

基于BSO-SVR的香蕉遥感时序估产模型研究

张海洋¹ 张瑶¹ 李民赞¹ 李修华² 王俊³ 田泽众¹

(1. 中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083;

2. 广西大学电气工程学院, 南宁 530004; 3. 河南科技大学电气工程学院, 洛阳 471003)

摘要: 为了提高有限样本下遥感时序估产效果, 本文提出一种基于BSO-SVR的香蕉遥感时序估产模型。该模型以广西壮族自治区扶绥县的71块香蕉田块为研究区, 利用时间序列Sentinel-2遥感影像数据, 结合实测产量数据, 对2019—2020年香蕉产量进行预测与分析。融合阈值分割和形态学开操作方法, 滤除香蕉关键生育期内遥感影像的厚云和云阴影区域; 引入头脑风暴优化算法(Brain storming optimization algorithm, BSO)自动搜寻支持向量回归算法(Support vector regression, SVR)的最优惩罚因子和核函数参数, 解决SVR模型的参数优化不足导致模型预测精度低的问题; 搭建基于BSO-SVR的时间序列遥感估产模型, 深入挖掘多时相遥感信息, 以提升香蕉估产准确度。结果表明, 相较于网格搜索算法(Grid search, GS)和灰狼优化算法(Grey wolf optimizer, GWO)搜寻SVR模型的最优参数, 本文提出的头脑风暴优化算法具有更高的预测精度和更快的预测速度, 在2019年和2020年BSO-SVR模型测试集的决定系数(Coefficient of determination, R^2)分别为0.777和0.793, 验证集 R^2 分别为0.765和0.636, 运行时间分别为0.320、0.331 s; 与传统的岭回归模型(Ridge regression, RR)和偏最小二乘回归模型(Partial least squares regression, PLSR)相比, BSO-SVR模型的预测性能最佳, 其次是RR模型, PLSR模型表现最差。本文提出的时序估产模型实现了香蕉田块产量的精准预估。

关键词: 香蕉; 遥感; 估产; 时间序列; 头脑风暴优化算法; 支持向量回归算法

中图分类号: TP79; S127 **文献标识码:** A **文章编号:** 1000-1298(2021)S0-0098-10

BSO-SVR-based Remote Sensing Time-series Yield Estimation Model for Banana

ZHANG Haiyang¹ ZHANG Yao¹ LI Minzan¹ LI Xiuhua² WANG Jun³ TIAN Zezhong¹

(1. Key Laboratory for Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China

2. College of Electrical Engineering, Guangxi University, Nanning 530004, China

3. College of Electrical Engineering, Henan University of Science and Technology, Luoyang 471003, China)

Abstract: Timely, comprehensive and accurate estimation of banana yield can provide growers with decisions on variable fertilization, irrigation, harvest planning, marketing and forward sales. To improve the accuracy of banana remote sensing yield estimation, totally 71 banana fields in Fusui County, Guangxi were used as the study area, and a remote sensing prediction model for banana yield in 2019—2020 was conducted by using time-series Sentinel-2 remote sensing image data, combined with field measured yield data. The method firstly obtained Sentinel-2 images during the key banana phenological period of 2019—2020, then the threshold segmentation and morphological open operation methods were used to remove cloud and cloud shadow coverage areas, the average normalized difference vegetation index (NDVI) values of each plot were extracted, and finally the BSO-SVR model was used to predict and evaluate the banana yield in combination with the actual measured data of banana yield. The results showed that compared with the grid search (GS) and grey wolf optimizer (GWO) algorithms to optimize

收稿日期: 2021-07-01 修回日期: 2021-08-15

基金项目: 广西创新驱动发展专项资金项目(桂科 AA18118037-3)、国家自然科学基金项目(41801245)和中央高校基本科研业务费专项资金项目(2021AC026)

作者简介: 张海洋(1993—),男,博士生,主要从事农作物遥感估产研究,E-mail: zhy_cau@163.com

通信作者: 张瑶(1988—),女,副教授,博士生导师,主要从事农业定量遥感应用研究,E-mail: zhangyao@cau.edu.cn

the penalty factor and kernel function parameters of the SVR model, the brain storming optimization algorithm proposed had higher prediction accuracy and faster prediction speed. The running times of the BSO-SVR model in 2019 and 2020 were 0.320 s and 0.331 s, respectively, and for the validation set, the R^2 of the BSO-SVR model was 0.777 and 0.793 in 2019 and 2020, respectively; for the test set, the R^2 of the BSO-SVR model was 0.765 and 0.636 in 2019 and 2020, respectively, except that the R^2 of the BSO-SVR model in 2019 is slightly lower than that of the GS-SVR model ($R^2 = 0.797$) in 2019, except that the R^2 of the BSO-SVR model was higher than that of the GWO-SVR model and the GS-SVR model, and in addition, the overall performance of the RMSE and MAE of the BSO-SVR model was optimal in 2019—2020 compared with that of the GWO-SVR model and the GS-SVR model, indicating that the prediction results of the BSO-SVR model were closer to the actual values and with higher forecasting accuracy. Compared with the traditional ridge regression (RR) and partial least squares regression (PLSR) models, in 2019, the BSO-SVR model had the highest R^2 , followed by the RR model, and the PLSR model was the worst, where the BSO-SVR model had R^2 above 0.75 for both the validation and test sets, which was 0.113 and 0.174 higher than that of the RR model, and 0.192 and 0.184 higher than that of the PLSR model, respectively. Meanwhile, the BSO-SVR model had the lowest RMSE and MAE compared with the RR model and PLSR model, indicating that the BSO-SVR model had good results in forecasting banana yield in 2019. In 2020, the BSO-SVR model had the best overall performance, with the average R^2 of 0.715 for the validation and test sets, and the R^2 of the validation and test sets were higher than that of the RR model by 0.035 and 0.014, respectively, and better than that of the PLSR model by 0.040 and 0.035, while the RMSE and MAE of the BSO-SVR model also had the best overall performance. The banana time-series yield estimation model proposed achieved accurate yield prediction of banana field plots, which can provide an effective way for field-scale crop yield estimation.

Key words: banana; remote sensing; yield estimation; time-series; brain storming optimization algorithm; support vector regression

0 引言

及时、准确、客观地估算香蕉产量,可为种植者提供变量施肥、灌溉、收获和销售等方面的决策支持^[1-3]。传统估产方法利用目测或抽样技术,结合研究区的天气状况、作物品种、土地类型以及病虫害发生状况等因素进行估算,该方法耗时费力且易受人为主观因素影响^[4-5]。遥感技术具有获取成本低、覆盖范围广和能连续动态监测等优势,在作物产量估算中得到广泛关注^[6-7]。基于遥感技术的产量估算方法主要有两种:①运用数据同化技术,将遥感反演参数信息集成到作物的生长模型中,实现作物生长状态监测和产量估算。但该方法需要大量的精确参数输入,计算成本高。此外,数据获取不完整也会造成预报结果偏差大、不稳定等问题^[8-10]。②利用经验统计模型方法预测作物产量,此方法无需事先对基础数据分布进行假设,而是利用特定生育期作物的特征植被指数等变量来预测产量,简单易行,在遥感估产中得到了广泛应用^[11-13]。然而,该方法仅在特定生产条件下,估产精度可以取得较好的结果,时空泛化能力弱。

