

潘月, 曹宏鑫, 齐家国, 等. 基于高光谱和数据挖掘的油菜植株含水率定量监测模型[J]. 江苏农业学报, 2022, 38(6): 1550-1558.  
doi: 10.3969/j.issn.1000-4440.2022.06.013

## 基于高光谱和数据挖掘的油菜植株含水率定量监测模型

潘月<sup>1,2</sup>, 曹宏鑫<sup>2</sup>, 齐家国<sup>1</sup>, 吴菲<sup>2,3</sup>, 韩旭杰<sup>1,2</sup>, 丁昊迪<sup>1,2</sup>, 葛道阔<sup>2</sup>, 张玲玲<sup>2</sup>, 张伟欣<sup>2</sup>, 张文字<sup>2</sup>

(1. 南京农业大学农学院/亚洲农业研究中心, 江苏 南京 210095; 2. 江苏省农业科学院农业信息研究所, 江苏 南京 210014; 3. 扬州大学农学院, 江苏 扬州 225009)

**摘要:** 为了构建监测效果更好、更具普适性的油菜植株含水率(Plant water content, *PWC*)定量监测模型,以油菜品种浙杂903、宁油22和宁杂1818为试验材料,设置2个施肥水平和3个水分处理,基于2019–2020年和2020–2021年生长季田间试验资料,在*PWC*的高光谱响应敏感波段范围采用逐步回归(Stepwise regression, SR)分析、连续投影算法(Successive projection algorithm, SPA)、竞争自适应加权算法(Competitive adaptive reweighted sampling, CARS)以及减量精细采样法(Reduced precise sampling method, RPSM)深度挖掘高光谱数据,通过筛选最优波段组合与光谱指数,基于线性回归(Linear regression, LR)、BP神经网络(Back-propagation neural network, BPNN)和支持向量机回归(Support vector regression, SVR)方法构建并比较油菜植株含水率监测模型。结果表明,针对油菜*PWC*监测,SR分析筛选的最优波段组合为730 nm、986 nm和1 071 nm,SPA法分析筛选的最优波段组合为686 nm、695 nm、707 nm、746 nm、964 nm、1 065 nm和1 069 nm,CARS法分析筛选的最优波段组合为694 nm、695 nm、696 nm、863 nm、864 nm、893 nm、973 nm、986 nm、1 050 nm和1 071 nm。RPSM筛选的最优光谱指数是归一化差值光谱指数(R981,R894)和比值光谱指数(R981,R894),其利用的波段均位于近红外波段。前述3个方法筛选的波段变量更多,蕴含的信息更全面,估测精度普遍优于光谱指数。建模分析结果表明,SPA-LR模型、SPA-BP模型、SPA-SVR模型均能实现油菜*PWC*的精确监测,经检验,其估测值和实测值的 $R^2$ 分别为0.693、0.940、0.841,均方根误差(RMSE)分别为1.623%、1.836%和1.227%。结果证明高光谱数据具备深度挖掘价值,运用全波段光谱分析方法能够在降维的同时保留有效信息,利用筛选出的波段组合构建线性或非线性模型,均能实现大田条件下全生育期油菜植株含水率的定量监测。

**关键词:** 高光谱; 油菜; 连续投影算法; 竞争自适应加权算法; BP神经网络

**中图分类号:** S126 **文献标识码:** A **文章编号:** 1000-4440(2022)06-1550-09

## Quantitative monitoring models of plant water content in rapeseed based on hyperspectrum and related data mining

PAN Yue<sup>1,2</sup>, CAO Hong-xin<sup>2</sup>, QI Jia-guo<sup>1</sup>, WU Fei<sup>2,3</sup>, HAN Xu-jie<sup>1,2</sup>, DING Hao-di<sup>1,2</sup>, GE Dao-kuo<sup>2</sup>, ZHANG Ling-ling<sup>2</sup>, ZHANG Wei-xin<sup>2</sup>, ZHANG Wen-yu<sup>2</sup>

(1. College of Agriculture/Asia Hub on Agriculture, Nanjing Agricultural University, Nanjing 210095, China; 2. Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China; 3. College of Agriculture, Yangzhou University, Yangzhou 225009, China)

收稿日期: 2022-04-01

基金项目: 国家自然科学基金项目(31471415、31871522); 江苏省农业科技自主创新资金项目[CX(19)2040-1]

作者简介: 潘月(1997-),女,安徽芜湖人,硕士研究生,主要从事农业定量遥感研究。(E-mail)panyue1007@163.com

通讯作者: 曹宏鑫, (E-mail) caohongxin@hotmail.com; 齐家国, (E-mail) qi@msu.edu

**Abstract:** To construct a quantitative monitoring model for plant water content (*PWC*) of rapeseed with relative better monitoring effect and more universality, rapeseed cultivars Zheza 903, Ningyou 22, and Ningza 1818 were used as the experimental materials in this study, two fertilization levels and three water treatments were set. Based on field test data in growing seasons of 2019–2020 and 2020–2021, the hyperspectral

data were deeply mined by stepwise regression (SR) analysis, successive projection algorithm (SPA), competitive adaptive reweighted sampling (CARS) and reduced precise sampling method (RPSM), within the sensitive band range of hyperspectral response of *PWC*. By screening the optimal band combination and spectral index, the monitoring models of the rapeseed *PWC* were constructed and compared based on linear regression (LR), back-propagation neural network (BPNN), and support vector machine regression (SVR). The results showed that, for the monitoring of rapeseed *PWC*, the optimal bands combination by SR analysis was 730 nm, 986 nm and 1 071 nm, the optimal bands combination by SPA method was 686 nm, 695 nm, 707 nm, 746 nm, 964 nm, 1 065 nm and 1 069 nm, and the optimal bands combination by CARS method was 694 nm, 695 nm, 696 nm, 863 nm, 864 nm, 893 nm, 973 nm, 986 nm, 1 050 nm and 1 071 nm. The optimal spectral indices screened by RPSM were reduced precise sampling method (*NDSI*) (R981, R894) and ratio spectral index (*RSI*) (R981, R894), all the utilized bands were located in the near-infrared band region. The above three methods screened relatively more band variables, contained relatively more comprehensive information, and the estimation accuracies were generally better than spectral indices. Analysis results of modeling showed that, the SPA-LR model, SPA-BP model and SPA-SVR model could realize accurate monitoring of rapeseed *PWC*. Through testing,  $R^2$  of estimated value and measured value were 0.693, 0.940 and 0.841, respectively, and root mean square error (*RMSE*) were 1.623%, 1.836% and 1.227%, respectively. This study proved that, the hyperspectral data have the value of deep mining, and spectral analysis method based on full-band can reduce dimensionality while retaining effective information. Quantitative monitoring of rapeseed *PWC* in the whole growth period under field conditions can all be realized by constructing a linear or nonlinear model using the selected band combinations.

