

农业遥感

基于支持向量回归的关中平原冬小麦估产研究*

曾妍¹, 王迪^{1*}, 赵小娟²

(1. 中国农业科学院农业资源与农业区划研究所 / 农业农村部农业遥感重点实验室, 北京 100081;
2. 青海省农牧业遥感中心, 西宁 810008)

摘要:【目的】小麦产量估测为有关部门制定政策和经济计划提供依据, 在粮食宏观调控中发挥重要作用。【方法】文章采用支持向量回归方法估测冬小麦产量, 以陕西省关中平原的5个市(西安市、宝鸡市、铜川市、渭南市、咸阳市)作为研究区, 将2011—2016年研究区内冬小麦4个生育时期(返青期、拔节期、抽穗—灌浆期、乳熟期)的条件植被温度指数、叶面积指数和每年的单产数据作为总样本, 划分训练集和试验集。基于MATLAB平台和LIBSVM3.23软件包, 建立研究区冬小麦产量回归预测模型, 得到产量预测结果并评价模型精度。【结果】回归模型的决定系数为0.88, 平均绝对百分比误差为6.12%, 均方根误差336.39 kg/hm²。【结论】支持向量回归模型拟合较为理想, 有较高的预测精度和较强的泛化能力。回归时的重要参数有惩罚因子C和核参数 σ , 其中核参数 σ 对模型精度影响更大。研究表明用该回归模型进行冬小麦产量预测是可行的, 支持向量回归方法在粮食产量预测领域有良好的应用前景。

关键词: 冬小麦产量; 支持向量回归; 估产

DOI: 10.12105/j.issn.1672-0423.20190602

0 引言

小麦在中国是仅次于水稻、玉米的第三大粮食作物^[1], 小麦估产可为有关部门制定政策和经济计划提供依据, 在粮食生产、流通和储备各环节的宏观调控中发挥作用。中国是全球最大的小麦生产国和消费国, 常年产量约占全球17%^[2], 中国小麦产量变化对全球小麦产量的预测有重要影响, 进而波及到国际粮价的稳定, 因此, 及时准确估测中国的小麦产量对保证全球粮食安全具有重要意义。

卫星遥感能及时高效地获取大范围地物信息, 广泛应用于农业监测领域, 其中产量估算更是农情遥感监测的重要内容。目前, 基于遥感信息的作物产量估测方法主要有3种: 经验模型或统计模型、机理模型和半经验半机理模型^[3-4]。机理模型能够反映作物产

收稿日期: 2019-10-25

第一作者简介: 曾妍(1998—), 硕士研究生。研究方向: 农业遥感。Email: iszengyan@163.com

* 通信作者简介: 王迪(1977—), 博士、副研究员。研究方向: 农业遥感。Email: wangdicaas@126.com

基金项目: 国家自然科学基金重点项目“基于“三位一体”空间抽样理论研究及其二联查找表研建”(41531179)

2019年12月

量形成过程,但输入参数多且校正困难,导致其应用受限。相较于机理模型,半经验半机理模型虽有简化,但模型输入中需要的2个关键变量,光能利用效率和作物收获指数,都难以在时空分布中进行准确地定量模拟^[5]。统计模型方法尽管对作物产量形成的机理解释性不强,但其结构简单、可操作性强,契合作物估产的业务化应用需求,因此仍然是当前作物遥感估产的常用方法。

经验模型以数学统计分析方法为基础,包括线性模型和非线性模型。而粮食产量受众多不确定因素的影响,是一个复杂的非线性系统,所以非线性估产模型的精度通常高于线性模型^[6]。支持向量回归(Support Vector Regression, SVR)方法是近年来广泛应用的一种非线性回归模型,其基本思想是,找到一个回归平面,让一个集合内所有数据到该平面的距离最近^[7]。相较于神经网络等机器学习方法,它的拟合和泛化推广能力更优异^[8-9],在解决粮食估产等非线性、小样本的实际问题上具有较大优势。