特定生育期遥感数据信息量有限,无法从作物生长的角度描述产量形成的全过程,致使大多数情

况下的产量估算精度较低^[14-15]。作物生长时序遥感数据蕴含作物整个生长季的生长发育动态信息^[16]。以香蕉作物为例,香蕉的出苗、孕蕾、抽蕾、挂果、采收等过程均与产量形成密切相关^[17-19]。构建基于多时相遥感数据的产量估算模型相较单一特定时相具有明显的优势^[20]。

Sentinel-2 卫星遥感影像以其高时间频率和高空间分辨率的特点,为香蕉作物时序估产提供了有力的数据支撑^[21-22]。然而,与其他作物产量估算研究面临同样的问题,即以田块为单位的实测产量数据获取成本较高、难度大,极大地限制了产量估算模型构建过程中的样本容量^[23-25]。

支持向量回归算法 (Support vector regression, SVR) 因其在小样本情况下仍具有泛化能力强及预测准确率高等优点,被国内外学者广泛地应用到农作物产量预测中^[26-27]。支持向量回归算法性能主要由惩罚因子和核函数参数确定,在使用支持向量回归方法建模的过程中,这两种参数的选取依赖人工经验,受人为主观因素影响而出现欠学习或过拟合现象^[28-30]。为提高 SVR 模型的预测准确性,需要合理选取优化算法对其参数进行调整。近几年来,国内外学者在惩罚因子和核函数参数的自适应确定方法上开展了大量研究^[31-32]。此外,还有学者

运用其他智能优化算法(如粒子群优化算法^[33]、鲸鱼优化算法^[34-35]、人工蜂群算法^[36])对 SVR 模型参数进行优化。与凭借经验设定 SVR 模型中的参数相比,这些智能优化算法能根据输入特征自动搜寻 SVR 模型的最佳设定参数,自适应性较好。但这些优化算法存在计算效率低,算法复杂度高和易陷入局部最优等缺陷。头脑风暴优化算法(Brain storming optimization algorithm, BSO)是一种基于模拟人类创造性解决问题过程的群体智能算法,该方法采用聚类思想搜索局部最优,降低了算法复杂性;通过局部最优的比较获得全局最优,避免算法陷入局部最优,为优化 SVR 参数提供一种有益思路^[37-39]。

本文以香蕉作为研究对象,获取香蕉营养生长期、孕蕾抽蕾期和幼果期等关键生育期的 Sentinel-2 遥感卫星影像。融合阈值分割和形态学开操作去除云和云阴影覆盖区域。引入 BSO 算法对 SVR 模型进行参数寻优,提出一种基于 BSO-SVR 的香蕉遥感时序估产模型,分别对 2019 年和 2020 年的

香蕉产量进行高精度预报。

1 研究区概况与数据源

1.1 研究区概况

研究区位于广西壮族自治区崇左市扶绥县香蕉种植基地(图 1),覆盖面积约 140.11 hm²,覆盖范围为 22°29′34.58″~22°30′48.32″N,107°46′7.03″~107°47′19.70″E。采用千寻位置网络有限公司的千寻星矩 SR6 网络 RTK 接收机和千寻寸技术服务,该技术服务基于北斗卫星系统基础定位数据,利用全国超 2 000 多个基站,能够为用户提供厘米级定位精度。研究区地势平坦,属亚热带季风气候,夏季高温多雨,光热充足,春季温暖湿润、秋季干燥少雨,冬季短促气候较低,年平均气温 21.3~22.4℃之间,冬季和夏季平均气温分别约为 14.1、27.9℃。该区域年降雨量较大,每年稳定在 1 200 mm 以上,年降雨天数在 130~200 d 之间,年日照时数平均为 1 600 h 左右。

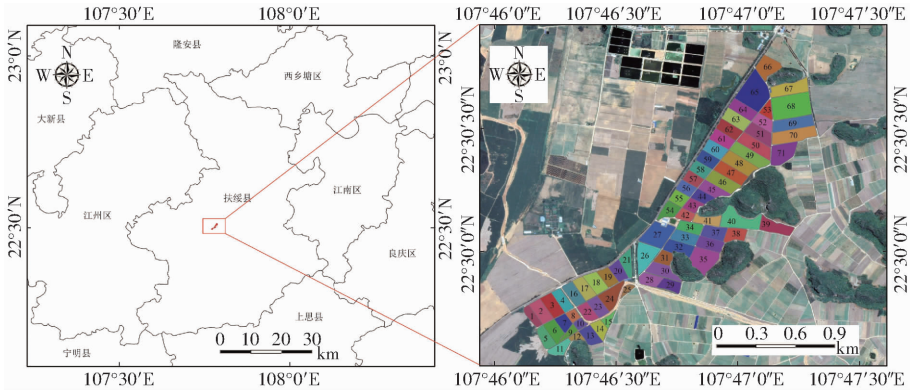


图 1 研究区位置及田块分布

Fig. 1 Location and field distribution of study area

研究区种植的香蕉品种为威廉斯 B6,栽植期为 9 月下旬或 10 月,其生长过程大致可分为定植期、营养生长期、孕蕾抽蕾期、幼果期和采收期等 5 个生育期。定植期是将香蕉苗移栽至大田或在老根上宿根的过程;一般每年 2—4 月对香蕉苗进行移裁定植;香蕉苗从定植期到 30 张叶片为香蕉营养生长期,时间为每年 5—6 月;孕蕾抽蕾期大致为每年的 7 月,该时期花芽开花及孕果穗;幼果期和采收期的香蕉果指肥度、饱满度逐渐增长,直到达到采收要求,时间为每年的 8—9 月。从营养生长期到幼果期是影响香蕉产量和质量形成的关键生育期,也是香蕉估产的最佳时期,因此本研究选取遥感影像成像时间为每年 5—9 月。

1.2 卫星遥感数据及预处理

选用的遥感数据为 Sentinel-2 影像数据, Sentinel-2 卫星搭载的多光谱传感器具有 13 个波

段,包括 10、20、60 m 3 种空间分辨率,其中分辨率 10 m 的波段为红波段(Red)、绿波段(Green)、蓝波段(Blue)、近红外波段(NIR);分辨率 20 m 的波段为 3 个红边波段(RE1、RE2、RE3)、近红外波段(Narrow NIR)、短波红外(SWIR1、SWIR2);分辨率 60 m 的波段为海岸波段(Coastal aerosol)、水汽波段(Water vapour)、卷云波段(SWIR cirrus)。

通过 Google Earth Engine (<https://earthengine.google.com/>) 平台来获取试验区 2019 年和 2020 年 5—9 月云量低于 35% 的 Sentinel-2 A/B L2A 级 MSI 影像, L2A 级影像是将大气顶层反射率数据(Top of atmosphere reflectance, TOA)经过正射校正、几何精校正以及大气校正转换后的地表反射率数据(Surface reflectance, SR)。因研究区处于中国南方地区,遥感探测过程中极易受到云、雨等气候条件影响,往往无法获取该地区全面、完整的 Sentinel-2