**Key words:** hyperspectral; rapeseed; successive projection algorithm; competitive adaptive reweighted sampling; BP neural network

油菜是中国种植面积最大的油料作物之一和国产植物油第一大油源,2020 年中国油菜种植面积约  $6.80 \times 10^6 \text{ hm}^2$ ,油菜籽产量  $1.400 \times 10^7 \text{ t}$ ,仅次于加拿大,居世界第二,发展油菜生产是保障中国食用油供给安全的根本途径<sup>[1]</sup>。油菜是需水量较大的作物<sup>[2-4]</sup>,植株水分状况不仅影响油菜生长<sup>[5]</sup>,还会影响其产量、氮素吸收能力以及油菜籽的品质<sup>[6-9]</sup>。而伴随水资源日益紧张,传统的“丰水高产”农业生产模式逐渐向现代“节水高产”模式转变,水分动态定量诊断与调控是实现这个转变的关键技术,其基础则是植株水分状况的定量监测。高光谱手段能够采集精细的光谱信息,其在监测植物水分状况中的应用十分广泛。运用遥感技术监测植物中的水分,学者更倾向于选择冠层和叶片指示水分状况<sup>[10]</sup>。随着传感传输技术的快速发展,遥感技术能够改变以往费时费力、破坏性大的诊断方法,实现植物水分实时监测、精确诊断。在单叶尺度方面,杨玉清等<sup>[11]</sup>用偏最小二乘法(Partial least squares, PLS)及逐步多元线性回归(Stepwise multiple linear regression, SMLR)法对玉米叶片水分含量进行建模分析,验证集决定系数分别为 0.975 和 0.980; Krishna 等<sup>[12]</sup>研究了 10 个水稻基因型在不同水分胁迫下的表现,通过光谱指数、多元技术和神经网络筛选最优波段,建立预测模型,结果表明,偏最小二乘回归-多元线性回归(PLSR-MLR)是预测叶片含水率

(RWC)的最优模型,验证  $R^2$  为 0.97。在冠层尺度方面,刘晓静等<sup>[13]</sup>的研究结果表明,在拔节期、抽穗期、开花期、灌浆前期和灌浆后期,冬小麦叶片相对含水量与归一化差值植被指数(*NDVI*)、优化土壤调整植被指数(*OSAVI*)、比值/归一化植被指数(*R/ND*)、冠气温差( $T_{\text{DC-a}}$ )和冠气温差( $T_{\text{DC-a}}$ )拟合效果较好,决定系数分别为 0.842、0.884、0.831、0.864 和 0.945。精确监测作物水分状况的主要步骤包括分析光谱响应规律、筛选模型变量、构建监测模型和检验监测模型。在油菜水分状况监测方面,张晓东等<sup>[14-15]</sup>基于高光谱数据,采用区间分段逐步回归法定量分析油菜含水率,模型预测值与实测值间的相关系数为 0.87,后续与多光谱图像和冠层温度多信息融合,将相关系数提高至 0.93;全春艳等<sup>[16]</sup>提出 2 种改进型角度指数  $ANI_{1450}$  和  $ASI_{1450}$ ,在不区分苗期、蕾薹期的情况下估算叶片等效水厚度效果最好, $R^2$  均达到 0.832;张君等<sup>[17]</sup>的研究结果表明,正交信号校正(OSC)+主成分分析(PCA)+SVR 模型估测叶片含水率效果最好,训练集决定系数( $R_c^2$ )与测试集决定系数( $R_v^2$ )分别高达 0.901 和 0.857。可见,已有研究多能够在单一生育期实现精确监测,而对于整个生育期的普适性有待提高,目前光谱分析方法在作物水分监测方面运用得较少,高光谱信息有待深度挖掘,机器学习方法在模型构建上应用得较少,且局限于实验室条件。本研究

旨在基于不同品种、氮素与水分处理下的油菜池栽田间试验,利用逐步回归分析法(Stepwise regression, SR)、连续投影算法(Successive projection algorithm, SPA)和竞争自适应加权算法(Competitive adaptive reweighted sampling, CARS)筛选最优波段组合,并利用减量精细采样法(Reduced precise sampling method, RPSM)筛选最优归一化差值光谱指数(*NDSI*)和比值光谱指数(*RSI*),为监测模型提供变量,利用线性回归(Linear regression, LR)、BP 神经网络(Back-propagation neural network, BPNN)和支持向量机回归(Support vector regression, SVR)方法构建油菜植株含水率(*PWC*)监测模型,以进一步提高油菜 *PWC* 监测精度,为油菜水分诊断调控提供依据。

## 1 材料与方法

### 1.1 试验设计

试验于 2019–2020 年和 2020–2021 年在江苏省农业科学院实验农场(32.03°N, 118.87°E)进行,前茬为休闲,两季试验的基本情况如表 1 所示。供试品种 3 个,分别为浙杂 903(C1)、宁油 22(C2)和宁杂 1818(C3),施肥水平分别为不施肥(N0)、施纯氮 180 kg/hm<sup>2</sup>(N2),移栽密度统一为 1 hm<sup>2</sup>1.2×10<sup>5</sup>株(行距 40 cm,株距 20 cm)。土壤基础养分:有机质含量 31.4 g/kg,全氮含量 2.03 g/kg,速效磷含量 20.3 mg/kg,速效钾含量 139 mg/kg,pH 7.31。

