

# 人工蜂群算法优化SVR的叶面积指数反演

周晓雪<sup>1,3</sup>, 李楠<sup>1,3</sup>, 潘耀忠<sup>1,2,3</sup>, 孙莉昕<sup>3</sup>

1. 遥感科学国家重点实验室 北京师范大学地理科学学部, 北京 100875;
2. 青海师范大学 地理科学学院, 西宁 810016;
3. 北京师范大学 地理科学学部 遥感科学与工程研究院, 北京 100875

**摘要:** 支持向量机回归SVR (Support Vector Regression) 方法作为叶面积指数反演的一种新思路, 在LAI反演中具有一定的应用价值和前景, 但SVR算法中惩罚系数 $C$ 、核函数宽度参数 $g$ 、不敏感损失函数参数 $\epsilon$ 的取值对回归精度有显著的影响。本文提出了一种基于人工蜂群算法ABC (Artificial Bee Colony) 优化SVR参数的遥感影像叶面积指数反演方法。研究数据为美国土壤水分实验 (SMEX02) 2002年LAI实测数据和同期的Landsat 7 ETM+地表反射率数据, 为了验证ABC算法优化SVR各个参数对反演精度的影响, 建立了未优化参数 (SVR)、优化单个参数 (ABC-SVR- $C$ , ABC-SVR- $g$ , ABC-SVR- $\epsilon$ )、优化3个参数 (ABC-SVR) 的3类LAI反演模型, 并比较了其回归拟合精度。在此基础上, 分析了3个关键参数对LAI反演模型精度的敏感性, 并对ABC算法优化SVR模型的精度进行显著性检验。研究表明: (1) 相比未优化参数模型, ABC算法优化模型具有更高的反演精度, 优化3个参数优于优化单个参数, 回归直线斜率 $k$ 达到0.797、决定系数 $r^2$ 达到0.775。(2) SVR的3个关键参数对模型精度都有影响, 相较参数 $C$ 和 $g$ , 参数 $\epsilon$ 引起模型精度的不确定性更高。(3) 95%的置信区间下, ABC-SVR模型与SVR模型的回归直线斜率 $k$ 、 $r^2$ 、RMSE的差异显著性检验 $P$ 值均小于0.005, ABC算法显著改善了SVR模型的精度。

**关键词:** 支持向量机回归SVR, 人工蜂群算法ABC, 参数优化, Landsat 7, 叶面积指数LAI

**引用格式:** 周晓雪, 李楠, 潘耀忠, 孙莉昕. 2022. 人工蜂群算法优化SVR的叶面积指数反演. 遥感学报, 26(4): 766-780

Zhou X X, Li N, Pan Y Z and Sun L X. 2022. Optimized SVR based on artificial bee colony algorithm for leaf area index inversion. National Remote Sensing Bulletin, 26(4): 766-780 [DOI:10.11834/jrs.20229298]

## 1 引言

叶面积指数LAI (Leaf Area Index) 是表征植被冠层结构的基本参数之一, 定义为单位地表面积上植物叶片单面总面积 (Chen和Black, 1992), 是反映植被群体长势状况的重要指标。目前, LAI广泛应用于作物长势监测、能量交换与平衡、精准农业以及全球气候变化研究等多方面 (邢著荣等, 2010; Latifi和Galos, 2010; 何友铸等, 2015)。

叶面积指数观测主要分为地面测量和遥感监测两种方式, 传统的地面测量在室内通过比叶重法或面积量算法来获取样本点的LAI, 这种方法精度较高, 但费时费力, 仅适用于小范围区域, 不能满足大范围作物长势监测、病虫害监测、全

球气候变化预测等的需求 (刘晓臣等, 2008)。遥感监测技术具有大面积实时观测、成本低等特点, 目前已经广泛应用于LAI的反演研究。遥感反演LAI的主流方法有两种: 物理模型法和统计模型法 (方秀琴和张万昌, 2003)。物理模型法主要是基于植被的辐射传输模型, 具有很强的物理基础, 不受植被类型影响, 有较好的适用性, 但是模型输入参数较多且获取困难, 遥感反演结果不确定, 存在所谓的遥感病态反演问题 (梁顺林等, 2016)。统计模型法通过建立地表反射率数据或植被指数与LAI实际观测值的经验关系模型 $LAI=f(x)$ 来反演LAI, 这种方法虽然对植被的类型、生长环境、土壤背景都具有依赖性, 但是该方法相对简单, 具有良好的操作性, 易于推广使用。如Durbha等 (2007)

收稿日期: 2019-08-20; 预印本: 2020-03-24

基金项目: 国家重点研发计划 (编号: 2018YFC1504603)

第一作者简介: 周晓雪, 研究方向为农业遥感。E-mail: 201821051195@mail.bnu.edu.cn

通信作者简介: 潘耀忠, 研究方向为农业遥感。E-mail: pyz@bnu.edu.cn

将统计模型方法中的支持向量回归方法SVR (Support Vector Regression) 与PROSAIL模型结合反演LAI, 采用站点实测LAI数据验证, LAI预测值与实测值的RMSE达到0.64。Tuia等(2011)使用支持向量回归方法反演遥感影像中的3种生物物理参数: 生物叶片叶绿素含量、叶面积指数和植被覆盖度, 结果表明使用非参数方法反演生物物理参数时, SVR方法是一个较好的选择方案。郭琳等(2010)用环境星CCD数据, 建立了归一化植被指数NDVI (Normalized Difference Vegetation Index) 与LAI的指数模型、对数模型、支持向量回归(SVR)模型, 用于估算甘蔗的LAI值, 结果发现支持向量回归模型反演精度最高。宰松梅等(2011)以甜椒叶片长度和最大宽度建立了估算甜椒叶面积的支持向量模型、线性回归模型、人工神经网络模型, 结果表明支持向量机模型能较真实地估算甜椒叶面积的实际大小, 估算精度达到了0.996。梁栋等(2013)在冬小麦不同的生育期, 分别利用支持向量机回归方法, 建立了植被指数NDVI、RVI (Ratio Vegetation Index) 和4个波段(蓝、绿、红和近红外)与LAI的关系模型, 结果表明支持向量机回归预测模型具有更好的拟合效果, 且对冬小麦的多个生育期均具有较好的适用性。

上述研究表明支持向量机回归算法应用到LAI的反演中具有可行性, 与其他反演算法相比具有较好的反演精度。但是其中大多数研究较少考虑到SVR参数的优化问题, 有研究表明, SVR中的惩罚系数 $C$ 、核函数宽度 $g$ 、不敏感损失函数 $\epsilon$ 等3个参数的取值对回归精度具有显著的影响, 如果不根据数据特点设定特定参数, 将显著影响模型精度(闫国华和朱永生, 2009; 成鹏和汪西莉, 2011)。但现有SVR反演LAI参数的研究仅优化 $C$ 、 $g$ 两个参数, 并未分析不同参数对模型精度的敏感性。

人工智能优化算法是目前流行的参数优化算法, 如蚁群算法(Dorigo和Gambardella, 1997)、遗传算法(Holland, 1975)、粒子群算法(Kennedy和Eberhart, 1995)、模拟退火算法(Steinbrunn等, 1997)、人工蜂群算法(Karaboga, 2005)等。在Karaboga等的研究中(Karaboga和Basturk, 2007;

Karaboga和Akay, 2009; Karaboga等, 2014), 通过大量的实验证明了人工蜂群算法ABC (Artificial Bee Colony) 相比其他智能优化算法具有更好的优化能力, 不易陷入局部最优。目前已有学者将人工蜂群算法用于优化支持向量机参数。如于明和艾月乔(2012)提出一种基于人工蜂群算法的SVM参数优化方法, 用于计算机笔迹鉴别, 并与遗传算法、蚁群算法、标准粒子群算法优化的SVM对比, 结果证明了ABC算法优化的SVM具有更高的分类正确率。王琳等(2014)基于短期交通流量数据, 采用人工蜂群算法、粒子群算法优化SVR预测模型参数, 实验结果表明, 与粒子群优化算法相比, 人工蜂群算法优化的SVR预测方法能够更有效地克服局部最优解, 预测精度较高, 运行时间短, 具有较强的学习能力和推广能力。李楠等(2018)提出了一种基于人工蜂群算法改进SVM参数的遥感分类方法, 并比较了人工蜂群算法、遗传算法和粒子群算法优化的SVM分类精度和效率, 结果显示人工蜂群算法优化的SVM具有更高的分类精度。综上, 本文选择ABC算法对SVR关键参数进行优化, 进一步支撑SVR方法在LAI领域中更好的应用。

基于上述问题, 本文利用人工蜂群算法优化支持向量机回归3个关键参数(惩罚系数 $C$ 、核函数宽度 $g$ 、不敏感损失函数 $\epsilon$ ), 构建人工蜂群算法优化的SVR回归模型, 以获得更精确的LAI反演结果。为了验证模型效果, 构建了3种LAI反演模型并对反演精度进行了综合分析: 未优化参数的SVR模型、ABC算法分别优化SVR单个参数的模型(ABC-SVR- $C$ 、ABC-SVR- $g$ 、ABC-SVR- $\epsilon$ )、ABC算法优化SVR的3个参数( $C$ 、 $g$ 、 $\epsilon$ )的模型(ABC-SVR)。此外分析了SVR的3个关键参数对LAI反演模型精度的敏感性, 并对ABC算法优化SVR模型的精度进行显著性检验。

## 2 实验区域及数据

### 2.1 实验区域概况

本研究基于美国农业部2002年在爱荷华州开展的土壤水分实验SMEX02 (Soil Moisture Experiment 2002), 研究区如图1所示, 该区域内95%的植被为农作物, 主要农作物包括大豆和玉米。



图1 实验区及LAI地面实测点分布

Fig. 1 Study area and the spatial distribution of ground-based LAI measurement site