卫星影像。因此,每个月仅选取 1 幅最佳影像(感兴趣区云量最少,若一个月中有几幅云量较少的影像,则选取靠近月中旬的影像)作为该月的选定影像,选取的影像数据成像时间如表 1 所示。

表 1 选取的 Sentinel-2 影像成像时间

Tab.1 Imaging time of acquired Sentinel-2 images

年份	日期/月份
2019	11/05, 30/06, 20/07, 09/08, 23/09
2020	05/05, 04/06, 09/07, 23/08, 02/09

为了消除研究区的云和云阴影等因素对香蕉估产结果的影响,本文结合阈值分割和形态学操作方法进行去云处理。首先通过对比云覆盖区域与无云区域的 Green 和 NIR 两个波段的光谱反射率,设定恰当的分割阈值。其中,通过 Green 波段可滤除白色厚云区域,利用 NIR 波段可去除云阴影区域。然后利用形态学开运算操作对云掩膜影像进行处理,以滤除云掩膜影像的异常像素点和平滑云覆盖区域。云覆盖区域去除效果如图 2c 所示。

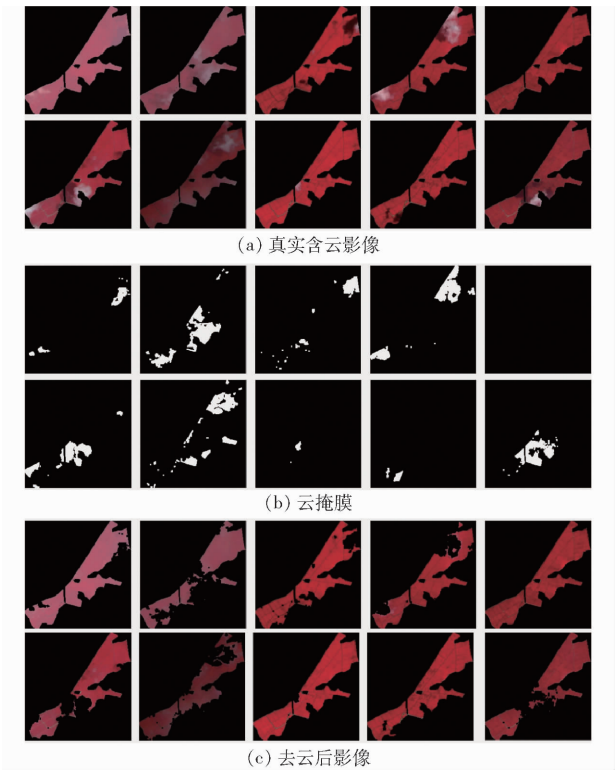


图 2 Sentinel-2 图像的云和云阴影去除结果

Fig.2 Clouds and cloud shadows removal result of Sentinel-2 images

1.3 香蕉产量

2019 年和 2020 年通过对种植户分批次采收香蕉时记录的售卖量进行统计得到香蕉各田块的实测产量,然后对每个田块的实际面积(hm²)进行测量,即可求取每个年份各田块的单产值(t/hm²),香蕉具体实测单产数据如表 2 所示。

表 2 香蕉实测单产

Tab.2 Measured banana yield in detail t/hm²

年份	最小值	最大值	均值	标准差	中位数
2019	36.04	55.20	46.975	5.007	48.905
2020	36.40	52.89	46.985	3.575	47.390

2 基于 BSO-SVR 的香蕉估产模型构建

2.1 SVR 理论基础

SVR 是一种基于结构风险最小化原理的用于解决小样本、非线性及高维问题的机器学习方法,其核心思想是利用核函数将非线性函数从低维空间映射到高维特征线性函数,然后进行线性拟合。

设样本集为 $(x_i, y_i), i = 1, 2, \dots, l$ 。其中, x_i 为第 i 个样本的输入变量, $x_i \in \mathbf{R}^w$; y_i 为第 i 个样本的输出变量, $y_i \in \mathbf{R}$; w 表示输入变量的维数。SVR 在高维特征空间的线性函数可表示为

$$f(x) = \omega \varphi(x) + b \tag{1}$$

式中 ω ——法向量 b ——偏置项

$\varphi(x)$ ——特征映射函数

针对回归问题,引入不敏感损失函数 ε ,在正、负松弛度分别为 ξ 和 ξ' 的水平下进行无误差拟合,以回归函数的复杂度和拟合误差之和最小值为目标函数,优化模型为

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi'_i) \\ \text{s. t.} \begin{cases} \omega \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i & (\xi_i \geq 0) \\ y_i - \omega \varphi(x_i) - b \leq \varepsilon + \xi'_i & (\xi'_i \geq 0) \end{cases} \end{cases} \tag{2}$$

式中 C ——惩罚因子

为求解约束最优化问题,引入 Lagrange 乘子,令 ω, b, ξ 和 ξ' 等变量的偏导为零,则可得到 SVR 的对偶形式

$$\begin{cases} \max \left(\sum_{i=1}^l y_i (\alpha_i - \alpha'_i) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha'_i) - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha'_i) (\alpha_j - \alpha'_j) \kappa(x_i, x_j) \right) \\ \text{s. t.} \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha'_i) = 0 \\ 0 \leq \alpha_i, \alpha'_i \leq C \end{cases} \end{cases} \tag{3}$$

其中 $\kappa(x_i, x_j) = \varphi(x_i) \varphi(x_j)$

式中 $\kappa(x_i, x_j)$ ——核函数,表示将 x_i, x_j 映射到高维特征空间后进行内积计算

α_i, α'_i ——Lagrange 乘子

此时,SVR 回归函数可表示为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha'_i) \kappa(x_i, x_j) + b \tag{4}$$

常用的核函数有线性核函数、多项式核函数、高斯核函数和 Sigmoid 核函数等。采用目前广泛选用的高斯核函数,其表达式为

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

式中 σ ——高斯核函数的宽度

2.2 头脑风暴优化算法参数寻优

采用 SVR 模型进行香蕉产量预测时,惩罚因子 C 和核函数参数 $g(g = 1/(2\sigma^2))$ 是 2 个需要确定的重要参数,直接影响最终预测结果的准确度。其中,惩罚因子 C 主要用来平衡支持向量的复杂度和误差率,若 C 取值过大,支持向量增加,导致模型变复杂;若 C 取值过小,支持向量减少,模型变简单。核函数参数 g 反映了单个样本对超平面的影响,若 g 较小,则单个样本对超平面的影响较小;若 g 较大,则单个样本对超平面的影响较大。因此,为了提高香蕉产量估测模型的精度,采用头脑风暴优化算法(BSO)对惩罚因子 C 和核函数参数 g 进行优化计算,建立基于头脑风暴优化算法优化支持向量回归(BSO-SVR)的香蕉产量估算模型。

BSO 算法需先产生一定数量的个体,然后利用 K-means 算法进行分类,找出每个类中最优的个体作为类中心,再按一定规则更新类中心和个体,达到停止条件停止迭代。结合 BSO 算法的计算流程,利用 BSO 算法优化 SVR 模型的参数 C 和 g 的具体步骤如下(图 3):