2019–2020 年供试品种为浙杂 903(C1)和宁杂 1818(C3),2020–2021 年供试品种为宁油 22(C2)

和宁杂 1818(C3)。采用裂区设计,主区为施肥,设为不施肥(N0)和施肥(N2,于移栽前施入有机肥 60 kg/hm<sup>2</sup>,于抽薹期叶面喷硼 15 kg/hm<sup>2</sup>,氮肥 180 kg/hm<sup>2</sup>按基肥:腊肥:薹肥=5:3:2 分配)处理,副区为水分与品种处理,水分处理设浇水 2 min(约小区水层 4.87 mm)(W1)、3 min(约小区水层 7.31 mm)(W2)、4 min(约小区水层 9.75 mm)(W3) 3 个水平。2020–2021 年浇水后不同处理 0~10 cm 耕层土壤体积含水率变化情况如图 1 所示,共 12 个处理,重复 3 次,随机排列。小区面积 3.77 m×2.77 m≈10.44 m<sup>2</sup>,行距 0.4 m,株距 20.0 cm。分别于 2019 年 10 月 1 日、2020 年 10 月 7 日播种,分别于 2019 年 11 月 2 日、2020 年 11 月 11 日移栽,其他栽培措施同高产大田管理。

表 1 2 个生长季试验的基本情况

Table 1 Basic situation of experiments in two growing seasons

年份	生育期	采样日期 (月-日)	小区数 (个)	样本数 (个)
2019–2020	越冬期	01-18	4	12
2020–2021	苗期	11-18	4	12
	苗期	11-30	4	12
	苗期	12-15	12	36
	苗期	01-03	12	36
	越冬期	01-19	12	36
	终花期	04-09	12	36

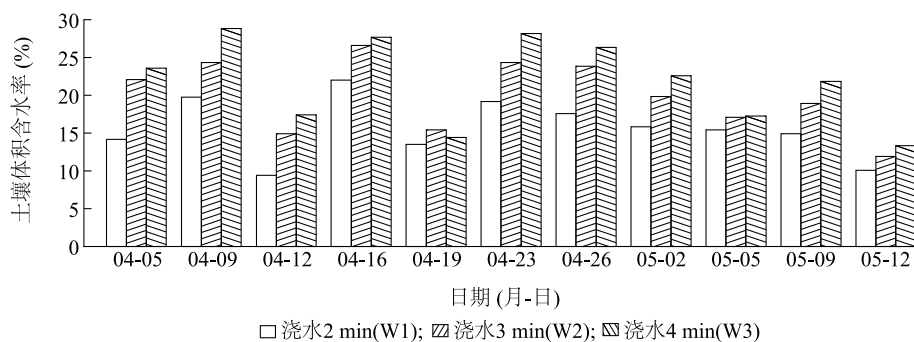


图 1 2020–2021 年不同处理 0~10 cm 耕层土壤体积含水率变化

Fig.1 Variation of volume water content of 0-10 cm topsoil under different treatments in 2020–2021

### 1.2 测定项目与方法

1.2.1 冠层光谱反射率 选择晴朗无云无风天气,分别于各个生育时期当天 10:00–14:00 在每个小区

选 3 个代表植株并挂牌标记序号,每个植株测定 5 个数据,取平均值代表该小区的光谱数据。用美国 ASD(Analytical Spectral Device)公司的 ASD Hand-

held 2 型手持式光谱分析仪距油菜冠层上方 50 cm 测定冠层光谱反射率,光谱仪视角为 25°,测定波段为 325~1 075 nm,光谱采样间隔为 1.0 nm,光谱分辨率为 3 nm,每次测量前用白板校正。

1.2.2 油菜植株含水率(*PWC*) 基本与光谱测定同步,分别在 2019-2020 年越冬期(移栽后 77 d)、2020-2021 年苗期(移栽后 7 d、19 d、34 d、53 d)、越冬期(移栽后 69 d)、终花期(移栽后 149 d)取样,每小区取上述挂牌标记植株 3 株。鲜质量利用称质量法,样本称量鲜质量后,恒温 105 ℃杀青 30 min,之后恒温 80 ℃烘干至恒质量,分别称量样本干质量。计算植株含水率,  $PWC = (PWF - PWD) / PWF \times 100\%$ ,式中 *PWF* 为植株地上部鲜质量总和,*PWD* 为植株地上部干质量总和。取 3 株平均值代表该小区油菜植株含水率。

### 1.3 数据处理与分析

1.3.1 监测模型变量的筛选 本研究主要选用特征波段反射率和光谱指数为模型变量,运用 Matlab 软件计算,用 Origin 作图。

1.3.1.1 SR 分析 SR 分析以线性回归方程为基础,采用双检验原则,在建立“最优”回归方程的过程中按照贡献度由大到小的顺序,逐步引入对油菜 *PWC* 影响显著的波段,剔除影响不显著的波段,即每一步都进行 *F* 检验,保留  $F \geq 4$  时引入的波段,剔除  $F \leq 3.8$  时引入的波段,以筛选出最优波段组合。

1.3.1.2 SPA 法 首先设定波长数(*N*),然后根据不同波长数下的均方根误差(RMSE 值)确定最佳的建模波长数量。每次向前循环选择时,都计算新引入的波长在剩余波长变量上的投影,保证它在选择的前一个波长变量正交子空间上具有最大投影值,从而最大限度地消除共线性信息<sup>[18]</sup>。