构建估产模型多采用与粮食产量具有高度相关性的参数,包括植被指数、太阳光合有效辐射、生物量等^[10],其中应用最广泛的是能够反映植被绿度状态的归一化植被指数(Normalized Difference Vegetation Index, NDVI)^[11-13]。然而,除植被绿度状态以外,影响农作物生长的土壤水分、气温等指标也与产量的形成紧密关联。为此,文章综合考虑作物生长过程中的水分胁迫指标、作物生长状态指标,选取条件植被温度指数(Vegetation Temperature Condition Index, VTCI)、叶面积指数(Leaf Area Index, LAI)作为模型输入变量,采用支持向量回归方法建立遥感估产模型,并对该方法的精度进行评价,为及时、准确的区域小麦估产提供新的解决途径。

1 材料与方 法

1.1 研究思路

回归的基本思路是利用有限的样本数据,建立输入和输出之间隐含的函数关系。该文在MATLAB环境下,通过LIBSVM3.23工具包^[14],采用支持向量回归方法,基于所选的核函数和最优参数构建产量估测模型,找到各影响因素对产量最佳的映射关系,利用实际产量数据对模型模拟结果进行精度评价,技术路线如图1所示。

1.2 研究区概况

研究区为陕西省中部关中平原(图2a)的5个市,西安市、宝鸡市、咸阳市、渭南市、铜川市,地处106°18'E~110°35'E, 33°35'N~35°52'N(图2b),总面积5.53万km²。关中平原三面环山,渭河横跨平原,地势西高东低,中部较为平坦宽阔,平均海拔约400m。地属大陆性季风气候,多年平均气温13.9℃,多年平均降水量595.3mm。该区域气候温润、水系纵横、土质疏松,是我国重要的粮食产区,也是重要的麦、棉产区,小麦占其耕地面积的50%左右。当地耕作制度为两年三熟,作物种植模式主要是在灌溉区内进行冬小麦、夏玉米轮作,在旱作区域进行夏季休闲式的冬小麦连作^[15]。