- (1)算法初始化,设置最大迭代次数 T ,初始随机产生的个体数量 n ,分类类别 m 和概率 p_a, p_b, p_c, p_d 。
- (2)计算这 n 个个体的适应度。其中,适应度为五折交叉验证下的均方误差。
- (3)利用 K-means 聚类算法将 n 个个体分成 m 类,将每类中最高适应度的个体作为该类的类中心。
- (4)以一定概率 p_a 进行类中心更新。随机选择一个类中心,产生随机数 $r_1 \in (0, 1)$,若 $r_1 < p_a$,则用随机生成的个体替换该类中心。
- (5)个体更新。产生随机数 $r_2 \in (0, 1)$ 。若 $r_2 < p_b$,则进行个体更新。产生随机数 $r_3 \in (0, 1)$,若 $r_3 < p_i$ 且 $r_3 < p_c$,则在该类中心加一个随机扰动产生新个体;若 $r_3 < p_i$ 且 $r_3 \geq p_c$,则在该类中随机选择一个个体,加一个随机扰动产生新个体。其中, p_i 表示第 i 类个体被选中的概率。若 $r_2 \geq p_b$,则随机选择两个类产生新个体。产生随机数 $r_4 \in (0, 1)$,若 $r_4 < p_d$,则融合两个类中心后加一个随机扰动产生新个体;若 $r_4 \geq p_d$,则先在这两类上分别随机选择一个个体进行融合,再加一个随机扰动产生新个体。

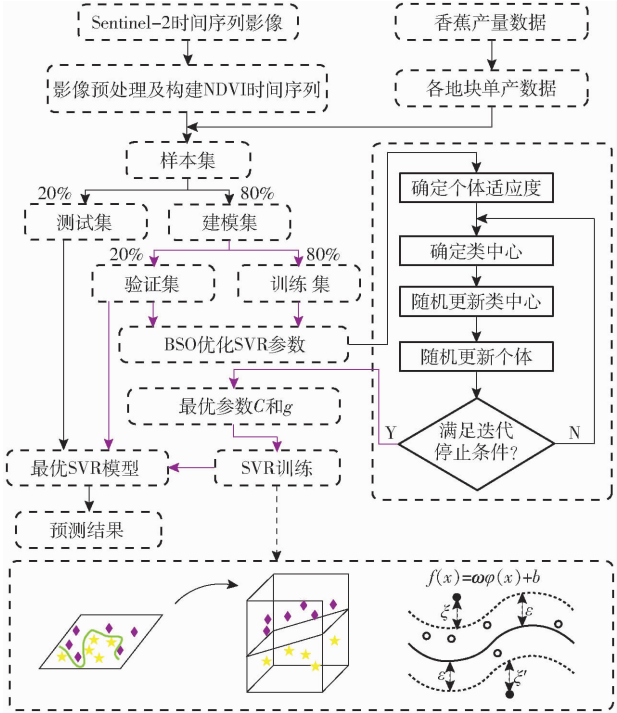


图 3 基于 BSO-SVR 的香蕉产量预测模型流程图

Fig.3 Flow chart of BSO-SVR based banana yield forecasting model

(6)比较新产生的个体和原个体,适应度小的作为下一次迭代的新个体。

(7) n 个个体逐一进行更新,若达到迭代停止条件,则停止迭代,否则返回步骤(3)。

2.3 基于 BSO-SVR 的香蕉产量预测模型

基于 BSO-SVR 模型的香蕉产量预测流程如图 3 所示,步骤如下:

- (1)获取 2019—2020 年研究区香蕉关键生育期内的多时相 Sentinel-2 遥感影像数据,结合阈值分割和形态学操作方法对获得的影像进行去云处理。随后,求取各地块区域内所有像素的平均 NDVI 值。若某一地块被云层全部覆盖,则使用该影像的 NDVI 均值进行缺失值填充。
- (2)获取 2019—2020 年研究区所有地块的产量数据,依据每个地块的面积,求取各地块的单产数据。
- (3)整合每年香蕉生育期内影像中各地块的时间序列 NDVI 值,生成一个二维矩阵 X ,横向代表时间维度,列向代表不同田块的 NDVI 均值。每个田块的时间序列信息可表示为
$$x_i^b = (x_i^1, x_i^2, \dots, x_i^B) \quad (i = 1, 2, \dots, t) \quad (6)$$
式中 i ——田块序号 b ——时序影像
 t ——田块总数量
 B ——每年各田块时序影像的总数量
每年各地块的香蕉单产为
$$y_i = (y_1, y_2, \dots, y_n) \quad (i = 1, 2, \dots, t) \quad (7)$$
- (4)借助 Python 中 sklearn 库的“training_test_

split”函数,将样本以 8:2 比例随机划分为建模集和测试集,再将建模集以 8:2 占比随机划分为训练集和验证集,则训练集有 45 个样本,验证集有 12 个样本,测试集有 14 个样本。然后,将训练集、验证集和测试集的输入变量进行标准化处理。

(5)以训练集和验证集作为模型控制量,设定 BSO 的初始参数,根据 2.2 节头脑风暴优化算法对 SVR 的惩罚因子 C 和核函数参数 g 进行优化,获得最优参数 C 和 g 。

(6)利用 BSO 算法优化后的 SVR 模型对验证集和测试集的样本进行预测,得到香蕉单产预测结果,并与产量实测值进行比较分析。

2.4 模型性能评价指标

选取决定系数 (Coefficient of determination, R^2)、均方根误差 (Root mean square error, RMSE) 和平均绝对误差 (Mean absolute error, MAE) 作为香蕉估产模型预测性能的评定指标。其中 RMSE 和 MAE 直接体现模型估算误差,决定系数 R^2 体现估算值和实测值之间变化趋势的拟合程度。决定系数越大、均方根误差和平均绝对误差越小,表明模型的预测效果越好。

3 预测结果对比及分析

3.1 参数寻优方法对比

为了验证 BSO 算法在 SVR 参数寻优上的优势,分别选取传统寻优算法:网格搜索法 (Grid search,

GS)^[40] 以及群智能优化算法:灰狼优化算法 (Grey wolf optimizer, GWO)^[41] 与 BSO 算法进行比较,并保证一致的计算机软硬件环境。其中操作系统为 Windows 10, CPU 采用 Intel i7-8700 处理器,主频为 3.2 GHz,内存容量为 8 GB,算法运行平台为 Python 3.7.6。

经过大量实验验证,确定 BSO、GS 和 GWO 方法的最优初始参数设置如下:BSO 的初始随机产生的个体数量设置为 20,分类类别设定为 5,最大迭代次数设置为 20, p_a 、 p_b 、 p_c 和 p_d 均设定为 0.5。GWO 的迭代次数设置为 20,狼群数量设定为 20。采用 GS 方法对 SVR 参数 C 和 g 寻优时, C 设置为 $[0, 10]$, g 设置为 $[0, 5]$,步长设为 0.2。

采用 GS、GWO 和 BSO 对 SVR 模型的参数 C 和 g 进行优化 (分别表示为 GS-SVR、GWO-SVR 和 BSO-SVR),最优参数和运行时间如表 3 所示。