1.3.1.3 CARS 法 首先利用蒙特卡罗采样法(Monte Carlo sampling, MSC)采样 50 次,运用指数衰减函数(Exponentially decreasing function, EDF)去掉权重较低的变量,运用自适应重加权采样技术(Adaptive reweighted sampling, ARS)筛选出偏最小二乘回归(Partial least regression, PLSR)模型回归系数绝对值大的变量,每次产生的新变量组合会再构建 PLS 回归模型并计算交叉验证均方差(RMSECV),RMSECV 最小的变量子集即为最优子集<sup>[19]</sup>。

1.3.1.4 RPSM 法 采取 RPSM 法<sup>[20]</sup>,在 680 ~

1 075 nm 波段范围内,首先每 10 nm 取光谱反射率均值,构建任意两波段组合的归一化光谱指数(*ND-SI*)和比值光谱指数(*RSI*),结合油菜植株含水率建立线性模型,计算决定系数,绘制等高线图以确定核心波段。对核心波段精细采样,以 1 nm 为间隔构建光谱指数。重复上述操作,选择决定系数最高的光谱指数作为模型变量。

$$NDSI = (R_{\lambda 1} - R_{\lambda 2}) / (R_{\lambda 1} + R_{\lambda 2}) \quad (1)$$

$$RSI = R_{\lambda 1} / R_{\lambda 2} \quad (2)$$

其中, $R_{\lambda 1}$ 、 $R_{\lambda 2}$  为 680~1 075 nm 范围内任意波段的光谱反射率。

### 1.3.2 监测模型的构建

1.3.2.1 线性回归 以筛选的光谱指数或敏感波段反射率作为自变量,油菜 *PWC* 作为因变量,运用最小二乘逼近拟合它们的定量关系,构建一元或多元线性回归方程。

1.3.2.2 BP 神经网络 本研究定义隐藏层神经元数量为 10,使用 Levenberg-Marquardt 反向传播算法进行训练。它是使用最广泛的非线性最小二乘算法,运用模型函数对估测数据在其领域内作线性近似分析,利用泰勒公式,忽略二阶以上导数项,能够使 BP 神经网络具有良好的非线性映射能力、容错性和较强的泛化能力。

1.3.2.3 支持向量机回归(SVR) 本研究使用支持向量机算法库 LIBSVM<sup>[21]</sup>,SVM 模型类型为 epsilon-SVR,选用 RBF 核函数,损失函数 *P* 值为 0.01,惩罚参数(*c*)和核函数参数(*g*)由穷举法确定。

1.3.3 监测模型的检验 采用常用的决定系数( $R^2$ )、均方根误差(RMSE)、平均绝对误差(MAE)和平均绝对误差占观测值的比值(*Dap*)对模型的精度进行评估, $R^2$ 为监测值与实际值的决定系数,RMSE 为监测的误差变幅, $R^2$ 越高,RMSE、MAE 和 *Dap* 越低,表明模型监测效果越理想。计算公式如下:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (P_i - Q_i)^2} \quad (3)$$

$$MAE = \frac{\sum_{i=1}^n |P_i - Q_i|}{n} \quad (4)$$

$$Dap = \frac{MAE}{\sum_{i=1}^n Q_i} \times 100 \quad (5)$$

其中,*n* 为模型测试的样本数, $P_i$  为模型的估测值, $Q_i$  为试验的实测值。



## 2 结果与分析

### 2.1 油菜植株含水率与冠层光谱反射率的相关性

2 个生长季油菜 *PWC* 与冠层光谱反射率的相关系数随波长的变化如图 2 所示。油菜 *PWC* 与冠层光谱反射率在 325~1 075 nm 波段大部分呈现极显著负相关性,即随着 *PWC* 增大冠层反射率降低,符合水分吸收太阳辐射使得光谱反射率降低的遥感原理,可见油菜冠层光谱反射率对 *PWC* 具有良好的指示作用。由于红边区域的 680~780 nm 反射率对水分有间接响应,近红外波段反射率对水分有直接响应,故选取红边波段和近红外波段进行后续分析。

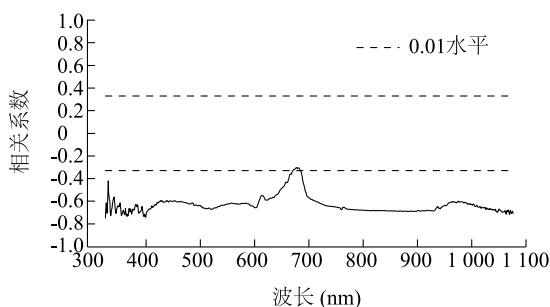


图 2 油菜植株含水率与冠层光谱反射率的相关系数随波长的变化

Fig.2 Variation of the correlation coefficient between rapeseed plant water content and canopy spectral reflectance as the wavelength changes

### 2.2 油菜植株含水率监测模型变量的选择

2.2.1 基于 SR 分析的特征波段反射率 针对油菜 *PWC* 利用全波段数据进行 SR 分析,结果如表 2 所示。共建立了 3 个回归模型,模型  $R^2$  随着波长数的增多逐渐增大,标准误差逐渐减小,所有模型均达到 0.01 显著性水平。第 3 个模型引入 1 071 nm、986 nm 和 730 nm 波长的光谱反射率,构建的回归模型  $R^2$  为 0.824,拟合度最佳,因此,将模型 3 作为最优波段组合用于构建油菜 *PWC* 监测模型。

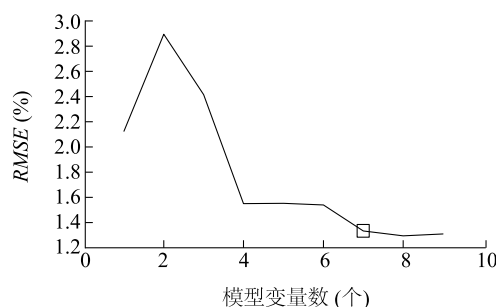
表 2 基于逐步回归分析筛选的油菜植株含水率 (*PWC*) 敏感波段

Table 2 Sensitive wavelengths screened by stepwise regression method for plant water content (*PWC*) of rapeseed