### 2.2 LAI野外实测数据

SMEX02实验过程中进行了地表植被参数样本数据的采集，其中在WC (Walnut Creek Watershed region) 实验区建立94个地面实测点，采集了包括叶面积指数等的作物生长数据和土壤数据，样本采集时间为2002-06-15到2002-07-09，其中LAI数据是使用LAI-2000进行5次测量的平均值。野外LAI测量点位分布如图1所示。

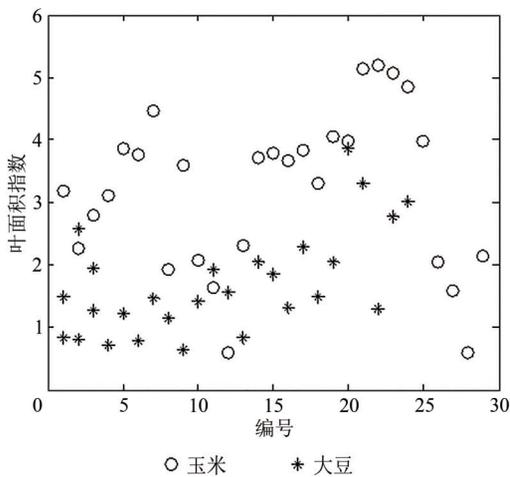


图2 LAI实际测量值及作物类别

Fig. 2 LAI measurement values and the category of crop

由于每个点位上的LAI测量数据不连续（每天只进行部分点位的LAI测量），因此，本实验选取了

LAI地面测量数据较多的2002-07-02和2002-07-08两个日期内的LAI数据进行实验，这两日的LAI地面观测总样本量分别为31个和25个，56个LAI实际测量值和对应作物类别如图2所示。由于样本数量限制，本研究将这两天数据合为一个共56个LAI实测数据的样本集，并根据随机抽样方法，划分训练集（70%）和测试集（30%）以构建和验证反演模型。

### 2.3 遥感影像数据

本研究根据LAI实测数据，选取两景接近其日期的Landsat 7 ETM+数据作为反演模型的遥感数据，进行叶面积指数（LAI）的反演，该过程需要准确获取每个影像像元的地表反射率值，因此，本文采用ENVI5.1中FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) 大气校正模型进行ETM+数据的大气校正。

两景数据分别为2002-07-01 (Path 26、Row 30) 和2002-07-08 (Path 27、Row 30)，图3为研究区域地表反射率影像。由于ETM+影像第6波段分辨率为60 m，与其他波段分辨率（30 m）不相符，故剔除该波段。此外，一些研究指出植被指数与LAI具有很强的相关性（武旭霞和谢巧云，2014；刘洋等，2013），因此本研究引入了NDVI、RVI、EVI 3种植被指数参与建立LAI反演模型，最终使用的所有遥感影像信息如表1。



(a) 2002-07-01 研究区域影像  
(a) The image of study area on July 1, 2002

(b) 2002-07-08 研究区域影像  
(b) The image of study area on July 8, 2002

图3 研究区域假彩色影像(4,3,2波段合成)

Fig. 3 False color image of study area (4,3,2 band synthesis)

表1 遥感影像数据信息

Table 1 Information of remote sensing image

来源	波段号	波段	波长/ $\mu\text{m}$	空间分辨率/m
ETM+	1	蓝	0.450—0.515	30
	2	绿	0.525—0.605	30
	3	红	0.630—0.690	30
	4	近红外	0.775—0.900	30
	5	中红外	1.550—1.750	30
	7	中红外	2.090—2.350	30
	—	NDVI	—	30
植被指数	—	RVI	—	30
—	—	EVI	—	30

注: 表中数据来源于 <https://eosps0.gsfc.nasa.gov/missions/landsat-7> [2019-08-20]。

### 3 实验方法

#### 3.1 支持向量机回归算法

支持向量机SVM (Support Vector Machine) 是Vpanik等所在的AT&T Bell实验室研究小组在1995年提出的一种基于统计学习理论的机器学习方法 (Cherkassky, 1997), SVM主要解决分类和回归分析两类问题 (王定成等, 2003; 臧淑英等, 2012)。由于实际应用中, 数据大部分都是非线性情况, 因此SVM算法中引入核函数将数据映射到高维线性空间, 分类则是指在该高维空间里构建一个超平面, 使得所有训练样本到该平面距离最大, 两侧距离超平面最近的样本点称为支持向量, 由支持向量可以计算超平面两侧的超平面 $\alpha_1$ 、 $\alpha_2$ , 这两个平面之间的距离越大, 表示分类结果越可靠。回归分析是指在该高维空间里构建超平面, 使得所有训练样本到该平面的距离最小, 超平面两侧

距离最远的样本点为支持向量, 同样可以计算得到两个超平面 $\alpha_1$ 、 $\alpha_2$ , 使得回归偏差最小。

SVM应用于回归问题被称为支持向量回归SVR (Support Vector Regression), 适用于样本数量有限、非线性的样本, 能较大程度解决过学习和局部极值等不足, 广泛应用于模式识别、图像处理、金融时序预测、故障检测等领域。支持向量回归分为线性回归和非线性回归 (Kwok, 1998), 以下对非线性回归进行解释。

对于给定的样本训练集:  $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$ ,  $n$ 为训练样本个数,  $x$ 为样本训练集中自变量,  $y$ 为样本训练集的因变量, 在本实验中为实测LAI值, 支持向量机非线性回归通过一个非线性映射函数 $\phi(x)$ , 将数据 $x$ 映射到一个高维线性空间, 再进行回归, 回归函数如下

$$f(x) = \omega \cdot \phi(x) + b \quad (1)$$

式中,  $f(x)$ 为回归超平面,  $\omega$ 和 $b$ 是回归超平面的系数。

引入Vapnik提出的 $\varepsilon$ -不敏感损失函数 (Cherkassky, 1997)

$$L_\varepsilon(f(x), y) = \begin{cases} 0 & , |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & , \text{其他} \end{cases} \quad (2)$$

$\varepsilon$ 为不敏感系数, 用于控制拟合精度, 式(2)可表达为, 如果 $x$ 的预测值 $f(x)$ 与实际值 $y$ 之差小于 $\varepsilon$ , 则认为预测值 $f(x)$ 是无损失的, 反之, 则认为 $f(x)$ 预测值有损失。

引入惩罚系数 $C$ 和松弛变量因子 $\xi_i, \xi_i^*$ , 最后的优化目标为

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

$$\text{s.t.} \begin{cases} y_i - (\omega \cdot \phi(x_i) + b) \leq \varepsilon + \xi_i \\ (\omega \cdot \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 (i = 1, 2, \dots, n) \end{cases} \quad (4)$$

式中,  $\xi_i, \xi_i^*$  为第  $i$  个样本的松弛系数。

利用拉格朗日乘子, 将上述函数转化为其对偶问题

$$\max \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \cdot K(x_i, x_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad (5)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (6)$$

式中,  $\alpha_i, \alpha_i^*$  为拉格朗日系数向量,  $K(x_i, x_j)$  为核函数。

通过对对偶问题的求解, 得到支持向量机的最终函数形式

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (7)$$

式中,  $K(x, x_i)$  为核函数, 在本实验中采用径向基核函数 RBF (Radial Basis Function), 函数形式如下

$$K(x, x_i) = \exp \left\{ -\frac{\|x - x_i\|^2}{g^2} \right\} \quad (8)$$

式中,  $g$  为径向基核函数宽度参数。

综上, 影响 SVR 模型回归性能的参数包括惩罚系数  $C$ 、核函数宽度参数  $g$ 、不敏感损失函数参数  $\varepsilon$  等。惩罚系数  $C$  反映算法对超出  $\varepsilon$  管道的样本数据的惩罚程度,  $C$  越大模型越复杂, 对数据的拟合程度越高; RBF 的宽度参数  $g$ , 控制函数径向作用范围; 不敏感损失函数参数  $\varepsilon$  通过控制回归误差管道的大小, 影响支持向量的数目,  $\varepsilon$  过大, 支持向量数就少, 可能导致模型过于简单, 学习精度不够;  $\varepsilon$  过小, 回归精度较高, 但可能导致模型过于复杂, 得不到很好的推广能力。合理的参数组合才能得高精度的回归结果, 因此, 本文以人工蜂群算法对优化 SVR 关键参数 ( $C, g, \varepsilon$ ), 建立高精度的 SVR 反演模型。

### 3.2 人工蜂群算法

人工蜂群算法 ABC (Artificial Bee Colony) 是由 Karaboga 于 2005 年提出的一种模拟蜂群采蜜行为的启发式算法 (Karaboga 等, 2014), 该方法模

拟蜂群内部不同的分工和合作来完成整个采蜜过程, 目标是寻找到最优蜜源 (问题的最优解)。

人工蜂群组成主要包括 4 个部分:

(1) 蜜源: 一个蜜源与问题的一组解一一对应, 蜜源的花粉数量表示解的适应值, 适应度越好, 在实际问题中对应解则越佳。

(2) 引领蜂: 一个引领蜂储存一个蜜源的有关信息并以一定概率进行分享, 其数量与蜜源数量相等, 为蜂群规模的一半。

(3) 跟随蜂: 跟随蜂等在蜂巢里面通过与引领蜂分享相关信息找到蜜源进行采蜜。

(4) 侦查蜂: 在原有的蜜源被遗弃之后搜索蜂巢附近的新蜜源。

人工蜂群算法实现流程图如图 4 所示。

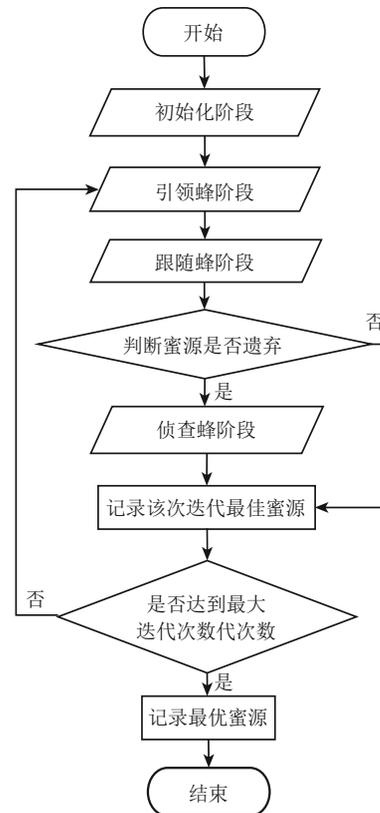


图 4 人工蜂群算法流程图

Fig. 4 The flow chart of artificial bee colony algorithm

初始化阶段: 设置求解问题解空间  $d$  维、蜂群规模  $NP$ 、引领蜂数目  $NP/2$ 、蜜源是否被遗弃控制参数  $limit$ 、算法迭代次数  $Max\ Cycle$ 。且 ABC 算法随机产生  $NP/2$  个初始解  $x_i, i=1, 2, \dots, NP/2$ , 每个  $x_i$  为一个  $d$  维向量, 并构造适应度函数, 判断每个蜜源的好坏程度。

引领蜂阶段: 引领蜂在蜜源邻域进行搜索,

搜索规则如式(9)所示:

$$v_{ij} = x_{ij} + r_{ij}(x_{ij} - x_{ij}) \quad (9)$$

式中,  $i, k \in (1, 2, \dots, NP/2), j \in (1, 2, \dots, d), k \neq i$ ,  $k$ 为随机选取的相邻蜜源, 要保证两蜜源不是同一蜜源,  $r_{ij}$ 为  $[-1, 1]$ 的随机数, 用于控制邻域范围。

跟随蜂阶段: 跟随蜂根据式(10)选择跟随的引领蜂

$$p_i = \frac{fit_i}{\sum_{j=1}^{NP/2} fit_j} \quad (10)$$

式中,  $fit_i$ 表示第 $i$ 解的适应度, 即第 $i$ 个蜜源的花蜜数量,  $p_i$ 第 $i$ 个蜜源被选择的概率。由上式可得, 当一个蜜源的适应度越大则被选择的概率越高, 跟随蜂选择引领蜂后, 同样按式(9)在蜜源邻域内进行搜索, 寻找更优蜜源, 即最优适应度对应的蜜源。

侦查蜂阶段: 在蜂群算法里面, 以参数 $limit$ 保存每个解未被更新的次数, 当引领蜂和跟随蜂完成搜索之后, 某蜜源未被更新次数达到 $limit$ , 放弃该蜜源, 对应的引领蜂变为侦查蜂, 假设被放弃的蜜源为 $x_i$ , 则侦查蜂按照式(11)产生一个新解代替旧解:

$$x_{ij} = \min_j(x_{ij}) + rand(0,1)(\max_j(x_{ij}) - \min_j(x_{ij})) \quad (11)$$

式中,  $j = (1, 2, \dots, d)$ ,  $x_{ij}$ 为解 $x_i$ 的第 $j$ 个元素值。

### 3.3 人工蜂群算法优化支持向量回归算法

支持向量回归算法中, 对于不同的问题需要根据实际情况选择合适的参数进行建模才能达到良好的训练效果, 如果参数选择不合适, 则SVR会出现“欠学习”或者“过拟合”情况。结合3.1节, 本文使用ABC算法对SVR的关键参数( $C, g, \varepsilon$ )进行优化, 在优化单个参数时, ABC算法中每个蜜源相当于一个一维解, 优化3个参数时每个蜜源相当于一个三维解, 算法目标是寻找到一组使SVR模型精度最高的参数。

与3.2节对应, 本文采用ABC算法优化SVR参数的具体步骤如下:

步骤1, 随机划分样本集为训练集和测试集(70%/30%);

步骤2, 初始化蜂群, 设置蜂群规模 $NP$ (蜜源数目等于引领蜂数目等于 $NP/2$ ), 蜜源未被更新最大次数 $limit$ , 以及最大迭代次数 $Max\ Cycle$ , 参数( $C, g, \varepsilon$ )的取值范围; 蜜源适应度函数设置为SVR算法对训练集采用 $k$ -折交叉验证(本文 $k=10$ )后的平均均方误差(PMSE)的倒数。 $k$ 折交叉验证方法将训练样本分割成 $k$ 个子样本, 一个单独子样本作验证, 其他 $k-1$ 个样本作训练。交叉验证重复 $k$ 次, 每个子样本验证一次, 平均 $k$ 次的验证结果作为最终的结果。本文中 $k$ -折交叉验证PMSE小, 蜜源对应适应度高, 反之则适应度低。

步骤3, ABC算法模拟蜂群进行采蜜, 完成引领蜂的邻域搜索, 跟随蜂选择引领蜂进行邻域搜索, 计算每组SVR参数对应的平均均方误差, 根据平均均方误差大小更新最佳参数解位置。比较每个参数解的未被更新次数 $trial$ 与 $limit$ 的大小, 判断该参数解是否丢弃, 并计算丢弃后新参数解的位置和适应度, 记录每次迭代最佳参数组合。

步骤4, 重复第3步, 直到迭代次数等于最大迭代次数 $Max\ Cycle$ , 记录最优蜜源位置即最佳参数组合解( $C, g, \varepsilon$ ), 跳出循环。

### 3.4 实验方案

在本研究中, 为了探究人工蜂群算法对SVR参数优化的效果, 建立3类模型: 第1类为未进行参数优化的SVR模型, 第2类为ABC算法优化单一参数的SVR模型, 即对 $C, g, \varepsilon$ 中一个参数进行优化, 第3类为利用ABC算法对SVR3个关键参数优化的SVR模型, 即对 $C, g, \varepsilon$ 3个参数同时进行优化。3类模型共建立5个模型, 5个模型采用相同的训练集和测试集, 根据参数敏感性分析结果(5.2节), 各参数的寻优范围设置见表2。

表2 实验参数设置

Table 2 Parameter setting of experiment

模型名称	优化参数	$C$ 取值范围	$g$ 取值范围	$\varepsilon$ 取值范围	$NP$	$Limit$	迭代次数
SVR	—	—	—	—	—	—	—
ABC-SVR-C	$C$	[0.01, 100]	—	—	30	20	100
ABC-SVR-g	$g$	—	[0.01, 10]	—	30	20	100
ABC-SVR- $\varepsilon$	$\varepsilon$	—	—	[0.0001, 0.05]	30	20	100
ABS-SVR	$C, g, \varepsilon$	[0.01, 100]	[0.01, 10]	[0.0001, 0.05]	30	20	100

## 4 结果

### 4.1 优化单个SVR参数的LAI反演模型

根据表2中的实验参数设置,使用人工蜂群算法优化SVR单个参数,优化后各参数取值如表3,没有进行优化的参数取默认值( $C=1$ ,  $g=1/k$ ,  $\varepsilon=0.1$ ) (Chang 和 Lin, 2011), ABC算法迭代结果分别如图6(a)(b)(c)所示。优化参数 $C$ 、 $g$ 算法迭代次数均小于20次时, SVR算法对训练样本集采用10折交叉验证后的平均均方误差(PMSE)已经趋于一个较小的稳定值,优化参数 $\varepsilon$ 算法迭代次数在介于30—40次PMSE也同样稳定在一个比较小的值。以优化后的参数分别建立研究区域LAI反演的第1类和第2类模型,验证结果如图5、7所示。

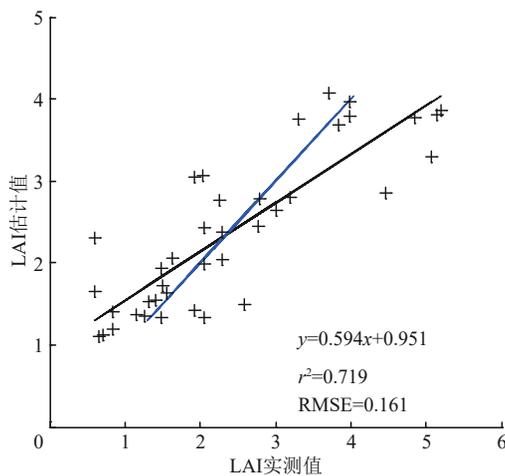
第1类未优化参数的SVR模型验证结果如图5(a)(b)所示,第2类优化SVR参数惩罚系

数 $C$ 的LAI反演模型ABC-SVR- $C$ 验证结果如图7(a)(b)所示。SVR模型验证结果表明,训练集和测试集的回归直线斜率(黑色)分别为0.594和0.630,决定系数分别为0.719和0.727, RMSE为0.161和0.137。将ABC-SVR- $C$ 模型验证结果图7(a)(b)与图5(a)(b)中未优化参数的SVR模型反演结果对比分析可得,训练集和测试集的回归直线斜率分别提高了0.087和0.038, RMSE下降到0.149和0.135,特别显著的是训练集验证结果中决定系数由0.719提高到0.745。

表3 参数优化结果

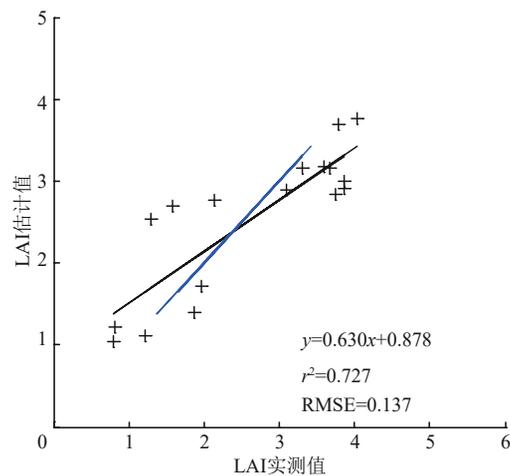
Table 3 The result of optimizing parameters