表 3 模型最优参数和运行时间						
Tab. 3 Optimal parameters and running time of models						
模型	C		g		运行时间/s	
	2019 年	2020 年	2019 年	2020 年	2019 年	2020 年
GS-SVR	5.4	9.2	0.2	0.2	5.161	5.164
GWO-SVR	9.145	9.018	0.680	0.091	0.959	0.980
BSO-SVR	9.616	3.333	0.181	0.196	0.320	0.331

利用 GS-SVR、GWO-SVR 和 BSO-SVR 模型对香蕉验证样本和测试样本的单产进行预测,预测结果的散点图如图 4 所示。由图 4 可知,对于 2019

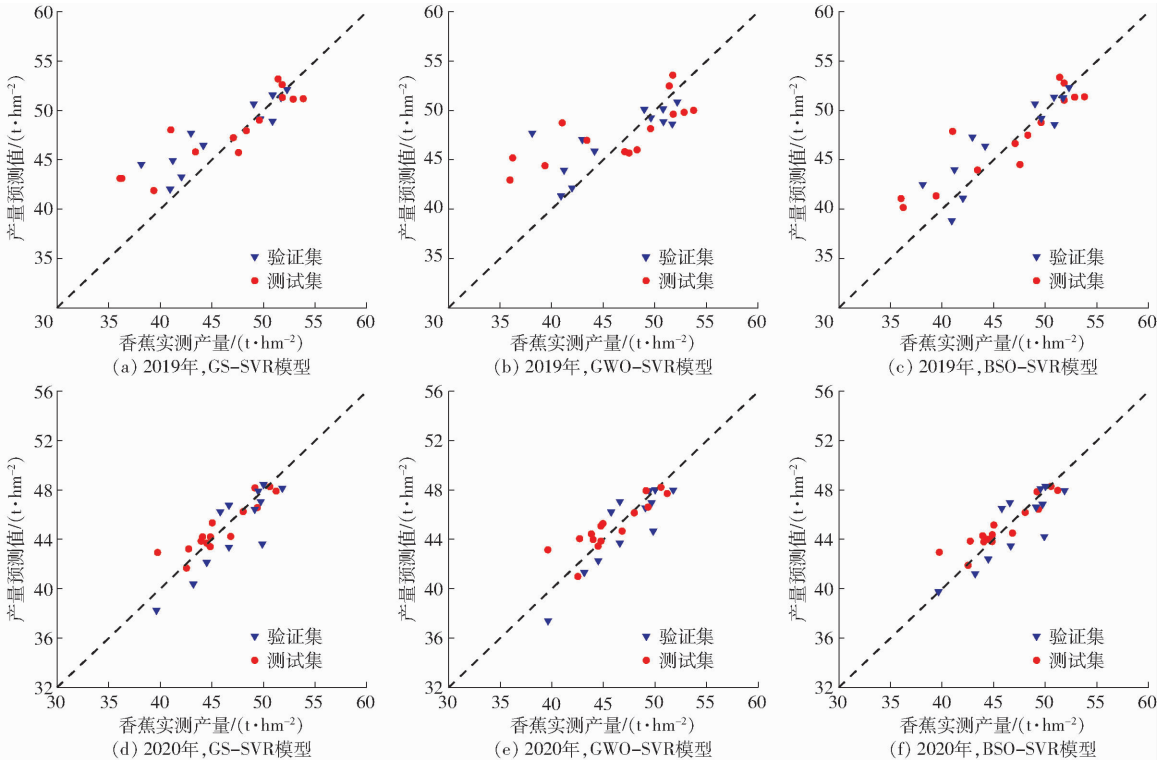


图 4 GS-SVR、GWO-SVR 和 BSO-SVR 模型的预测结果

Fig. 4 Prediction results of GS-SVR, GWO-SVR and BSO-SVR models

年和 2020 两个年度的香蕉产量估算,依靠 BSO-SVR 模型得到的验证集和测试集的预测结果均更加逼近于实测值,相比其他两种模型,BSO-SVR 模型体现了更好的预测效果。

采用指标 R^2 、RMSE 和 MAE 对此 3 种模型的预测结果进行评价,如图 5 所示。由图 5 可知,对于验证集,BSO-SVR 模型的 R^2 在 2019 年和 2020 年分别为 0.777 和 0.793;对于测试集,BSO-SVR 模型

的 R^2 在 2019 和 2020 年分别为 0.765 和 0.636,除 2019 年 BSO-SVR 模型的 R^2 略低于 GS-SVR 模型 ($R^2=0.797$) 外,其他均优于 GWO-SVR 模型和 GS-SVR 模型,此外,与 GWO-SVR 模型和 GS-SVR 模型相比,2019—2020 年 BSO-SVR 模型的 RMSE 和 MAE 整体表现最优,表明 BSO-SVR 模型的预测结果误差小,更接近实际值,具有较高的预测精度。

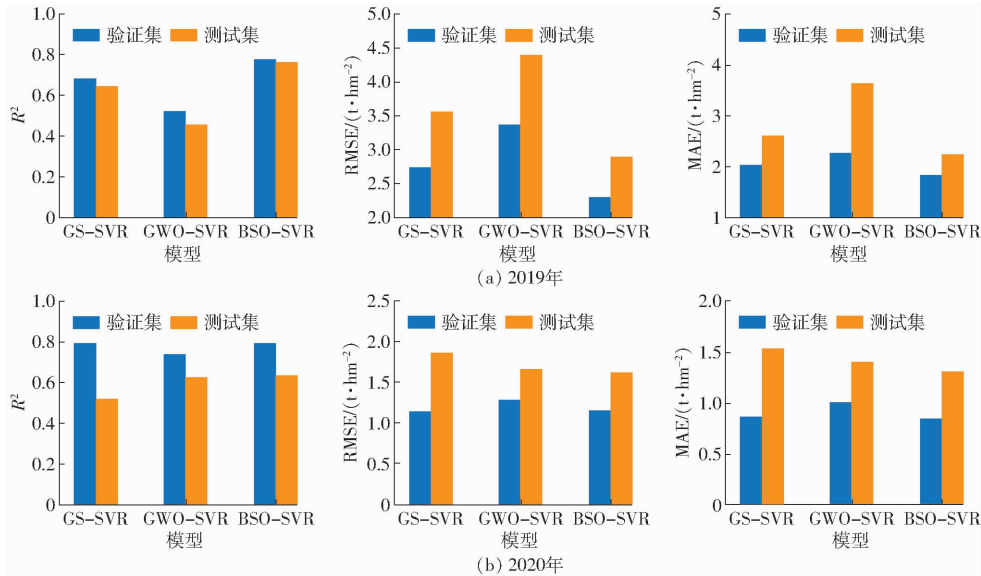


图 5 GS-SVR、GWO-SVR 和 BSO-SVR 模型性能对比
Fig. 5 Performance comparison of GS-SVR, GWO-SVR and BSO-SVR models