模型	波长 (nm)	$R^2$	标准误差	显著性
1	1 071	0.545	0.018 2	<0.001
2	1 071、986	0.664	0.015 9	<0.001
3	1 071、986、730	0.824	0.011 6	<0.001

$R^2$ : 决定系数。

2.2.2 基于 SPA 的特征波段反射率 对油菜 *PWC* 运行 SPA 算法,其 *RMSE* 值随模型变量数(波长数)的变化趋势如图 3 所示,先上升后下降,波长数量大于 7 个时,*RMSE* 值的变化不再显著,此时 *RMSE* 为 1.33%。因此选取 7 个波长作为特征波长组合,分别为 686 nm、695 nm、707 nm、746 nm、964 nm、1 065 nm 和 1 069 nm,其中 4 个位于可见光波段,3 个位于近红外波段。



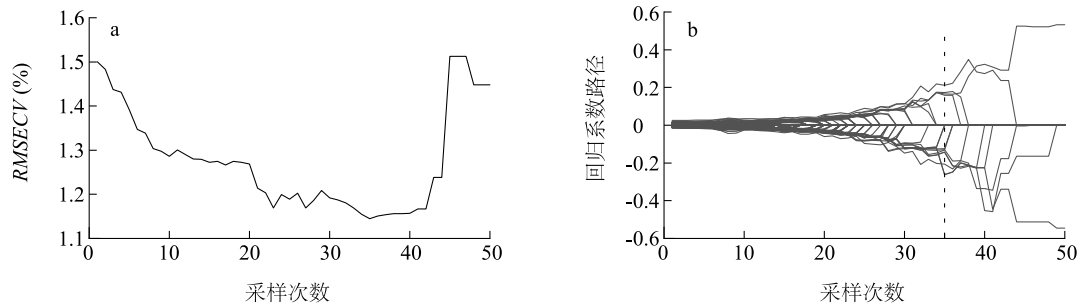
*RMSE*: 均方根误差。

图 3 基于 SPA 算法的油菜植株含水率 (*PWC*) 的均方根误差 (*RMSE*) 随模型变量数变化趋势

Fig.3 Variation trend of root mean square error (*RMSE*) for rapeseed plant water content (*PWC*) with number of model variables based on SPA algorithm

2.2.3 基于 CARS 的特征波段反射率 对油菜 *PWC* 运行 CARS 算法的筛选过程如图 4 所示。随着采样次数增加,在粗选过程中 *RMSE* 值迅速下降,然后下降速度逐渐变缓,转换到精选过程。剔除了与油菜 *PWC* 无关的波长,在 35 次采样后 *RMSECV* 逐渐增大,说明在筛选过程中丢失了有关信息。因此选择第 35 次采样筛选的 10 个波长为最优波长组合,分别为 694 nm、695 nm、696 nm、863 nm、864 nm、893 nm、973 nm、986 nm、1 050 nm 和 1 071 nm,其中,3 个波长位于可见光波段且十分集中,7 个波长位于近红外波段。

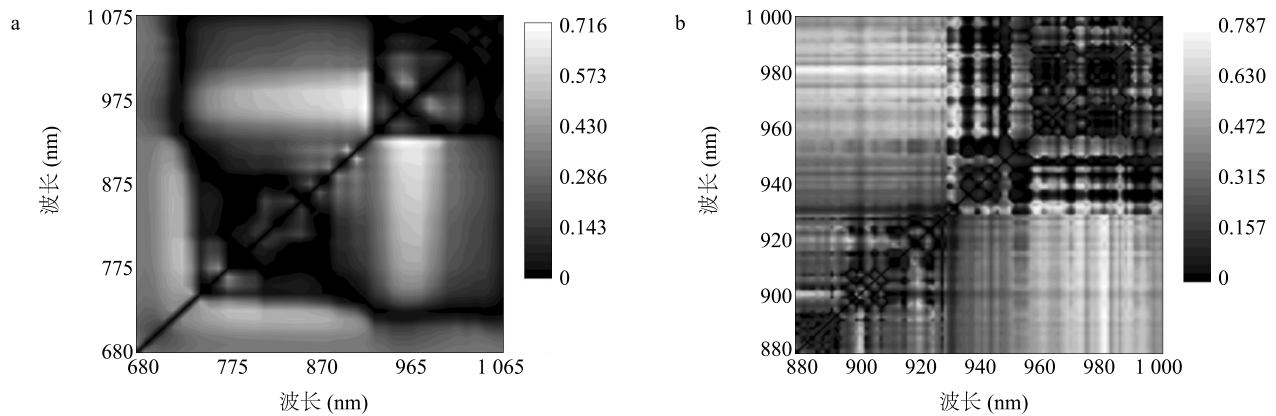
2.2.4 基于 RPSM 的光谱指数比较 基于油菜 *PWC* 与任意两波段组合的 *NDSI* 决定系数如图 5 所示。初步采样全波段循环的决定系数如图 5a 所示,*NDSI* 决定系数最大值均为 0.716,近红外波段 880~1 000 nm 的部分决定系数超过 0.500,因此选取 880~1 000 nm 的光谱反射率进行精细采样,结果如图 5b 所示,*NDSI*( $R_{981}, R_{894}$ ) 为监测油菜 *PWC* 所得最优 *NDSI*,决定系数为 0.787。*RSI* 决定系数情况与 *NDSI* 相似,*RSI*( $R_{981}, R_{894}$ ) 为最优 *RSI*,决定系数为 0.786,略低于 *NDSI*( $R_{981}, R_{894}$ ),选其作为变量构建油菜 *PWC* 监测模型。



a:模型交叉验证均方根误差变化趋势;b:各波长回归系数随采样次数变化趋势。 $RMSECV$ :交叉验证均方差。

图4 基于CARS筛选植株含水率(PWC)监测模型变量的过程

Fig.4 Process of selecting variables of plant water content (PWC) monitoring model based on CARS



a:在680~1 075 nm 范围内每隔 10 nm 采样;b:在880~1 000 nm 范围内每隔 1 nm 采样。

图5 油菜植株含水率(PWC)与任意两波段组合的归一化差值光谱指数(NDSI)的线性模型决定系数等高线

Fig.5 Contour map of linear model determination coefficient ( $R^2$ ) of rapeseed plant water content (PWC) and normalized difference spectral index (NDSI) by any two band combinations