模型	参数		
	$C$	$g$	$\varepsilon$
ABC-SVR- $C$	4.324	0.111(默认)	0.1(默认)
ABC-SVR- $g$	1(默认)	0.457	0.1(默认)
ABC-SVR- $\varepsilon$	1(默认)	0.111(默认)	0.037
ABC-SVR	45.0904	0.01	0.00014



(a) SVR模型训练集的验证结果

(a) Validation result of SVR model with training set



(b) SVR模型测试集的验证结果

(b) Validation result of SVR model with test set

—— 回归直线

——  $k=1$

图5 未优化参数的SVR模型验证结果

Fig. 5 Validation results of SVR model without optimized parameters

优化SVR核函数宽度参数 $g$ 的ABC-SVR- $g$ 模型验证结果如图7(c)(d)所示,与未优化参数的SVR模型验证结果图5(a)(b)对比分析,结果表明,优化了核函数宽度参数 $g$ 的LAI反演模型,其训练集和测试集的回归直线斜率分别提高了0.102和0.029,决定系数也有较大的提高,达

到了0.765和0.735, RMSE下降到0.144和0.132。

优化SVR损失函数参数 $\varepsilon$ 的ABC-SVR- $\varepsilon$ 模型验证结果如图7(e)(f)所示,与未优化参数的SVR模型验证结果图5(a)(b)进行对比分析可得,ABC-SVR- $\varepsilon$ 模型预测值与实测值更为接近,回归直线更接近于1,训练集回归斜率由0.594提

高到0.646, 决定系数和RMSE没有显著变化。但是在测试集验证结果, 回归斜率0.630提高到

0.728, 决定系数由0.727提高到了0.757, RMSE下降到0.126。

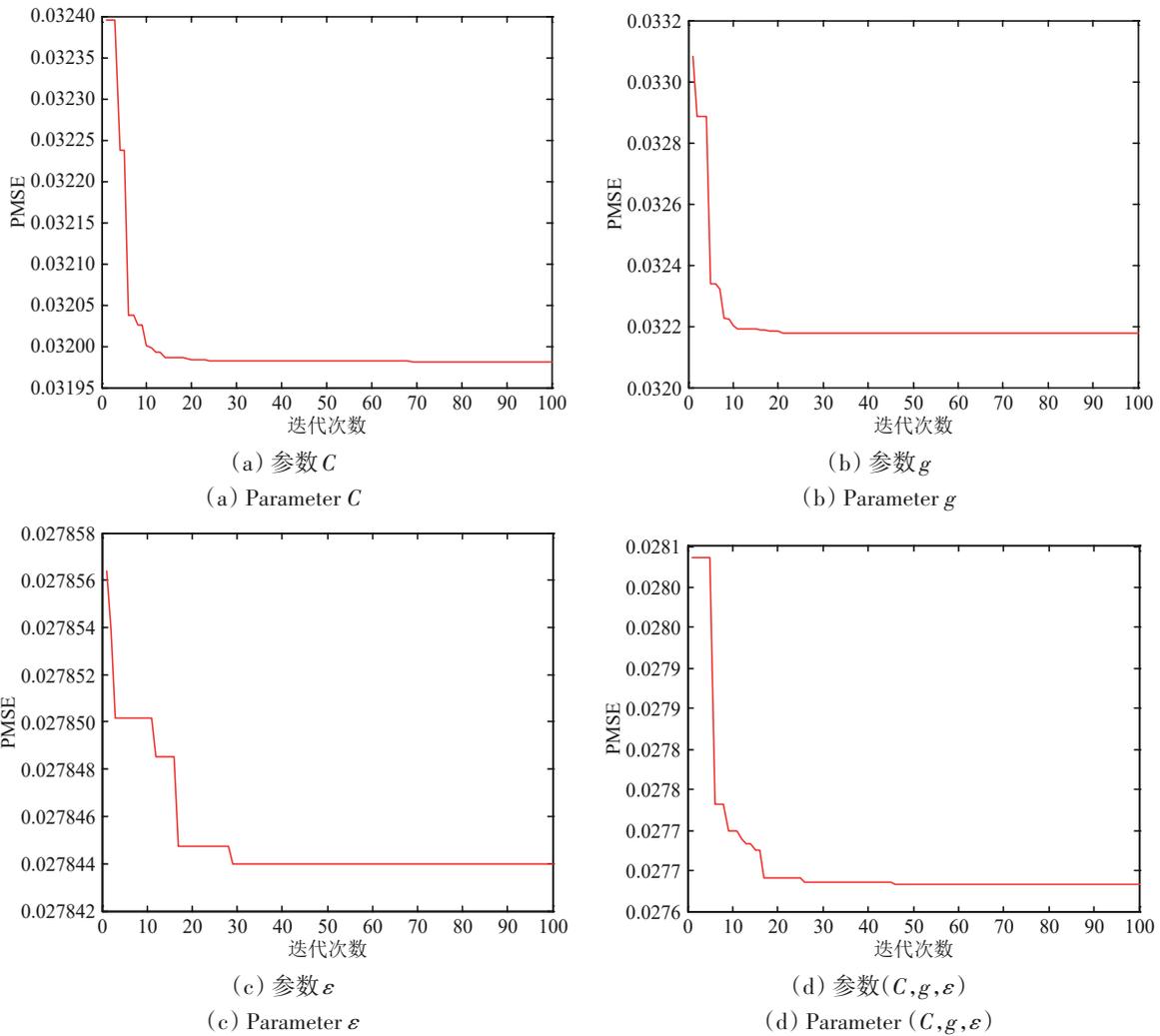
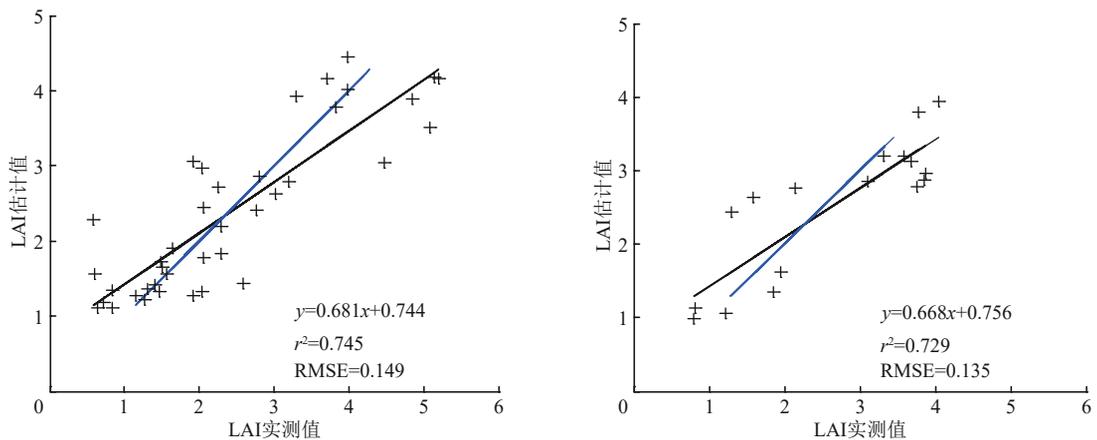