尽管 2019 年存在 BSO-SVR 模型的香蕉产量拟合效果逊色于 GS-SVR 模型的情况,但对于不同年份的估算误差均优于 GS-SVR。此外,网格算法是一种暴力搜索最优点方法,效率低且运行时间长,在本研究中,利用 GS 算法寻找最佳参数所需的时间比 BSO 算法高出 10 倍以上(表 3)。GWO-SVR 模型在运行时间方面和 BSO-SVR 模型相当,但在香蕉产量的预测效果上明显劣于其他两种寻优算法。总体而言,BSO-SVR 模型的性能整体表现最佳,其次是 GS-SVR 模型,而 GWO-SVR 模型的整体拟合效果最不理想。

3.2 预测模型性能比较

为了验证 BSO-SVR 模型在香蕉估产上的预测效果和使用 SVR 算法在小样本情况下的优越性和适用性,采用传统的岭回归模型(Ridge regression, RR)^[42-43]和偏最小二乘回归模型(Partial least squares regression, PLSR)^[44-45]与本文提出的方法进行对比。

RR 模型和 PLSR 模型均在 Python 软件的 scikit-learn 库中运行,参数采用默认的参数设置。使用训练集对 RR 模型和 PLSR 模型进行训练,将验证集和测试集代入训练好的 RR 模型和 PLSR 模型

中,对预测模型进行验证和精度评定,3 种预测模型的预测结果如图 6 所示。为评定模型预测效果,利用香蕉单产实测值与预测值计算 R^2 、RMSE 和 MAE,计算结果如图 7 所示。

由图 7 可知,在 2019 年,BSO-SVR 模型的 R^2 最高,其次是 RR 模型,PLSR 模型最差,其中 BSO-SVR 模型验证集和测试集的 R^2 均达到 0.75 以上,分别比 RR 模型提高 0.113 和 0.174,分别比 PLSR 模型提高 0.192 和 0.184。同时,相较于 RR 模型和 PLSR 模型,BSO-SVR 模型的 RMSE 和 MAE 最低,表明 BSO-SVR 模型在 2019 年香蕉产量预测方面有很好的效果。在 2020 年,BSO-SVR 模型总体表现最佳,验证集和测试集的平均 R^2 为 0.715,验证集和测试集的 R^2 分别高于 RR 模型 0.035 和 0.014,优于 PLSR 模型 0.040 和 0.035,同时 BSO-SVR 模型的 RMSE 和 MAE 也整体表现最优。BSO-SVR 模型在 2019 年的预测效果优于 2020 年的预测效果,这可能是因为与 2020 年香蕉各地块的单产数据相比,2019 年香蕉的单产数据的差异性和波动性较大(表 2),而 SVR 模型更适用于空间变异性较大的样本。

综合图 7 中对香蕉单产的各项预测精度评价指

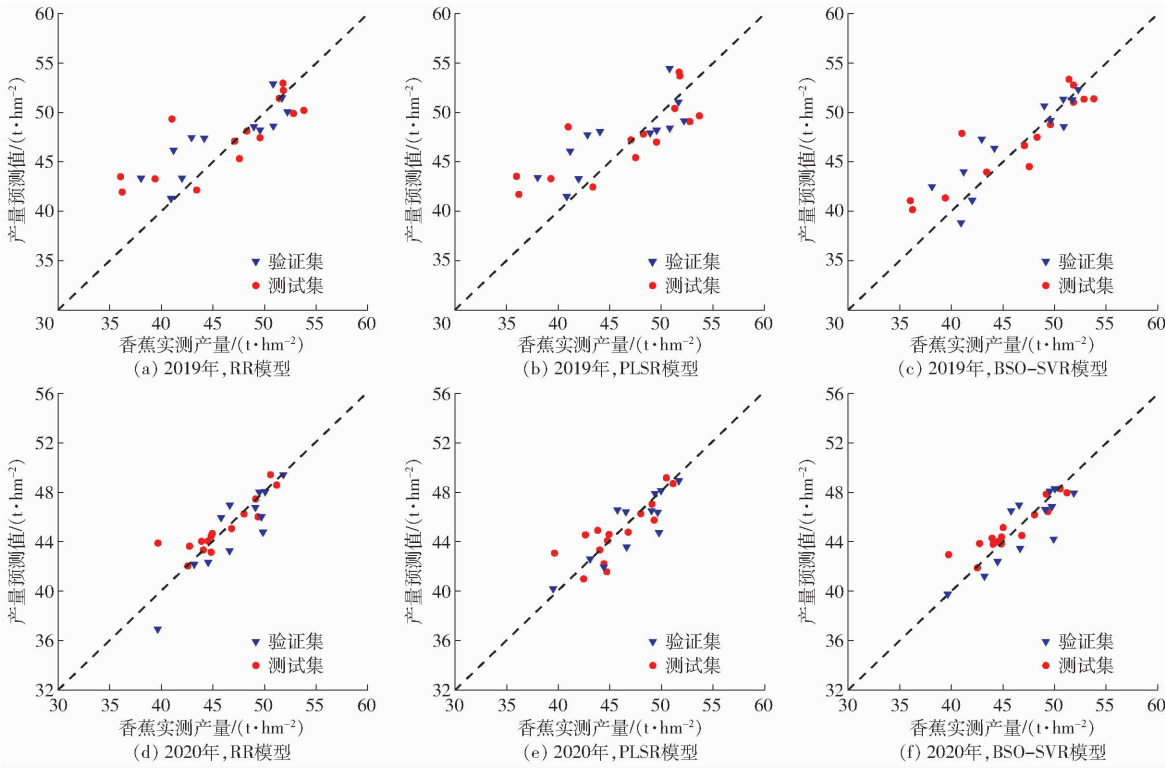


图 6 RR、PLSR 和 BSO-SVR 模型的预测结果
Fig. 6 Prediction results of RR, PLSR and BSO-SVR models

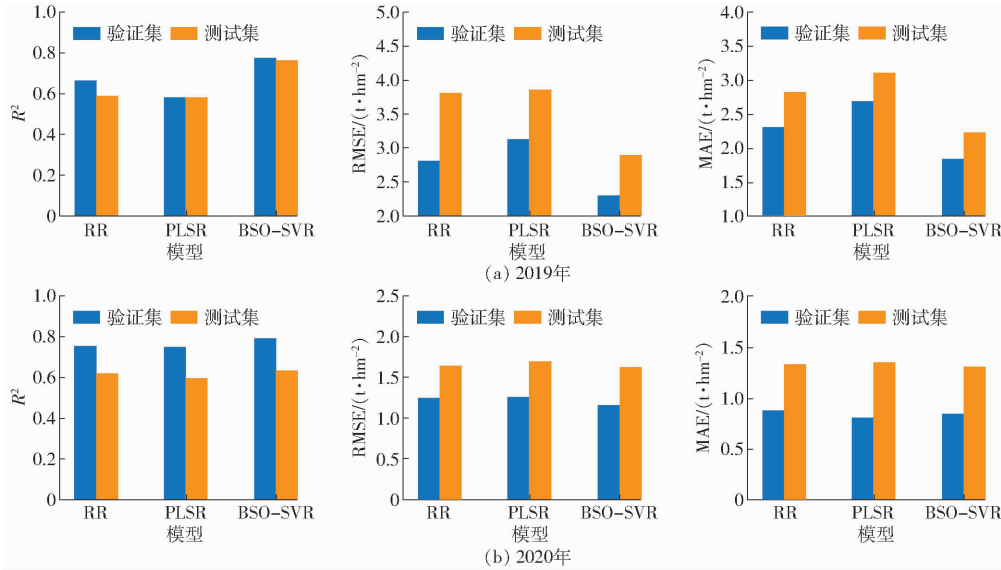


图 7 RR、PLSR 和 BSO-SVR 模型性能对比
Fig. 7 Performance comparison of RR, PLSR and BSO-SVR models