## 2.3 油菜植株含水率监测模型的构建及检验

**2.3.1 线性回归模型** 对于采集的 60 个油菜样本, 42 个用于线性回归模型构建, 18 个用于检验, 模型效果如表 3 所示, 建模  $R^2$  由高到低依次为 SPA、CARS、SR、RPSM-NDSI、RPSM-RSI。基于 NDSI、RSI 构建的

模型建模和检验效果均相似, 有较好的监测能力, 但相比于 SR、SPA、CARS 筛选的敏感波段反射率略有不足, 因为参与的波段数较少, 丢失了一些有效信息。SPA-LR 的建模  $R^2$  和检验  $R^2$  最高, 分别为 0.878 和 0.693, 检验  $RMSE$  和  $Dap$  分别为 1.623% 和 7.530%。

表3 油菜植株含水率(PWC)的线性回归模型及检验

Table 3 Linear regression models for plant water content (PWC) in rapeseed and their performances

筛选方法	估测模型	建模 $R^2$	评价指标				
			$r$	$R^2$	$RMSE$ (%)	$MAE$ (%)	$Dap$ (%)
SR	$Y = 0.899 - 0.367 \times R_{1071} + 0.765 \times R_{986} - 0.471 \times R_{730}$	0.824	0.783 **	0.614	1.803	1.386	8.767
SPA	$Y = 0.885 + 3.763 \times R_{686} - 5.33 \times R_{695} + 1.704 \times R_{707} - 0.238 \times R_{746} + 0.162 \times R_{964} - 0.339 \times R_{1065} + 0.293 \times R_{1069}$	0.878	0.833 **	0.693	1.623	1.191	7.530
CARS	$Y = 0.902 + 1.606 \times R_{694} - 1.914 \times R_{696} - 0.403 \times R_{864} + 0.253 \times R_{973} + 0.292 \times R_{986} + 0.223 \times R_{1050} - 0.256 \times R_{1071}$	0.873	0.760 **	0.577	1.935	1.241	7.850
RPSM-NDSI	$Y = 0.895 + 0.667 \times NDSI(R_{981}, R_{894})$	0.787	0.687 **	0.472	2.112	1.42	8.978
RPSM-RSI	$Y = 0.533 + 0.362 \times RSI(R_{981}, R_{894})$	0.786	0.690 **	0.476	2.108	1.419	8.976

建模  $R^2$  为模型拟合度的决定系数,  $r$  为相关系数,  $R^2$  为决定系数,  $RMSE$  为均方根误差,  $MAE$  为平均绝对误差,  $Dap$  为平均绝对误差占观测值的比值。\*\* 表示在 0.01 水平影响显著,  $r(0.01, 16) = 0.59$ 。SR: 逐步回归分析; SPA: 连续投影算法; CARS: 竞争自适应加权算法; RPSM-NDSI: 减量精细采样法-归一化差值光谱指数; RPSM-RSI: 减量精细采样法-比值光谱指数。

2.3.2 BP 神经网络模型 对于采集的 60 个油菜样本,42 个用于训练 BP 神经网络模型,9 个用于验证,9 个用于测试,模型效果如表 4 所示,训练  $R^2$  由高到低依次为 SPA、RPSM-NDSI、RPSM-RSI、SR 及

CARS。以 SPA-BP 模型的训练  $R^2$  最高,为 0.964,SR-BP 模型的验证  $R^2$  最高,为 0.977,SPA-BP 模型的测试  $R^2$  最高,为 0.940, $RMSE$  和  $Dap$  分别为 1.836%和 17.566%。

表 4 油菜植株含水率(PWC)的 BP 神经网络模型及检验

Table 4 BP neural network models for plant water content (PWC) in rapeseed and their performances

筛选方法	训练 $R^2$	验证 $R^2$	评价指标				
			$r$	$R^2$	$RMSE(\%)$	$MAE(\%)$	$Dap(\%)$
SR	0.871	0.977	0.903 **	0.815	1.332	1.126	14.332
SPA	0.964	0.940	0.970 **	0.940	1.836	1.391	17.566
CARS	0.858	0.858	0.909 **	0.827	1.259	0.983	12.438
RPSM-NDSI	0.908	0.884	0.894 **	0.800	1.499	1.302	16.186
RPSM-RSI	0.873	0.905	0.936 **	0.877	1.288	0.901	11.369

\*\* 表示 0.01 水平显著性, $r(0.01,7)=0.798$ 。SR:逐步回归分析;SPA:连续投影算法;CARS:竞争自适应加权算法;RPSM-NDSI:减量精细采样法-归一化差值光谱指数;RPSM-RSI:减量精细采样法-比值光谱指数。

2.3.3 SVR 模型 对于采集的 60 个油菜样本,42 个用于 SVR 模型训练,18 个用于检验,模型效果如表 5 所示,训练  $R^2$  由高到低依次为 SPA、CARS、SR、

RPSM-NDSI 和 RPSM-RSI。SPA-SVR 模型的训练  $R^2$  最高,为 0.910,检验  $R^2$  亦最高,为 0.841, $RMSE$  和  $Dap$  分别为 1.227%和 5.368%。

表 5 油菜植株含水率(PWC)的 SVR 模型及检验

Table 5 SVM models for plant water content (PWC) in rapeseed and their performances

筛选方法	训练 $R^2$	评价指标				
		$r$	$R^2$	$RMSE(\%)$	$MAE(\%)$	$Dap(\%)$
SR	0.857	0.784 **	0.614	1.809	1.361	8.606
SPA	0.910	0.917 **	0.841	1.227	0.849	5.368
CARS	0.894	0.845 **	0.713	1.569	1.150	7.275
RPSM-NDSI	0.788	0.688 **	0.473	2.109	1.407	8.900
RPSM-RSI	0.788	0.690 **	0.476	2.101	1.379	8.723