图6 ABC算法优化SVR参数迭代结果

Fig. 6 The iteration result of optimizing parameters of SVR by ABC algorithm



(a) Validation result of ABC-SVR-C model with training set

(b) Validation result of ABC-SVR-C model with test set

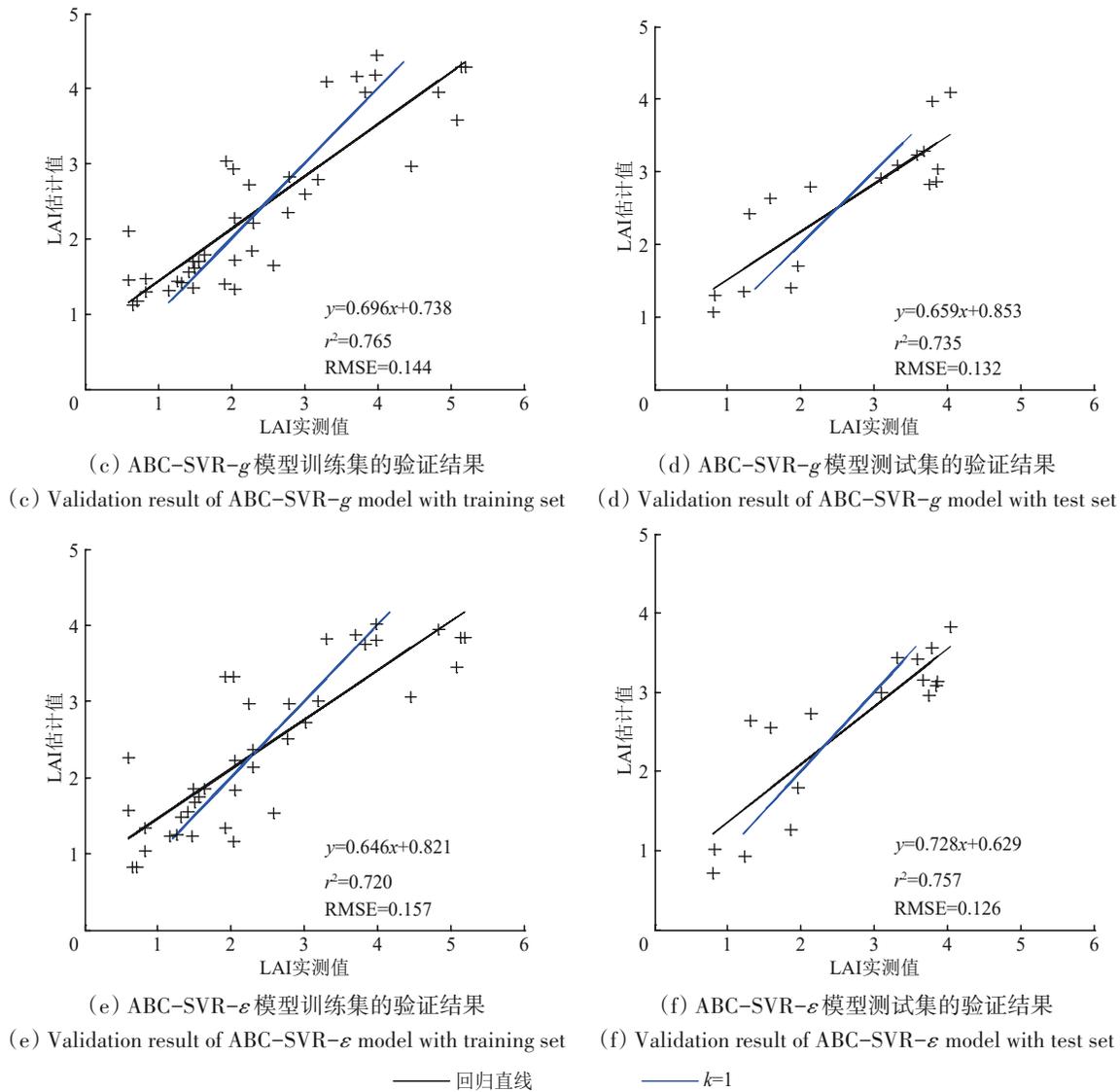


图7 优化单个参数的SVR模型验证结果

Fig. 7 Validation results of SVR model with optimized single parameters

最后, 选择更能反映模型性能和回归拟合能力的测试集, 对比分析第2类模型 (ABC-SVR- $C$ 、ABC-SVR- $g$ 、ABC-SVR- $\varepsilon$ ) 的精度, 即对比分析图7 (b) (d) (f)。实验结果表明, 在这3个模型中ABC-SVR- $\varepsilon$ 模型精度最佳, 其测试集回归直线斜率最高 ( $k=0.728$ ), 回归直线偏移最小 ( $b=0.629$ ), 决定系数最高 ( $r^2=0.757$ ), 均方根误差最小 ( $RMSE=0.126$ )。

#### 4.2 优化3个SVR参数的LAI反演模型

由4.1节可知, 优化SVR算法惩罚系数 $C$ 、核函数宽度参数 $g$ 、不敏感损失函数 $\varepsilon$ 等3个参数之一, 均可以提高LAI反演模型精度, 因此, 根据表2中的参数设置, 利用ABC算法同时优化这3个参数,

各参数优化后的取值如表3, ABC算法迭代结果如图7 (d), 迭代次数为40—50次时趋于稳定。以表3中3个参数 ( $C$ ,  $g$ ,  $\varepsilon$ ) 优化后的取值建立ABC-SVR模型, 并对该模型进行验证, 结果如图8。

第3类优化SVR3个参数的模型验证结果 (图8) 与第1类未优化参数的SVR模型验证结果 (图5) 对比分析可知, 训练集回归直线斜率达到0.713, 相比SVR模型 (0.594) 提高了0.119, 而决定系数为0.733, 比SVR模型0.719提高了0.014, RMSE由0.161下降到0.151。从测试集的验证结果上看, ABC-SVR模型反演精度提高的更为显著, 回归直线斜率达到0.797, 更接近于1, 相比SVR模型

(0.630) 提高了0.167, 而决定系数更是有了明显的提高, 达到了0.775, RMSE相比SVR模型 (0.137),

也降低到了0.122。

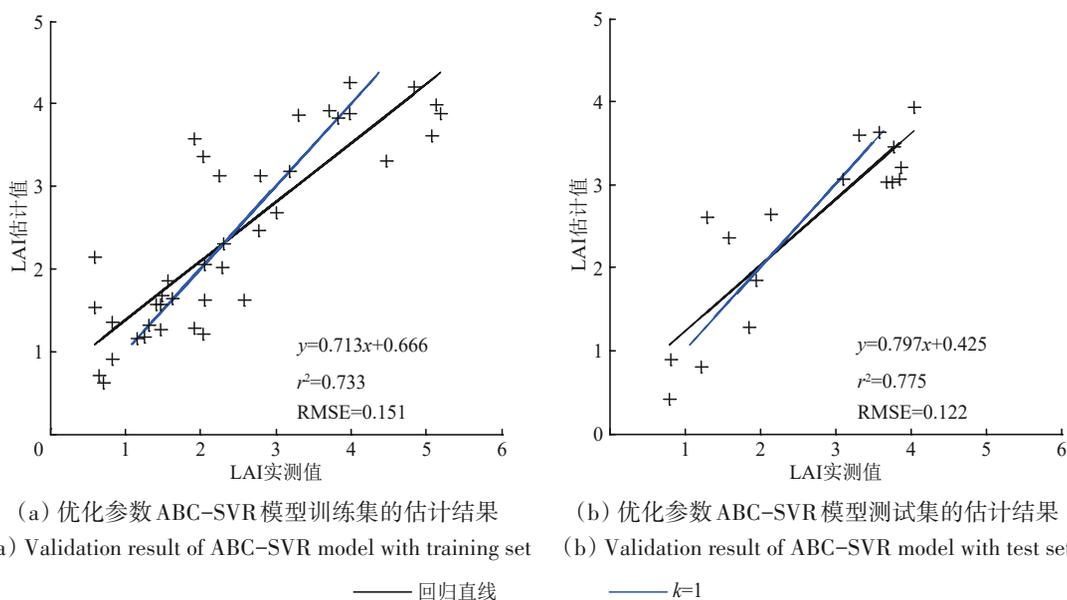


图8 优化3个参数的SVR模型验证结果

Fig. 8 Validation results of SVR model with optimized three parameters

同样以测试集验证结果 (图8 (b)、图6 (b) (d) (f)) 对比 ABC-SVR 模型与 ABC-SVR- $C$ 、ABC-SVR- $g$ 、ABC-SVR- $\varepsilon$  模型精度, 结果表明, 在这4个LAI反演模型中ABC-SVR模型的精度最好, 其测试集回归直线斜率最高 ( $k=0.797$ ), 回归直线偏移最小 ( $b=0.425$ ), 决定系数最高 ( $r^2=0.775$ ), 均方根误差最小 (RMSE=0.122)。

## 5 讨论

### 5.1 3类反演模型精度比较

根据4.1节结果分析可得ABC算法优化SVR单个关键参数建立的第2类模型ABC-SVR- $C$ 、ABC-SVR- $g$ 、ABC-SVR- $\varepsilon$  均比第1类未优化参数的SVR模型精度高, 且相较于优化 $C$ 、 $g$ 两个参数的模型, 优化 $\varepsilon$ 参数的模型精度最佳, 回归直线斜率达到0.728、决定系数达到0.757, 是第2类模型中精度最高的模型, 这证明参数 $\varepsilon$ 是SVR中一个比较重要的参数, 而大部分优化SVR参数的研究中只优化了参数 $C$ 、 $g$ , 没有对参数 $\varepsilon$ 进行太多研究, 本研究证明了参数 $\varepsilon$ 对SVR回归性能具有较大的影响。因此, 与SVR的有关研究中对参数 $\varepsilon$ 的优化是非常有必要的。

第4.2节结果分析可得优化SVR的3个参数的

第3类模型 (ABC-SVR模型) 相比第1类的SVR模型 ( $k=0.630$ ,  $r^2=0.727$ ) 和第2类的3个模型ABC-SVR- $C$  ( $k=0.668$ ,  $r^2=0.729$ ), ABC-SVR- $g$  ( $k=0.659$ ,  $r^2=0.735$ ), ABC-SVR- $\varepsilon$  ( $k=0.728$ ,  $r^2=0.757$ ), ABC-SVR模型精度最高, 模型预测值与LAI地面实测值具有很高的相关性, 是LAI反演的优选模型。这说明同时优化3个参数能更显著地提高SVR模型的精度。ABC-SVR模型精度最高, 回归直线斜率达到0.797、决定系数达到0.775。

### 5.2 参数敏感性分析

对第4节结果进行综合分析可得, 参数 $C$ 、 $g$ 、 $\varepsilon$ 的取值均对模型预测精度有一定的影响, 本节旨在选定各参数的寻优范围以及分析各参数对模型精度的影响。基于第4节的训练数据和测试数据, 参考杨攻等 (2009)、王云飞和郑晓娟 (2010) 对SVR参数敏感性分析方法, 分别探讨SVR模型精度 ( $r^2$ 和RMSE) 对每个参数的敏感性, 当分析某个参数的敏感性时, 其余两个参数均取默认值 (表3)。

图9 (a) (b) 为惩罚系数 $C$ 对 $r^2$ 的影响, 其中图9 (b) 局部放大图9 (a), 范围0—100; 图9 (c) (d) 为惩罚系数 $C$ 对RMSE的影响, 其中图9 (d) 局部放大图9 (c), 范围0—100。综合图9 (a)

(b) (c) (d) 对惩罚系数  $C$  进行敏感性分析可知, 惩罚系数  $C$  对模型预测精度较大的影响。图 9 (a) 显示  $C$  值在 (0, 800) 的范围内决定系数  $r^2$  波动较大,  $C$  值大于 800 以后维持在最低 0.64—0.65, 此时模型预测值与实际值相关程度较低。图 9 (b) 局部分析可得,  $C$  值在 (0, 100) 内是决定系数波动最大的区间, 但决定系数的最大值, 即模型预测值与实际值的相关程度最高出现在这个区间,  $C$  值在 (100—800) 区间, 决定系数呈现下降趋

