱指数在反演玉米不同生育期 LWC 时的精度存在显著差异。研究同时证实,光谱指数响应效能存在显著的空间异质性和基因型依赖性。张海威等^[33]通过分析 8 种光谱指数与植被 LWC 的关系,得出 GVMi 指数反演效果最佳,而本研究中 GVMi 指数预测效果并不理想,这可能是由于不同植物种类的叶片在光谱波段(特别是在近红外和红边区域)的反射率存在差异。

即使是相同植物,不同品种的光谱响应也可能不同,因此,针对不同作物或品种,可能需要不同的光谱指数或模型来进行LWC的预测。尽管已有许多水情指数被开发用于估测LWC,但很难找到一种在所有生育时期均能提供稳定、准确预测的水情指数。刘仁华等^[34]研究表明,玉米生长过程中叶片含水量、绿度、结构均在变化,光谱指数预测精度也随之波动。因此,光谱指数的选择和预测模型的适用性需要考虑不同试验地点、气候和品种的影响,导致其无法反映叶片含水量的动态变化,影响预测的时效性和准确性。

3.3 机器学习算法在玉米叶片含水量预测中的稳定性

机器学习技术凭借其非线性建模优势及数据驱动特性,已成为作物LWC高精度反演的核心方法。现有研究表明,基于灰色关联分析筛选的敏感光谱指数结合偏最小二乘回归与反向传播神经网络,可实现冬小麦LWC的定量预测(决定系数 $R^2=0.72\sim 0.80$)^[35-37]。然而,传统特征选择方法对高维光谱数据的非线性交互作用捕捉不足,且易受冗余波段干扰,导致模型泛化能力受限。本研究创新性地引入ReliefF特征权重算法,通过评估特征与目标变量的非线性关联度,筛选了重要值为正值的9种光谱指数,其中重要值排名第一的MDATT指数与本研究中玉米LWC相关性最佳的光谱指数一致,说明ReliefF技术能够准确地筛选出对目标变量最有影响的特征,并且能有效降低特征空间的维度,提高模型的训练效率和预测性能^[24]。随后,将全波段光谱和筛选后的9种光谱指数作为输入特征,分别投入到3种机器学习算法中对玉米LWC进行建模,并比较3种机器学习模型的预测性能。本研究发现输入变量维度与模型架构的匹配度明显影响了机器学习模型在估测玉米LWC中的性能。前人的研究表明利用合适的光谱指数与随机森林结合可以成功预测小麦生物量^[38],但由于其未能对输入特征进行敏感指数筛选而导致预测精度不高。在本研究中,与使用全波段2151条光谱波段相比,经过筛选后的9种光谱指数仅用15个光谱波段就能达到97%的解释能力,且验证集的RMSE、RE也相对较低。并且,不同机器学习模型的性能表现也存在差异。PLSR模型对特征冗余敏感,全波段输入时因多重共线性导致决定系数 R^2 仅为0.59,而光谱指数输入后提升至0.77(决定系数 $\Delta R^2=+0.18$),但其非线性建模能力不足,性能仍低于随机森林(RF)与高斯过程回归(GPR)。有研究表明与BP和SVR模型相比,PLSR算法在预测小麦LWC方面表现最佳(决定系数 $R^2=0.85$)^[9]。而本研究中RF和GPR模型为玉米LWC监测提供了最佳预测,PLSR表现较差。同样本研究发现RF和GPR所构建的模型中,不同特征的表现差异相对较小,而PLSR在不同特征下的建模效果差异较大,可能是由于PLSR在一定程度上解决了模型构建过程中建模因子之间的共线性问题,而在解决非线性问题方面的能力仍有待提高。而RF和GPR模型都具有较强的鲁棒性,能够有效地处理大量特征且避免过拟合,并且在高维度数据中,往往能够自动选择最有影响力的特征,从而提高预测精度。其与光谱指数的结合可以有效地降低机器学习输入变量的维数和多重共线性问题,是预测LWC最有前景的方法。

在作物水分状况的预测中,选择合适的输入特征对机器学习算法非常重要。本研究利用光谱指数结合机器学习构建的玉米LWC估测模型来源于一个为期2年的田间试验数据。其实用性和稳定性仍然需要进行多年多点多生育时期的田间试验分析,未来将结合多年度数据进一步优化模型,并针对年际环境差异设计自适应校准模块,利用足够量的数据进一步构建玉米LWC的估测模型,以提高光谱技术对植株LWC的诊断能力。