标得出,BSO-SVR 模型的预测效果最佳,PLSR 模型的预测结果最差,RR 模型的预测精度介于两种模型之间。BSO-SVR 模型取得较好预测效果的原因可能是 SVR 模型能将回归问题转换为二次规划问题,摆脱了容易陷入局部最优的问题,很适合处理田块尺度的小样本数据,而且,SVR 模型能捕捉 NDVI 时间序列特征间的非线性关系,从而提升了香蕉估产性能。对于 RR 模型和 PLSR 模型,尽管 RR 模型具有较强的预测稳定性,但该方法是通过

放弃最小二乘的无偏性,以损失部分信息、降低精度为代价提高拟合稳健性。PLSR 模型预测结果的波动性较低,而且对共线性起到了一定的消除作用,但该方法在建模过程中损失了数据信息,并降低了计算效率。

4 结论

(1)结合阈值分割和形态学操作方法,能有效去除研究区田块尺度的云和云阴影覆盖区域,消除

了云和云阴影等因素对香蕉估产结果的影响。

(2)与网格搜索法和灰狼优化算法相比,头脑风暴优化算法能精准快速地获取 SVR 模型惩罚因子和核函数参数的最优值,BSO-SVR 模型验证集的 R^2 在 2019 年和 2020 年分别为 0.777 和 0.793,测试集的 R^2 在 2019 年和 2020 年分别为 0.765 和 0.636,2019 年和 2020 年运行时间分别为 0.320 s 和 0.331 s。

(3)相较于传统的岭回归模型和偏最小二乘回归模型,BSO-SVR 模型具有最高的预测精度,其次是 RR 模型,PLSR 模型的表现最不理想。2019 年 BSO-SVR 模型验证集和测试集的 R^2 分别比 RR 模型提高 0.113 和 0.174,分别比 PLSR 模型提高 0.192 和 0.184。2020 年 BSO-SVR 模型验证集和测试集的 R^2 分别高于 RR 模型 0.035 和 0.014,优于 PLSR 模型 0.040 和 0.035。

参 考 文 献

- [1] CHEN Yanlu, XIE Ling, LONG Yanyan, et al. A new species and two new Chinese records of *Ochroconis* from sugarcane and banana rhizosphere in Guangxi, China[J]. *Mycoscience*, 2020, 61(6): 307-314.
- [2] SINHA P, ROBSON A, SCHNEIDER D, et al. The potential of in-situ hyperspectral remote sensing for differentiating 12 banana genotypes grown in Uganda[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 167: 85-103.
- [3] 陈仲新, 任建强, 唐华俊, 等. 农业遥感研究应用进展与展望[J]. *遥感学报*, 2016, 20(5): 748-767.
CHEN Zhongxin, REN Jianqiang, TANG Huajun, et al. Progress and perspectives on agricultural remote sensing research and applications in China[J]. *Journal of Remote Sensing*, 2016, 20(5): 748-767. (in Chinese)
- [4] SOARES J D R, PASQUAL M, LACERDA W S, et al. Comparison of techniques used in the prediction of yield in banana plants[J]. *Scientia Horticulturae*, 2014, 167: 84-90.
- [5] SOARES J D R, PASQUAL M, LACERDA W S, et al. Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants[J]. *Scientia Horticulturae*, 2013, 155: 24-29.
- [6] 史舟, 梁宗正, 杨媛媛, 等. 农业遥感研究现状与展望[J]. *农业机械学报*, 2015, 46(2): 247-260.
SHI Zhou, LIANG Zongzheng, YANG Yuanyuan, et al. Status and prospect of agricultural remote sensing[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46(2): 247-260. (in Chinese)
- [7] VAN K T, KASSAHUN A, CATAL C. Crop yield prediction using machine learning: a systematic literature review[J]. *Computers and Electronics in Agriculture*, 2020, 177: 105709.
- [8] CHENG Zhiqiang, MENG Jihua, WANG Yiming, et al. Improving spring maize yield estimation at field scale by assimilating time-series HJ-1 CCD data into the WOFOST model using a new method with fast algorithms[J]. *Remote Sensing*, 2016, 8(4): 303.
- [9] HUANG Jianxi, GOMEZ-DANS J L, HUANG Hai, et al. Assimilation of remote sensing into crop growth models: current status and perspectives[J]. *Agricultural and Forest Meteorology*, 2019, 276(277): 107609.
- [10] 徐新刚, 吴炳方, 蒙继华, 等. 农作物单产遥感估算模型研究进展[J]. *农业工程学报*, 2008, 24(2): 290-298.
XU Xin'gang, WU Bingfang, MENG Jihua, et al. Research advances in crop yield estimation models based on remote sensing[J]. *Transactions of the CSAE*, 2008, 24(2): 290-298. (in Chinese)
- [11] SHARIFI A. Yield prediction with machine learning algorithms and satellite images[J]. *Journal of the Science of Food and Agriculture*, 2020, 101(3): 891-896.
- [12] NEVAVUORI P, NARRA N, LIPPING T. Crop yield prediction with deep convolutional neural networks[J]. *Computers and Electronics in Agriculture*, 2019, 163: 104859.
- [13] HAGHVERDI A, WASHINGTON-ALLEN R A, LEIB B G. Prediction of cotton lint yield from phenology of crop indices using artificial neural networks[J]. *Computers and Electronics in Agriculture*, 2018, 152: 186-197.
- [14] SCHWALBERT R A, AMADO T, CORASSA G, et al. Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern Brazil[J]. *Agricultural and Forest Meteorology*, 2020, 284: 107886.
- [15] ARCHONTOULIS S V, CASTELLANO M J, LICHT M A, et al. Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt[J]. *Crop Science*, 2020, 60(2): 721-738.
- [16] 高中灵, 徐新刚, 王纪华, 等. 基于时间序列 NDVI 相似性分析的棉花估产[J]. *农业工程学报*, 2012, 28(2): 148-153.
GAO Zhongling, XU Xin'gang, WANG Jihua, et al. Cotton yield estimation based on similarity analysis of time-series NDVI[J]. *Transactions of the CSAE*, 2012, 28(2): 148-153. (in Chinese)
- [17] ZHANG Jiangzhou, LI Baoshen, ZHANG Junling, et al. Organic fertilizer application and Mg fertilizer promote banana yield and quality in an Udic Ferralsol[J]. *Plos One*, 2020, 15(3): e0230593.
- [18] STEVENS B, DIELS J, BROWN A, et al. Banana biomass estimation and yield forecasting from non-destructive measurements for two contrasting cultivars and water regimes[J]. *Agronomy*, 2020, 10(9): 1435.
- [19] DE S A V, BONINI N A, CABRERA P J, et al. Artificial neural network modelling in the prediction of bananas' harvest[J]. *Scientia Horticulturae*, 2019, 257: 108724.
- [20] AGHIGHI H, AZADBAKHT M, ASHOURLLOO D, et al. Machine learning regression techniques for the silage maize yield prediction using time-series images of landsat 8 OLI[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(12): 4563-4577.
- [21] MERCIER A, BETBEDER J, BAUDRY J, et al. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 163: 231-256.
- [22] ASHOURLLOO D, SHAHRABI H S, AZADBAKHT M, et al. Automatic canola mapping using time series of Sentinel 2 images