势。由图 9 (c) 分析 RMSE 变化可知,  $C$  值在 (0, 200) 内, RMSE 逐渐趋于一个低稳定值, 此时模型误差较低, 大于 200 时 RMSE 也逐渐增大, 模型误差逐渐增大; 从局部变化图 9 (d) 可知, 但  $C$  值在 70 左右时, RMSE 趋于一个很稳定的低值, 此时模型误差较低。因此综合分析两个模型精度指标的变化, 在实验的过程中建议设置参数  $C$  的寻优范围为 (0, 100)。

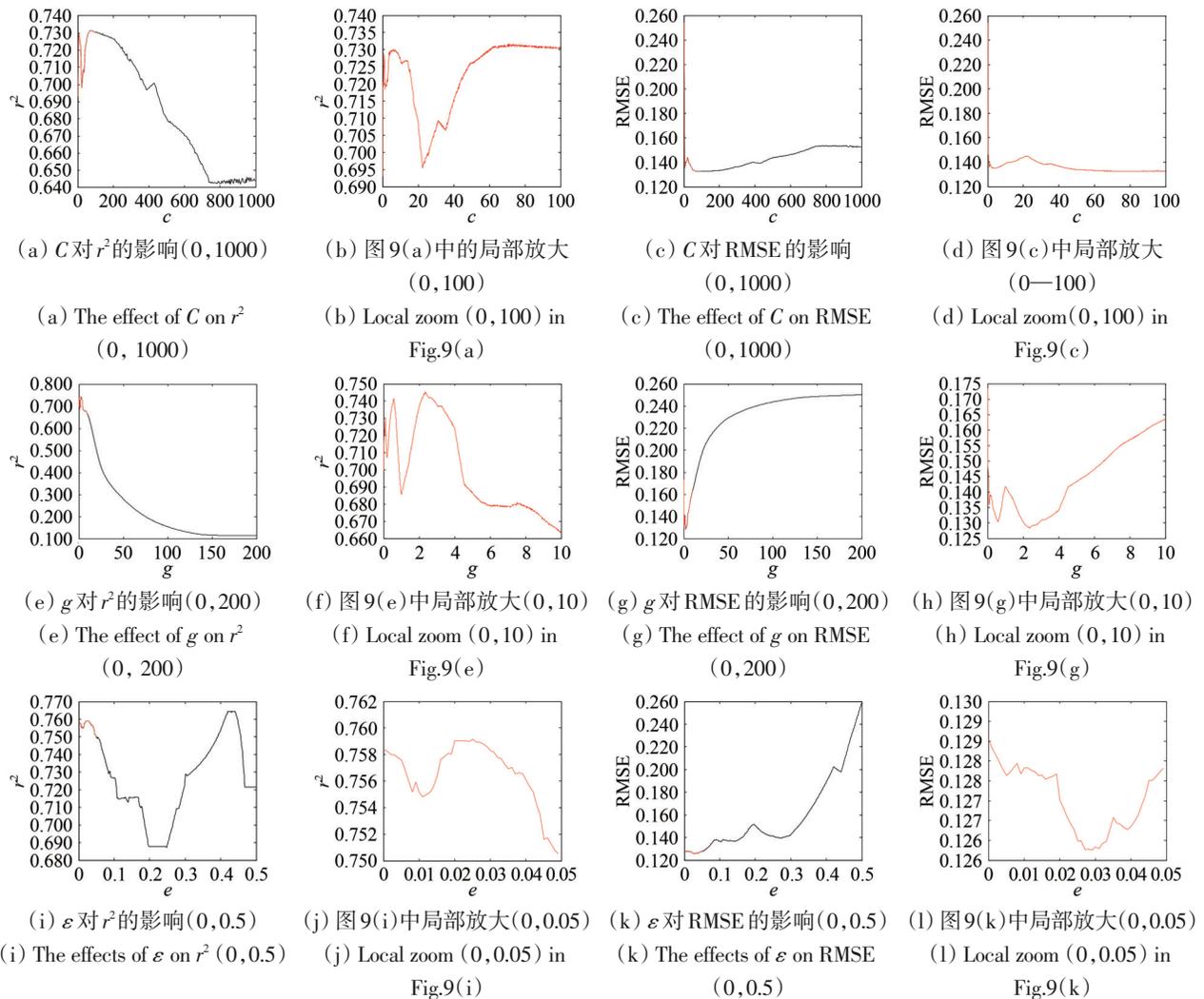


图 9 SVR 关键参数对模型精度的影响

Fig. 9 The influence of SVR key parameters on model accuracy

图 9 (e) (f) 为核函数宽度参数  $g$  对  $r^2$  的影响, 其中图 9 (f) 局部放大图 9 (e), 范围 0—10; 图 9 (g) (h) 为核函数宽度参数  $g$  对 RMSE 的影响, 其中图 9 (h) 局部放大图 9 (g), 范围 0—10。由图 9 (e) (f) (g) (h) 对核函数宽度参数  $g$  进行敏感性分析可知, 该参数取值对模型精度的影响很大。由图 9 (e) (f) 可

知, 决定系数在 (0, 10) 区间出现决定系数最大值 (约 0.75), 在 (10, 140) 区间内决定系数整体上持续下降, 大于 140 以后决定系数接近 0.1, 此时预测值和实际值之间几乎没有相关性, SVR 基本丧失预测能力, 出现明显的过学习现象, 此时模型没有推广能力, 可见参数  $g$  的选择对 SVR 模

型的预测能力有很大的影响。由图9 (g) (h) 中可知,当参数 $g$ 在(0, 10)区间内RMSE出现最低值,即模型的最小预测误差,在(10, 140)区间内RMSE持续上升,模型预测误差逐渐增大,大于140以后稳定在0.245左右,此时模型误差较大。因此,经过实验可以基本确定核函数宽度参数 $g$ 的理想寻优范围应该设置为(0, 10)。

图9 (i) (j) 为参数 $\varepsilon$ 对 $r^2$ 的影响,其中图9 (j) 局部放大图9 (i),范围0—0.05;图9 (k) (l) 为参数 $\varepsilon$ 对RMSE的影响,其中图9 (l) 局部放大图9 (k),范围0—0.05。从图9 (i) (j) (k) (l) 从总体上分析可得,随着参数 $\varepsilon$ 增大,表示模型预测精度的两个指标值变化均比较大,说明该参数的取值对模型的预测精度具有很大影响,局部微小的变化会引起模型精度的波动,因此对该参数的优化显得尤为重要。由图9 (i) (j) 分析可得,在(0, 0.5)区间内,决定系数波动很大。由图9 (k) (l) 分析可知,随着参数 $\varepsilon$ 的增大, RMSE也逐步增大,模型误差逐步增大,模型最低预测误差出现在(0, 0.05)区间内,因此,综合两个精度指标的敏感性可以合理设置参数 $\varepsilon$ 的取值范围在(0, 0.05)

区间内。

结合前述实验结果和分析,可以看出3个参数对SVR模型的精度( $r^2$ , RMSE)影响都比较大。相对于参数 $C$ (0, 1000)和参数 $g$ (0, 200)在较大的区间内变化时模型精度在整体上呈现单调的趋势,参数 $\varepsilon$ 局部微小的变化会引起模型精度的不确定性,因此,优化参数 $\varepsilon$ 更能降低模型不确定性。

### 5.3 ABC-SVR模型精度改善的显著性检验

从4.2节分析得知,ABC-SVR模型是整个实验过程中的最佳模型( $k=0.797$ ,  $r^2=0.775$ , RMSE=0.122),相比未优化参数的SVR模型精度( $k=0.630$ ,  $r^2=0.727$ , RMSE=0.137)精度有了一定的改善,但是为了避免单次实验的偶然性,以及研究ABC算法优化SVR参数模型精度提升的显著性,本小节将56个样本分别重新随机划分为20组不同的训练数据与测试数据的组合(70%训练集,30%测试集),基于这20个不同的训练集分别使用优化3个参数的ABC-SVR模型和未优化参数的SVR建立共40个模型,并使用相应测试集进行验证,所有模型精度如图10所示。

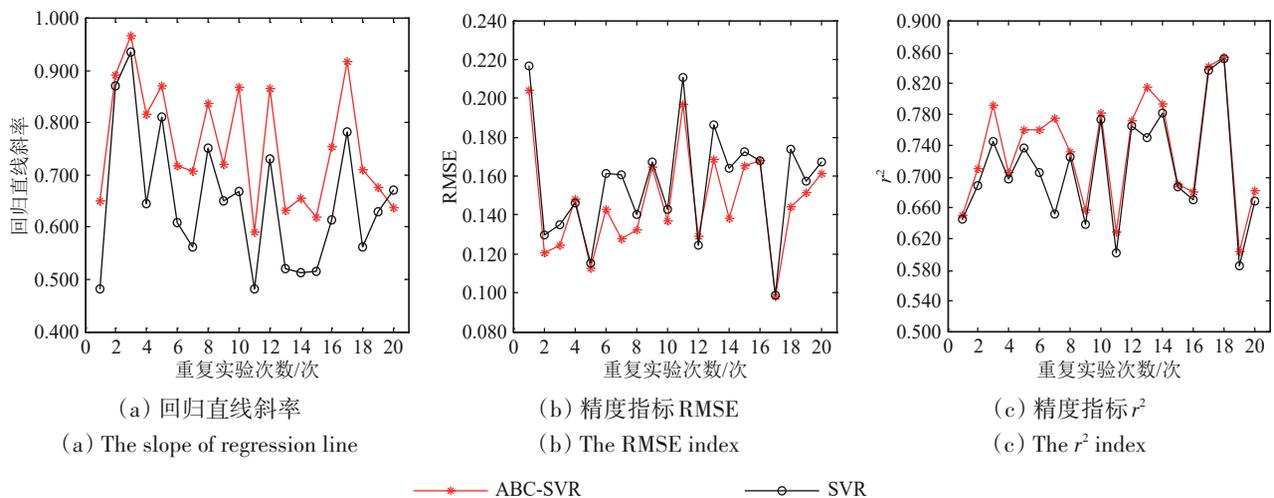


图10 20次重复实验的回归模型精度

Fig. 10 The accuracy of regression model of 20 repeat experiments

由图10可以分析得出,在20次重复实验中ABC-SVR模型预测值与实测值回归直线斜率95%都高于SVR模型, RMSE误差均小于SVR模型,决定系数均高于SVR模型,说明ABC算法能改善SVR模型的精度。