\*\* 表示 0.01 水平显著性, $r(0.01,16)=0.59$ 。SR:逐步回归分析;SPA:连续投影算法;CARS:竞争自适应加权算法;RPSM-NDSI:减量精细采样法-归一化差值光谱指数;RPSM-RSI:减量精细采样法-比值光谱指数。

### 3 讨论

作物水分状况的光谱监测研究始于 20 世纪 70 年代,Thomas 等<sup>[22]</sup>的研究结果表明,植物叶片的光谱反射率随叶片含水量下降而增加,1 450 nm 和 1 930 nm 波段的光谱反射率和叶片含水量呈显著相关( $P=0.01$ )。后续 Holben 等<sup>[23]</sup>、Jackson 等<sup>[24]</sup>以及很多学者在不同植物上的研究结果都表明,主要水分吸收波段为 0.97  $\mu\text{m}$ 、1.19  $\mu\text{m}$ 、1.45  $\mu\text{m}$ 、1.97  $\mu\text{m}$  和 2.70  $\mu\text{m}$ <sup>[25]</sup>。高光谱能够更加精细地体现光谱的变化,因此除了主要水分吸收波段,亦存在其他能够反映作物水分差异的敏感波段,例如 680~780 nm 区域的红边光谱,PWC 越低近红外波

段反射率越高,红边区域的光谱反射曲线会更陡峭<sup>[26]</sup>,将这些波段加以利用,可以实现作物水分状况的精确监测,故本研究选取具备理论支撑的680~1 075 nm 波段进行分析。然而高光谱数据量大,需要采取合适的手段提取关键波段,随着光谱分析技术的发展,全波段光谱快速分析成为可能。SR 分析通过逐一引入光谱变量,剔除对监测目标影响不显著的变量,保留影响显著的变量,来筛选最优光谱变量集。SPA 法是一种前向变量选择算法,能够使矢量空间共线性最小化,消除原始光谱数据中的冗余信息,该方法在叶绿素、生物量的高光谱定量监测中广泛应用<sup>[27-30]</sup>。CARS 法基于“适者生存”的原则,以迭代和竞争的方式筛选最优波段组合。这些方法

均能够在降低数据维度、筛选最优波段组合的同时充分挖掘数据、提取有效信息,从而监测并诊断作物水分状况<sup>[31-33]</sup>。光谱指数简单且具有较好的推广性,其中 *NDSI* 和 *RSI* 广泛用于获取植被生物物理特征属性<sup>[34]</sup>,减量精细采样法能够有效融合降采样法和精细采样法筛选最优光谱指数<sup>[20]</sup>。已有研究大多是运用一个方法筛选敏感波段,这些方法遵循的筛选原则不同,适配的建模方法不同,有必要和不同的建模方法组合,寻找筛选变量-建模效果最好的组合。

前述 3 个方法筛选的波段反射率变量更多、更分散,因此蕴含的信息更全面,在建模效果中即有体现,光谱指数构建模型的建模效果、检验效果低于全波段光谱分析。本研究旨在提高油菜水分状况的监测精度以及普适性,所用 SR、SPA 和 CARS 3 种全波段分析方法筛选的变量数量有所不同,但拟合精度最高,研究更侧重于比较不同变量筛选方法搭配不同建模方法构建的油菜 *PWC* 定量监测模型,今后的研究中可以采用更多的光谱分析方法,采集全波段有效信息,实现光谱数据的充分利用。

已有油菜水分状况的监测研究大多是运用线性或者非线性普通回归模型,在监测单一生育期作物水分状况时效果较好,而对全生育期的普适性有待提高。本研究引入 BP 神经网络、SVR 进行非线性拟合,与线性回归模型相比较,模型拟合优度明显提高,能够为油菜水分诊断提供依据。本研究设置的水分梯度处理为常规油菜生存环境所面临的干旱、正常和渍水情况,对于特殊气候地区,有必要进一步扩大油菜 *PWC* 范围,以扩大模型的监测适应性。

## 4 结 论

本研究基于不同品种、氮素与水分处理的油菜池栽田间试验,运用 RPSM 法、SR 分析法、SPA 法和 CARS 算法筛选监测模型变量,充分挖掘高光谱数据,构建线性回归模型、BP 神经网络模型、SVR 模型监测油菜 *PWC*,模型的估测值和实测值均呈现极显著相关关系。其中 SPA-LR 模型估测值与实测值的检验  $R^2$  为 0.693,模型的  $RMSE$  为 1.623%;SPA-BP 模型估测值与实测值的检验  $R^2$  为 0.940,模型的  $RMSE$  为 1.836%;SPA-SVR 模型估测值与实测值的检验  $R^2$  为 0.841,模型的  $RMSE$  为 1.227%。上述模型均能够实现油菜 *PWC* 的精确监测,以 SPA-SVR

模型的监测效果为最好,这可为油菜水分诊断提供依据。

致谢: 本研究所用试验材料宁油 22 和宁杂 1818 由江苏省农业科学院经济作物研究所张洁夫老师和浦惠明老师提供,浙杂 903 由江苏省兴化市农业技术推广中心王洁站长提供,特致谢忱!