4 结论

作物叶片含水量是衡量作物生长过程重要的营养指标之一。为了深入探究光谱指数结合机器学习在LWC估测中的应用。本研究评估了水情指数在不同生育时期和不同层位的预测效果,并将光谱指数和全波段光谱投入3种机器学习中进行建模,以期进一步提高叶片LWC的预测能力。从结果来看,以特征挑选后的光谱指数在模型构建中作为输入变量相比于全波段光谱在玉米LWC估测中具有更好的表现,并且3种机器学习方法中RF和GPR算法在非线性回归中都展现了较高的估测精度,RF训练集决定系数 R^2 为0.97,验证集 R^2 为0.89, RMSE和RE也最低,分别为1.95%和2.01%,GPR模型次之,验证集决定系数 R^2 为0.88, RMSE和RE分别为2.04%和2.09%。综合评价结果证明,RF和GPR模型与光谱指数的结合可以很好地解决农田尺度中LWC估测中非线性问题。

参考文献 References:

- [1] Ferraretto L F, Shaver R D, Luck B D. Silage review: Recent advances and future technologies for whole-plant and fractionated corn silage harvesting. *Journal of Dairy Science*, 2018, 101(5): 3937–3951.
- [2] Rouse Jr J W, Haas R H, Deering D W, *et al.* Monitoring the vernal advancement and retrogradation. TX, USA: Texas A & M University: College Station, 1974.
- [3] Inoue Y, Morinaga S, Shibayama M. Non-destructive estimation of water status of intact crop leaves based on spectral reflectance measurements. *Japanese Journal of Crop Science*, 1993, 62(3): 462–469.
- [4] Zhao S H, Wang Q, Yao Y J, *et al.* Estimating and validating wheat leaf water content with three MODIS spectral indexes: A case study in Ningxia Plain, China. *Journal of Agricultural Science and Technology*, 2016, 18(2): 387–398.
- [5] Rapaport T, Hochberg U, Cochavi A, *et al.* The potential of the spectral ‘water balance index’ (WABI) for crop irrigation scheduling. *New Phytologist*, 2017, 216(3): 741–757.
- [6] Kunz K, Hu Y, Schmidhalter U. Carbon isotope discrimination as a key physiological trait to phenotype drought/heat resistance of future climate-resilient German winter wheat compared with relative leaf water content and canopy temperature. *Frontiers in Plant Science*, 2022, 13: 1043458.
- [7] Zhang J, Zhang W, Xiong S, *et al.* Comparison of new hyperspectral index and machine learning models for prediction of winter wheat leaf water content. *Plant Methods*, 2021, 17(1): 1–14.
- [8] Ndlovu H S, Odindi J, Sibanda M, *et al.* A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)—based proximal and remotely sensed data. *Remote Sensing*, 2021, 13(20): 4091.
- [9] Guo F, Feng Q, Yang S, *et al.* Estimation of potato canopy leaf water content in various growth stages using UAV hyperspectral remote sensing and machine learning. *Frontiers in Plant Science*, 2024, 15: 1458589.
- [10] Yang C, Tan Y, Bruzzone L, *et al.* Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. *Remote Sensing*, 2017, 9(8): 782.
- [11] Gao B C. Normalized difference water index for remote sensing of vegetation liquid water from space. // *Proceedings of the Imaging Spectrometry Conference*. SPIE, 1995, 2480: 225–236.
- [12] Peñuelas J, Pinol J, Ogaya R, *et al.* Estimation of plant water concentration by the reflectance water index WI (R900/R970). *International Journal of Remote Sensing*, 1997, 18(13): 2869–2875.
- [13] Helskanen J. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *International Journal of Remote Sensing*, 2006, 27(6): 1135–1158.
- [14] Zarco-Tejada P J, Rueda C, Ustin S. Water content estimation in vegetation with MODIS reflectance data and model inversion method. *Remote Sensing of Environment*, 2003, 85(1): 109–124.
- [15] Apan A, Held A, Phinn S, *et al.* Formulation and assessment of narrow-band vegetation indices from EO-1 Hyperion imagery for discriminating sugarcane disease. // *Proceedings of the Spatial Sciences Institute Biennial Conference (SSC 2003): Spatial Knowledge Without Boundaries*, Canberra, Australia: Spatial Sciences Institute, 2003.
- [16] Gamon J, Penuelas J, Field C. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 1992, 41(1): 35–44.
- [17] Wang Z, Sun Z, Lu S. Optimal vegetation index for assessing leaf water potential using reflectance factors from the adaxial and abaxial surfaces. *Computers and Electronics in Agriculture*, 2020, 172: 105337.
- [18] Peñuelas J, Filella I, Biel C, *et al.* The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*, 1993, 14(10): 1887–1905.
- [19] Seelig H D, Hoehn A, Stodieck L S, *et al.* Plant water parameters and the remote sensing R1300/R1450 leaf water index: Controlled condition dynamics during the development of water deficit stress. *Irrigation Science*, 2009, 27(5): 357–365.
- [20] Hunt Jr E R, Daughtry C S T, Qu J J, *et al.* Comparison of hyperspectral retrievals with vegetation water indices for leaf and canopy water content. // *Proceedings of the Remote Sensing and Modeling of Ecosystems for Sustainability VIII*. SPIE, 2011, 8156: 66–76.
- [21] Wang L, Qu J J. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters*, 2007, 34(20): 20405.
- [22] Ceccato P, Flasse S, Gregoire J M. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote Sensing of Environment*, 2002, 82(2/3): 198–207.
- [23] Hunt Jr E R, Rock B N. Detection of changes in leaf water content using near-and middle-infrared reflectances. *Remote Sensing*