为了进一步验证ABC-SVR模型的精度的是否

显著提升,本文基于上述20次重复实验的结果,利用spss软件计算了ABC-SVR模型与SVR模型之间精度差异的显著性(回归直线斜率 $k$ 、RMSE、 $r^2$ ),结果如表4。在95%的置信区间下,分析回归直线斜率差异显著性可得,ABC-SVR模型的回归直线斜率( $0.756\pm 0.114$ )比SVR模型( $0.651\pm 0.130$ )

提升了0.089, 该差异具有统计学意义,  $t(19) = 0.813$ ,  $P < 0.001$ 。同理分析决定系数和RMSE可得, ABC-SVR模型的决定系数(0.734±0.071)比SVR模型(0.711±0.071)提升了0.023,  $t(19) = 3.599$ ,  $P < 0.005$ ; ABC-SVR模型的RMSE(0.146±0.026)比SVR模型(0.156±0.029)下降了0.013,  $t(19) = -4.417$ ,  $P < 0.001$ ; 最终分析结果说明两个精度指标的差异均具有统计学意义, 两者之间差异显著, ABC算法显著改善了SVR模型的精度。

表4 精度差异的显著性检验

模型精度	均值	标准差	$t$	$p$
ABC-SVR- $k$	0.756	0.114	0.813	0.000
SVR- $k$	0.651	0.130		
ABC-SVR- $r^2$	0.734	0.071	3.599	0.002
SVR- $r^2$	0.711	0.071		
ABC-SVR-RMSE	0.146	0.026	-4.417	0.000
SVR-RMSE	0.156	0.029		

## 6 结论

本文基于美国土壤水分实验SMEX02的LAI野外实测数据和Landsat 7 ETM+影像, 采用人工蜂群算法(ABC)优化SVR参数( $C$ ,  $g$ ,  $\varepsilon$ ), 建立3类模型: 未优化参数的SVR模型, 优化单个参数的ABC-SVR- $C$ 、ABC-SVR- $g$ 、ABC-SVR- $\varepsilon$ 模型以及优化3个参数的模型ABC-SVR, 对3类模型进行验证, 并分析SVR的3个关键参数对LAI反演模型精度的影响, 验证ABC-SVR模型精度改善的显著性。得出结论如下, (1) ABC算法优化SVR参数能明显提高LAI反演模型的精度, 且同时优化SVR3个参数能更显著地提高LAI反演模型的精度; (2) 相比参数 $C$ 和 $g$ , 参数 $\varepsilon$ 局部微小的变化会增加模型精度的不确定性, 因此在应用中优化参数 $\varepsilon$ 是十分必要的; (3) ABC-SVR模型与SVR模型的精度差异均具有统计学意义, 说明ABC算法显著改善了SVR模型的精度。

但研究实验还存在几点不足, 首先, 反演时需要样本量较大的训练集(夏天等, 2013; 杨敏等, 2015), 由于本研究区域LAI实测数据数量限制, 因此本研究将间隔一周的两组数据合并构建训练集和测试集, 因此由于样本的获取时间不一致, 这可能给模型带来误差; 其次, 本研究输入

9类特征变量(Landsat 7 ETM+6个波段、NDVI、RVI、EVI)进行实验, 没有考虑数据冗余是否会限制模型的精度(Forsati等, 2012; Ghamisi等, 2015); 最后, 本文中ABC算法迭代次数统一设置为100次, 所有迭代结果(图6)显示100次迭代后最优适应度(MSE)都能趋于稳定, 在后续研究中可以考虑是否可以设置合适的迭代次数以提高算法效率。对于上述不足, 后续研究中一方面可以选择适合的研究区域及相应的LAI地面实测数据保证输入数据的准确性, 另一方面可以对输入的多个特征变量进行特征选择, 去除冗余信息, 并根据不同情况设置不同的ABC算法迭代次数, 以建立更高效、更精确的LAI反演模型。

## 参考文献(References)

- Chang C C and Lin C J. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3): 1-27 [DOI: 10.1145/1961189.1961199]
- Chen J M and Black T A. 1992. Defining leaf area index for non-flat leaves. *Plant, Cell and Environment*, 15(4): 421-429 [DOI: 10.1111/j.1365-3040.1992.tb00992.x]
- Cheng P and Wang X L. 2011. Influence of SVR parameter on non-linear function approximation. *Computer Engineering*, 37(3): 189-191, 194 (成鹏, 汪西莉. 2011. SVR参数对非线性函数拟合的影响. *计算机工程*, 37(3): 189-191, 194) [DOI: 10.3969/j.issn.1000-3428.2011.03.067]
- Cherkassky V. 1997. The nature of statistical learning theory. *IEEE Transactions on Neural Networks*, 8(6): 1564 [DOI: 10.1109/TNN.1997.641482]
- Dorigo M and Gambardella L M. 1997. Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1): 53-66 [DOI: 10.1109/4235.585892]
- Durbha S S, King R L and Younan N H. 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, 107(1/2): 348-361 [DOI: 10.1016/j.rse.2006.09.031]
- Fang X Q and Zhang W C. 2003. The application of remotely sensed data to the estimation of the leaf area index. *Remote Sensing for Land and Resources*, 15(3): 58-62 (方秀琴, 张文昌. 2003. 叶面积指数(LAI)的遥感定量方法综述. *国土资源遥感*, 15(3): 58-62) [DOI: 10.3969/j.issn.1001-070X.2003.03.014]
- Forsati R, Moayedikia A, Keikha A and Shamsfard M. 2012. A novel approach for feature selection based on the bee colony optimization. *International Journal of Computer Applications*, 43(8): 13-16 [DOI: 10.5120/6122-8329]
- Ghamisi P, Couceiro M S and Benediktsson J A. 2015. A novel feature selection approach based on FODPSO and SVM. *IEEE Transac-*