## 参考文献:

- [1] 沈金雄,傅廷栋. 我国油菜生产、改良与食用油供给安全[J]. 中国农业科技导报, 2011, 13(1): 1-8.
- [2] 瞿益民,唐合年,葛妹兰,等. 油菜需水量试验分析[J]. 江苏水利, 2005(10): 20-21,23.
- [3] 张永忠. 平凉市冬油菜需水量试验结果分析[J]. 甘肃水利水电技术, 2003(4): 331-333.
- [4] 谢素华,杨明高. 人民渠平原灌区油菜需水量及需水规律研究[J]. 四川水利, 2001(1): 33-35.
- [5] 黄纯倩,朱晓义,张亮,等. 干旱和高温对油菜叶片光合作用和叶绿素荧光特性的影响[J]. 中国油料作物学报, 2017, 39(3): 342-350.
- [6] 邹小云,刘宝林,宋来强,等. 施氮量与花期水分胁迫对不同氮效率油菜产量性能及氮肥利用效率的影响[J]. 华北农学报, 2015, 30(2): 220-226.
- [7] 宋丰萍. 渍水时间对油菜生长及产量的影响[J]. 作物学报, 2010, 36(1): 170-176.
- [8] MAMNABI S, NASROLLAHZADEH S, GHASSEMI-GOLEZANI K, et al. Improving yield-related physiological characteristics of spring rapeseed by integrated fertilizer management under water deficit conditions[J]. Saudi Journal of Biological Sciences, 2020, 27(3): 797-804.
- [9] 赵丽英,王伟,宋玉伟. 土壤水分胁迫下油菜光合特性变化和膜伤害研究[J]. 河南农业科学, 2010(8): 33-35.
- [10] QUEMADA C, PÉREZ-ESCUADERO J M, GONZALO R, et al. Remote sensing for plant water content monitoring: a review[J]. Remote Sensing, 2021, 13(11): 2088.
- [11] 杨玉清,张甜甜,李军会,等. 近红外高光谱的活体玉米叶片水分成像研究[J]. 光谱学与光谱分析, 2018, 38(12): 3743-3747.
- [12] KRISHNA G, SAHOO R N, SINGH P, et al. Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing[J]. Agricultural Water Management, 2019, 213: 231-244.
- [13] 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[J]. 麦类作物学报, 2018, 38(7): 854-862.
- [14] 张晓东,毛罕平,左志宇,等. 干旱胁迫下油菜含水率的高光谱遥感估算研究[J]. 安徽农业科学, 2011, 39(30): 18451-18452,18487.
- [15] 张晓东,李立,毛罕平,等. 基于 PCA-BP 多特征融合的油菜



- 水分胁迫无损检测[J]. 江苏大学学报(自然科学版), 2016, 37(2): 174-182.
- [16] 全春艳, 马 驿, 杨振忠, 等. 基于角度指数的油菜叶片等效水厚度估算研究[J]. 核农学报, 2019, 33(1): 187-198.
- [17] 张 君, 蔡振江, 张东方, 等. 基于机器学习与光谱技术的油菜叶片含水率估测研究[J]. 河北农业大学学报, 2021, 44(6): 122-127.
- [18] YE S F, WANG D, MIN S G. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection [J]. *Chemometrics and Intelligent Laboratory Systems*, 2007, 91(2): 194-199.
- [19] LI H D, LIANG Y Z, Xu Q S, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. *Analytica Chimica Acta*, 2009, 648(1): 77-84.
- [20] 姚 霞, 刘小军, 王 薇, 等. 基于减量精细采样法估算小麦叶片氮积累量的最佳归一化光谱指数[J]. 应用生态学报, 2010, 21(12): 3175-3182.
- [21] CHANG C C, LIN C J. LIBSVM: a library for support vector machine[J]. *ACM Transactions on Internet Systems and Technology*, 2011, 2: 1-27.
- [22] THOMAS J R, NAMKEN L N, OERTHER G F, et al. Estimating leaf water content by reflectance measurements 1[J]. *Agronomy Journal*, 1971, 63(6): 845-847.
- [23] HOLBEN B N, SCHUTT J B, MCMURTREY J. Leaf water stress detection utilizing thematic mapper bands 3, 4 and 5 in soybean plants[J]. *International Journal of Remote Sensing*, 1983, 4(2): 289-297.
- [24] JACKSON R D, EZRA C E. Spectral response of cotton to suddenly induced water stress[J]. *International Journal of Remote Sensing*, 1985, 6(1): 177-185.
- [25] JENSEN J R. *Remote sensing of the environment*[M]. New York: Pearson Education Limited, 2007.
- [26] LIU L Y, WANG J H, HUANG W J, et al. Estimating winter wheat plant water content using red edge parameters[J]. *International Journal of Remote Sensing*, 2003, 25(17): 1688-1691.
- [27] 郭 松, 常庆瑞, 崔小涛, 等. 基于光谱变换与 SPA-SVR 的玉米 SPAD 值高光谱估测[J]. 东北农业大学学报, 2021, 52(8): 79-88.
- [28] 王玉娜, 李粉玲, 王伟东, 等. 基于连续投影算法和光谱变换的冬小麦生物量高光谱遥感估算[J]. 麦类作物学报, 2020, 40(11): 1389-1398.
- [29] 董 哲, 杨武德, 朱洪芬, 等. 基于连续投影算法与 BP 神经网络的玉米叶片 SPAD 值高光谱估算[J]. 山西农业科学, 2019, 47(5): 751-755.
- [30] 易 翔, 张立福, 吕 新, 等. 基于无人机高光谱融合连续投影算法估算棉花地上部生物量[J]. 棉花学报, 2021, 33(3): 224-234.
- [31] 孙 俊, 丛孙丽, 毛罕平, 等. 基于高光谱的油麦菜叶片水分 CARS-ABC-SVR 预测模型[J]. 农业工程学报, 2017, 33(5): 178-184.
- [32] 王松磊, 吴龙国, 王彩霞, 等. 可见近红外高光谱快速诊断番茄叶片含水量及其分布[J]. 光子学报, 2019, 30(9): 941-950.
- [33] 陈秀青, 杨 琦, 韩景晔, 等. 基于叶冠尺度高光谱的冬小麦叶片含水量估算[J]. 光谱学与光谱分析, 2020, 40(3): 891-897.
- [34] 张 峰, 周广胜. 植被含水量高光谱遥感监测研究进展[J]. 植物生态学报, 2018, 42(5): 517-525.

(责任编辑: 张震林)