- [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 156: 63–76.
- [23] RAHMAN M M, ROBSON A. Integrating landsat-8 and Sentinel-2 time series data for yield prediction of sugarcane crops at the block level[J]. Remote Sensing, 2020, 12(8): 1313.
- [24] LIU Jianguai, SHANG Jiali, QIAN Budong, et al. Crop yield estimation using time-series MODIS data and the effects of cropland masks in Ontario, Canada[J]. Remote Sensing, 2019, 11(20): 2419.
- [25] AMELINE M, FIEUZAL R, BETBEDER J, et al. Estimation of corn yield by assimilating SAR and optical time series into a simplified agro-meteorological model: from diagnostic to forecast [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(12): 4747–4760.
- [26] DRUCKER H, BURGESS C J C, KAUFMAN L, et al. Support vector regression machines[M]//MOZER M C, JORDAN M I, PETSCHKE T. Advances in neural information processing systems, 1997: 155–161.
- [27] MOUNTRAKIS G, IM J, OGOLE C. Support vector machines in remote sensing: a review [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 247–259.
- [28] 张建伟, 江琦, 刘轩然, 等. 基于 PSO-SVM 算法的梯级泵站管道振动响应预测[J]. 农业工程学报, 2017, 33(11): 75–81.
- ZHANG Jianwei, JIANG Qi, LIU Xuanran, et al. Prediction of vibration response for pipeline of cascade pumping station based on PSO-SVM algorithm[J]. Transactions of the CSAE, 2017, 33(11): 75–81. (in Chinese)
- [29] 蔡振闹, 吕信恩, 陈慧灵. 基于反向细菌优化支持向量机的躯体化障碍预测模型[J]. 吉林大学学报(工学版), 2018, 48(3): 936–942.
- CAI Zhenao, LÜ Xinen, CHEN Huiling. Prediction model of somatization disorder based on an oppositional bacterial foraging optimization based support vector machine[J]. Journal of Jilin University (Engineering and Technology Edition), 2018, 48(3): 936–942. (in Chinese)
- [30] 张婷婷, 于明, 李宾, 等. 基于 Wavelet 降噪和支持向量机的锂离子电池容量预测研究[J]. 电工技术学报, 2020, 35(14): 3126–3136.
- ZHANG Tingting, YU Ming, LI Bin, et al. Capacity prediction of lithium-ion batteries based on Wavelet noise reduction and support vector machine[J]. Transactions of China Electrotechnical Society, 2020, 35(14): 3126–3136. (in Chinese)
- [31] CHEN Weitao, LI Xianju, WANG Lizhe. Fine land cover classification in an open pit mining area using optimized support vector machine and WorldView-3 imagery[J]. Remote Sensing, 2020, 12(1): 82.
- [32] TAN Nian, WANG Xueshun, HUANG Anmin, et al. Wood density prediction of *Cunninghamia lanceolata* based on gray wolf algorithm SVM and NIR[J]. Scientia Silvae Sinicae, 2018, 54(12): 137–141.
- [33] GARCIA N P J, GARCIA-GONZALO E, ARBAT G, et al. A new predictive model for the filtered volume and outlet parameters in micro-irrigation sand filters fed with effluents using the hybrid PSO-SVM-based approach[J]. Computers and Electronics in Agriculture, 2016, 125: 74–80.
- [34] TIKHAMARINE Y, MALIK A, PANDEY K, et al. Monthly evapotranspiration estimation using optimal climatic parameters: efficacy of hybrid support vector regression integrated with whale optimization algorithm [J]. Environmental Monitoring and Assessment, 2020, 192(11): 1–19.
- [35] ZHENG Jing, FAN Junliang, ZHANG Fucang, et al. Estimation of rainfed maize transpiration under various mulching methods using modified Jarvis-Stewart model and hybrid support vector machine model with whale optimization algorithm [J]. Agricultural Water Management, 2021, 249: 106799.
- [36] WANG Yingzhou, NI Yulong, LU Shuai, et al. Remaining useful life prediction of lithium-ion batteries using support vector regression optimized by artificial bee colony[J]. IEEE Transactions on Vehicular Technology, 2019, 68(10): 9543–9553.
- [37] SHI Yuhui. Brain storm optimization algorithm[M]//TAN Y, SHI Y, CHAI Y, et al. Advances in swarm intelligence, Pt I. 2011: 303–309.
- [38] 沈林. 基于头脑风暴算法优化的 v-SVR 的研究及应用[D]. 兰州: 兰州大学, 2014.
- SHEN Lin. The study and application of v-SVR based brain storming algorithm—a case of stock market [D]. Lanzhou: Lanzhou University, 2014. (in Chinese)
- [39] CHENG Shi, QIN Qande, CHEN Junfeng, et al. Brain storm optimization algorithm: a review[J]. Artificial Intelligence Review, 2016, 46(4): 445–458.
- [40] 杜聪, 邵建华, 杨薇, 等. 网格搜索法优化的支持向量机室内可见光定位[J]. 激光杂志, 2021, 42(3): 104–109.
- DU Cong, SHAO Jianhua, YANG Wei, et al. Support vector machine indoor visible light positioning optimized by grid search method[J]. Laser Journal, 2021, 42(3): 104–109. (in Chinese)
- [41] MIRJALILI S, MIRJALILI S M, LEWIS A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46–61.
- [42] 张智韬, 王海峰, ARNON K, 等. 基于岭回归的土壤含水率高光谱反演研究[J]. 农业机械学报, 2018, 49(5): 240–248.
- ZHANG Zhitao, WANG Haifeng, ARNON K, et al. Inversion of soil moisture content from hyperspectra based on ridge regression[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(5): 240–248. (in Chinese)
- [43] 周敏姑, 邵国敏, 张立元, 等. 无人机多光谱遥感反演冬小麦 SPAD 值[J]. 农业工程学报, 2020, 36(20): 125–133.
- ZHOU Mingu, SHAO Guomin, ZHANG Liyuan, et al. Inversion of SPAD value of winter wheat by multispectral remote sensing of unmanned aerial vehicles[J]. Transactions of the CSAE, 2020, 36(20): 125–133. (in Chinese)
- [44] 张智韬, 王海峰, 韩文霆, 等. 基于无人机多光谱遥感的土壤含水率反演研究[J]. 农业机械学报, 2018, 49(2): 173–181.
- ZHANG Zhitao, WANG Haifeng, HAN Wenting, et al. Inversion of soil moisture content based on multispectral remote sensing of UAVs[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(2): 173–181. (in Chinese)
- [45] ZHANG Yao, LI Minzan, ZHENG Lihua, et al. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm[J]. Geoderma, 2019, 333: 23–34.