- tions on Geoscience and Remote Sensing, 53(5): 2935-2947 [DOI: 10.1109/TGRS.2014.2367010]
- Guo L, Pei Z Y, Zhang S L, Sun J Y, Liang Z L and Teng D J. 2010. Estimation method of sugarcane leaf area index using HJ CCD images. Transactions of the CSAE, 26(10): 201-205 (郭琳, 裴志远, 张松龄, 孙娟英, 梁自力, 滕冬建. 2010. 基于环境星 CCD 图像的甘蔗叶面积指数反演方法. 农业工程学报, 26(10): 201-205) [DOI: 10.3969/j.issn.1002-6819.2010.10.034]
- He Y Z, Zhang Z Q and Guan C Y. 2015. Application and prospect of hyperspectral remote sensing technology in precision agriculture monitoring. Crop Research, 29(1): 96-100 (何友铸, 张振乾, 官春云. 2015. 高光谱遥感技术在精细农业监测上的应用及展望. 作物研究, 29(1): 96-100) [DOI: 10.3969/j.issn.1001-5280.2015.01.23]
- Holland J H. 1975. Adaption in Natural and Artificial Systems. Ann Arbor: University of Michigan Press
- Karaboga D. 2005. An Idea Based on Honey Bee Swarm for Numerical Optimization. Erciyes University
- Karaboga D and Akay B. 2009. A comparative study of artificial bee colony algorithm. Applied Mathematics and Computation, 214(1): 108-132 [DOI: 10.1016/j.amc.2009.03.090]
- Karaboga D and Basturk B. 2007. A powerful and efficient algorithm for numerical function optimization: Artificial Bee Colony (ABC) algorithm. Journal of Global Optimization, 39(3): 459-471 [DOI: 10.1007/s10898-007-9149-x]
- Karaboga D, Gorkemli B, Ozturk C and Karaboga N. 2014. A comprehensive survey: Artificial Bee Colony (ABC) algorithm and applications. Artificial Intelligence Review, 42(1): 21-57 [DOI: 10.1007/s10462-012-9328-0]
- Kennedy J and Eberhart R. 1995. Particle swarm optimization//Proceedings of 1995 IEEE International Conference on Neural Networks. Perth, WA, Australia: IEEE: 1942-1948 [DOI: 10.1109/ICNN.1995.488968]
- Kwok J T Y. 1998. Support vector mixture for classification and regression problems//Proceedings of the 14th International Conference on Pattern Recognition. Brisbane, Queensland: IEEE [DOI: 10.1109/ICPR.1998.711129]
- Latifi H and Galos B. 2010. Remote sensing-supported vegetation parameters for regional climate models: a brief review. iForest-Biogeosciences and Forestry, 3(4): 98-101 [DOI: 10.3832/ifor0543-003]
- Li N, Zhu X F, Pan Y Z and Zhan P. 2018. Optimized SVM based on artificial bee colony algorithm for remote sensing image classification. Journal of Remote Sensing, 22(4): 559-569 (李楠, 朱秀芳, 潘耀忠, 詹培. 2018. 人工蜂群算法优化的 SVM 遥感影像分类. 遥感学报, 22(4): 559-569) [DOI: 10.11834/jrs.20187176]
- Liang D, Guan Q S, Huang W J, Huang L S and Yang G J. 2013. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat. Transactions of the Chinese Society of Agricultural Engineering, 29(7): 117-123 (梁栋, 管青松, 黄文江, 黄林生, 杨贵军. 2013. 基于支持向量机回归的冬小麦叶面积指数遥感反演. 农业工程学报, 29(7): 117-123) [DOI: 10.3969/j.issn.1002-6819.2013.07.015]
- Liang S L, Cheng J, Jia K, Jiang B, Liu Q, Liu S H, Xiao Z Q, Xie X H, Yao Y J, Yuan W P, Zhang X T and Zhao X. 2016. Recent progress in land surface quantitative remote sensing. Journal of Remote Sensing, 20(5): 875-898 (梁顺林, 程洁, 贾坤, 江波, 刘强, 刘素红, 肖志强, 谢先红, 姚云军, 袁文平, 张晓通, 赵祥. 2016. 陆表定量遥感反演方法的发展新动态. 遥感学报, 20(5): 875-898) [DOI: 10.11834/jrs.20166258]
- Liu X C, Fan W J, Tian Q J and Xu X R. 2008. Comparative analysis among different methods of leaf area index inversion. Acta Scientiarum Naturalium Universitatis Pekinensis, 44(2): 827-834 (刘晓臣, 范闻捷, 田庆久, 徐希孺. 2008. 不同叶面积指数反演方法比较研究. 北京大学学报(自然科学版), 44(2): 827-834) [DOI: 10.3321/j.issn:0479-8023.2008.05.025]
- Liu Y, Liu R G, Chen J M, Cheng X and Zheng G. 2013. Current Status and Perspectives of Leaf Area Index Retrieval from Optical Remote Sensing. Journal of Geo-information Science, 15(05): 734-743 (刘洋, 刘荣高, 陈镜明, 程晓, 郑光. 2013. 叶面积指数遥感反演研究进展与展望. 地球信息科学学报, 15(05): 734-743) [DOI: 10.3724/SP.J.1047.2013.00734]
- Steinbrunn M, Moerkotte G and Kemper A. 1997. Heuristic and randomized optimization for the join ordering problem. The VLDB Journal, 6(3): 191-208 [DOI: 10.1007/s007780050040]
- Tuia D, Verrelst J, Alonso L, Perez-Cruz F and Camps-Valls G. 2011. Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geoscience and Remote Sensing Letters, 8(4): 804-808 [DOI: 10.1109/LGRS.2011.2109934]
- Wang D C, Fang T J, Tang Y and Ma Y J. 2003. Review of support vector machines regression theory and control. Pattern Recognition and Artificial Intelligence, 16(2): 192-197 (王定成, 方廷健, 唐毅, 马永军. 2003. 支持向量机回归理论与控制的综述. 模式识别与人工智能, 16(2): 192-197) [DOI: 10.3969/j.issn.1003-6059.2003.02.012]
- Wang L, Zhang Y, Peng W H, Xu B and Wang Q C. 2014. SVR approach based on artificial bee colony optimization. Systems Engineering and Electronics, 36(2): 326-330 (王琳, 张赞, 彭文辉, 徐波, 王前程. 2014. 基于人工蜂群优化的支持向量回归预测方法. 系统工程与电子技术, 36(2): 326-330) [DOI: 10.3969/j.issn.1001-506X.2014.02.20]
- Wu X X, Xie Q Y. 2014. Research progress on inversion methods of leaf area index use remote sensing. China Agriculture Information, (07): 285-286 (武旭霞, 谢巧云. 2014. 植被叶面积指数遥感反演方法研究进展. 中国农业信息, (07): 285-286) [DOI: 10.3969/j.issn.1672-0423.2014.04.230]
- Wang Y F and Zheng X J. 2010. Sensitivity analysis of model parameters and v-SVR model of slope deformation due to excavating from open-pit to underground mining. Chinese Journal of Rock Mechanics and Engineering, 29(S1): 2902-2907 (王云飞, 郑晓娟. 2010. 露天转地下边坡变形 v-SVR 模型及参数敏感性分析. 岩石力学与工程学报, 29(S1): 2902-2907)
- Xia T, Wu W B, Zhou Q B and Zhou Y. 2013. Comparison of two inversion methods for winter wheat leaf area index based on hyperspectral remote sensing. Transactions of the Chinese Society of Agricultural Engineering, 29(3): 139-147 (夏天, 吴文斌, 周清波, 周勇. 2013. 冬小麦叶面积指数高光谱遥感反演方法对比. 农业

- 工程学报, 29(3): 139-147) [DOI: 10.3969/j.issn.1002-6819.2013.03.019]
- Xing Z R, Feng Y G, Li W M, Wang P and Yang G J. 2010. The research status of inversion of leaf area index with hyperspectral remote sensing. *Science of Surveying and Mapping*, 35(S1): 162-164, 62 (邢著荣, 冯幼贵, 李万明, 王萍, 杨贵军. 2010. 高光谱遥感叶面积指数(LAI)反演研究现状. *测绘科学*, 35(S1): 162-164, 62)
- Yan G H and Zhu Y S. 2009. Parameters selection method for support vector machine regression. *Computer Engineering*, 35(14): 218-220 (闫国华, 朱永生. 2009. 支持向量机回归的参数选择方法. *计算机工程*, 35(14): 218-220) [DOI: 10.3969/j.issn.1000-3428.2009.14.076]
- Yang M, Lin J, Gu Z Y, Tong G C, Wong Y B, Zhang J C and Lu X Z. 2015. Leaf area index retrieval based on Landsat 8 OLI multi-spectral image data and BP neural network. *Science of Soil and Water Conservation*, 13(4): 86-93 (杨敏, 林杰, 顾哲衍, 佟光臣, 翁永兵, 张金池, 鲁小珍. 2015. 基于 Landsat 8 OLI 多光谱影像数据和 BP 神经网络的叶面积指数反演. *中国水土保持科学*, 13(4): 86-93) [DOI: 10.3969/j.issn.1672-3007.2015.04.013]
- Yang M, Liu Y and Kong B. 2009. Research on parameters selection method of SVR model. *Computer ERA*, (11): 53-55 (杨玫, 刘瑜, 孔波. 2009. SVR 模型参数选择方法的研究. *计算机时代*, (11): 53-55) [DOI: 10.3969/j.issn.1006-8228.2009.11.020]
- Yu M and Ai Y Q. 2012. SVM parameter optimization and application based on artificial bee colony algorithm. *Journal of Optoelectronics·Laser*, 23(2): 374-378 (于明, 艾月乔. 2012. 基于人工蜂群算法的支持向量机参数优化及应用. *光电子·激光*, 23(2): 374-378)
- Zai S M, Wen J, Guo D D, Han Q B, Deng Z, Sun H and Zhao D B. 2011. Determination of leaf area of sweet pepper based on support vector machine model and image processing. *Transactions of the Chinese Society of Agricultural Engineering*, 27(3): 237-241 (宰松梅, 温季, 郭冬冬, 韩启彪, 邓忠, 孙浩, 赵东彬. 2011. 基于支持向量机模型和图像处理技术的甜椒叶面积测定. *农业工程学报*, 27(3): 237-241) [DOI: 10.3969/j.issn.1002-6819.2011.03.045]
- Zang S Y, Zhang C, Zhang L J and Zhang Y H. 2012. Wetland remote sensing classification using support vector machine optimized with genetic algorithm: a case study in Honghe nature national reserve. *Scientia Geographica Sinica*, 32(4): 434-44 (臧淑英, 张策, 张丽娟, 张玉红. 2012. 遗传算法优化的支持向量机湿地遥感分类——以洪河国家级自然保护区为例. *地理科学*, 32(4): 434-441) [DOI: 10.13249/j.cnki.sgs.2012.04.434]

## Optimized SVR based on artificial bee colony algorithm for leaf area index inversion

ZHOU Xiaoxue<sup>1,3</sup>, LI Nan<sup>1,3</sup>, PAN Yaozhong<sup>1,2,3</sup>, SUN Lixin<sup>3</sup>

1.State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Beijing Normal University and Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100875, China;

2.School of Geographical Sciences, Qinghai Normal University, Xining 810016, China;

3.Institute of Remote Sensing Science and Engineering, Faculty of Geographical Sciences, Beijing Normal University, Beijing 100875, China

**Abstract:** Support Vector Regression (SVR) method as a new idea in LAI inversion has certain application value and prospect. However, the value of penalty coefficient  $C$ , width parameter  $g$  of kernel function and insensitive loss function parameter  $\epsilon$  in the SVR algorithm have a significant impact on regression accuracy. This paper proposed a method for Leaf Area Index (LAI) inversion using remote sensing images based on ABC (Artificial Bee Colony) algorithm to optimize SVR parameters. In addition, the LAI measurement values were from the Soil Moisture Experiment 2002 in US (SMEX02) and Landsat 7 ETM + surface reflectance data at the same time. In order to verify the effect of SVR optimized by ABC, this paper established three types of LAI inversion models with non-optimized parameters (SVR), optimized single parameter (ABC-SVR- $C$ , ABC-SVR- $g$ , ABC-SVR- $\epsilon$ ), and optimized three parameters (ABC-SVR), and compared the accuracy of the three kinds of models. Based on this, we analyzed the sensitivity of LAI inversion model of three key parameters of SVR, and did a significant test on the accuracy of the ABC algorithm optimized SVR model. The study showed: (1) Compared with the model without optimizing parameters, the four models with the SVR parameters optimized by ABC algorithm had higher accuracy, and the optimized three parameters model had better accuracy than the model with optimizing single parameter, the slope of regression straight line reaching 0.797 and decision coefficient reaching 0.775. (2) The three key parameters of SVR have an influence on the accuracy of the LAI model, and compared with the parameters  $C$  and  $g$ , the parameter  $\epsilon$  is more uncertain to the accuracy of the model. (3) At the confidence interval of 95%, the  $P$  value of difference significance test on the slope  $k$ ,  $r^2$ , and RMSE between ABC-SVR model and SVR model all less than 0.005, indicated that the ABC algorithm significantly improved the accuracy of the SVR model.

**Key words:** Support Vector Regression (SVR), Artificial Bee Colony (ABC) algorithm, parameter optimization, Landsat 7, Leaf Area Index (LAI)

**Supported by** National Key Research and Development Program of China (No. 2018YFC1504603)