- of Environment, 1989, 30(1): 43–54.
- [24] Zhang B S, Li Y Y, Chai Z. A novel random multi-subspace based ReliefF for feature selection. Knowledge-Based Systems, 2022, 252: 109400.
- [25] Cheng B, Wu X. A modified PLSR method in prediction. Journal of Data Science, 2006, 4(3): 257–274.
- [26] Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24–31.
- [27] Nguyen-Tuong D, Seeger M, Peters J. Model learning with local Gaussian process regression. Advanced Robotics, 2009, 23(15): 2015–2034.
- [28] El-Hendawy S E, Al-Suhaibani N A, Elsayed S, *et al.* Potential of the existing and novel spectral reflectance indices for estimating the leaf water status and grain yield of spring wheat exposed to different irrigation rates. Agricultural Water Management, 2019, 217: 356–373.
- [29] Wang X, Zhao C, Guo N, *et al.* Determining the canopy water stress for spring wheat using canopy hyperspectral reflectance data in loess plateau semiarid regions. Spectroscopy Letters, 2015, 48(7): 492–498.
- [30] Wijewardana C, Alsajri F A, Irby J T, *et al.* Physiological assessment of water deficit in soybean using midday leaf water potential and spectral features. Journal of Plant Interactions, 2019, 14(1): 533–543.
- [31] Pasqualotto N, Delegido J, Van Wittenberghe S, *et al.* Retrieval of canopy water content of different crop types with two new hyperspectral indices: Water absorption area index and depth water index. International Journal of Applied Earth Observation and Geoinformation, 2018, 67: 69–78.
- [32] Liu X J, Chen G Q, Wang L, *et al.* Monitoring leaf relative water content of winter wheat based on hyperspectral index at different growth stages. Journal of Triticeae Crops, 2018, 47(7): 854–862.
刘晓静, 陈国庆, 王良, 等. 不同生育时期冬小麦叶片相对含水量高光谱监测. 麦类作物学报, 2018, 47(7): 854–862.
- [33] Zhang H W, Zhang F, Zhang X L, *et al.* Inversion of vegetation leaf water content based on spectral index. Spectroscopy and Spectral Analysis, 2018, 38(5): 1540–1546.
张海威, 张飞, 张贤龙, 等. 光谱指数的植被叶片含水量反演. 光谱学与光谱分析, 2018, 38(5): 1540–1546.
- [34] Liu R H, Zhou G S, Zhou L, *et al.* Remote sensing inversion of leaf and canopy water content in different growth stages of summer maize. Journal of Applied Meteorological Science, 2020, 31(1): 52–62.
刘仁华, 周广胜, 周莉, 等. 夏玉米不同生育期叶片和冠层含水量的遥感反演. 应用气象学报, 2020, 31(1): 52–62.
- [35] Jin X, Xu X, Song X, *et al.* Estimation of leaf water content in winter wheat using grey relational analysis-partial least squares modeling with hyperspectral data. Agronomy Journal, 2013, 105(5): 1385–1392.
- [36] Hasan U, Sawut M, Kasim N, *et al.* Hyperspectral estimation model of leaf water content in spring wheat based on grey relational analysis. Spectroscopy and Spectral Analysis, 2018, 38(12): 3905–3911.
- [37] Zhang J, Tian H Q, Zhao Z Y, *et al.* Moisture content detection in silage maize raw material based on hyperspectrum and improved discrete particle swarm. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(1): 285–293.
张珏, 田海清, 赵志宇, 等. 基于改进离散粒子群算法的青贮玉米原料含水率高光谱检测. 农业工程学报, 2019, 35(1): 285–293.
- [38] Wang L, Zhou X, Zhu X, *et al.* Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal, 2016, 4(3): 212–219.