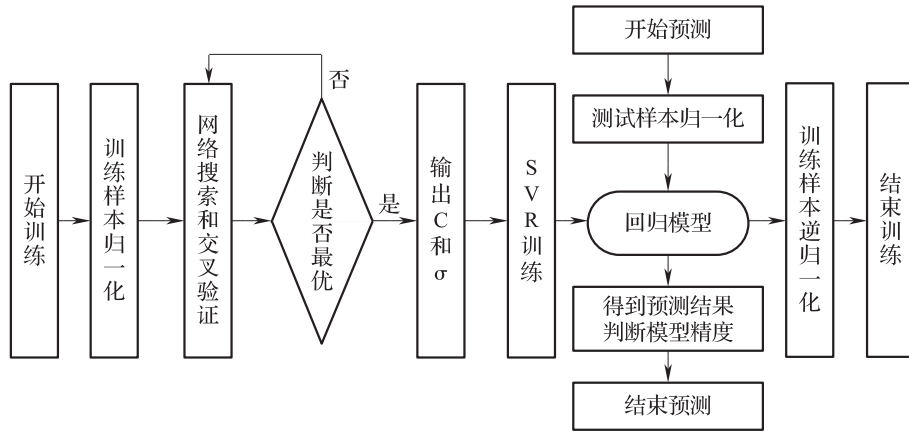


图 1 技术路线图

Fig.1 Technical flowchart

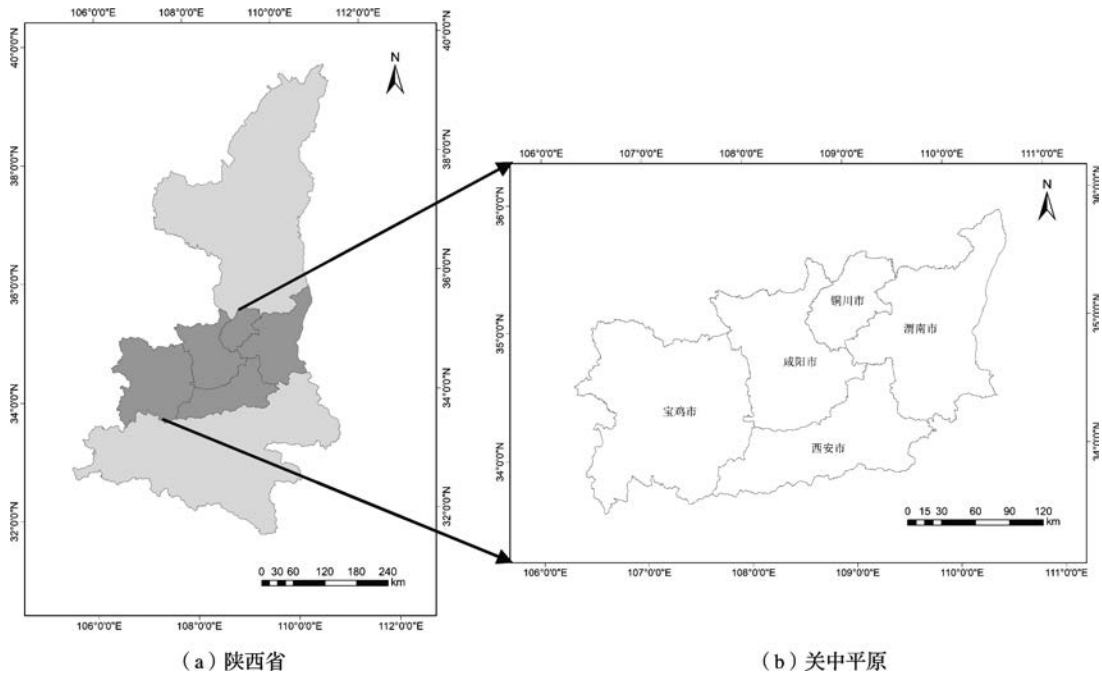


图 2 研究区地理位置

Fig.2 Location of study area

1.3 数据源

该文采用的遥感数据是研究区内 2011—2016 年冬小麦主要生育期（3 月上旬至 5 月下旬）的 Aqua-MODIS 日地表反射率产品（MYD09GA）、日地表温度产品（MYD11A1）、叶面积指数产品（MCD15A3）。所用的产量数据来源于 2011—2016 年的《陕西省统计年鉴》。将各市的小麦种植面积以及总产量汇总，表 1 中列出了计算后得到的对应年份各市冬小麦单产数据。

2019年12月

表1 2011—2016年研究区5市冬小麦单产

地区	2011	2012	2013	2014	2015	2016
西安市	4 399	4 665	4 153	4 510	4 937	4 856
宝鸡市	4 070	4 365	3 879	4 215	4 485	4 410
铜川市	2 598	2 955	2 549	2 768	3 052	2 950
渭南市	3 303	3 780	3 307	3 555	3 948	3 829
咸阳市	4 168	4 500	3 966	4 283	4 658	4 532

1.4 研究方法

1.4.1 VTCI 计算

条件植被温度指数 (VTCI) 是基于归一化植被指数 (NDVI) 和地表温度 (LST) 特征空间呈三角形区域分布的特点提出的, 能体现植物生长过程中受到的水分胁迫, 是一种近实时的干旱监测方法, 计算方法为^[16]:

$$VTCI = \frac{LST_{\max}(NDVI_i) - LST(NDVI_i)}{LST_{\max}(NDVI_i) - LST_{\min}(NDVI_i)} \quad (1)$$

$$LST_{\max}(NDVI_i) = a + bNDVI_i \quad (2)$$

$$LST_{\min}(NDVI_i) = a' + b'NDVI_i \quad (3)$$

式(1)~(3)中, $LST_{\max}(NDVI_i)$ 、 $LST_{\min}(NDVI_i)$ 分别被称作热、冷边界, 表示在研究区内, 当 $NDVI_i$ 值为某一特定值时, 所有像素地表温度的最大值和最小值; $LST(NDVI_i)$ 表示某一像素的 $NDVI$ 值为 $NDVI_i$ 时的地表温度; a , b , a' , b' 是待定系数, 由研究区内的 $NDVI$ 和 LST 散点图近似得到。

计算 VTCI 采用的遥感数据是研究区内 2011—2016 年冬小麦主要生育期 (3 月上旬至 5 月下旬) 的 Aqua-MODIS 日地表反射率产品 (MYD09GA)、日地表温度产品 (MYD11A1)。利用 MODIS 重投影工具 MRT (MODIS Reprojection Tools) 对原始数据进行拼接、重采样、投影转换等预处理操作后, 得到研究区日 LST 和日地表反射率数据。利用日地表反射率数据计算日 NDVI, 用最大值合成方法分别生成逐像素的旬 LST 和旬 NDVI 最大值合成数据, 并按照 VTCI 定义计算出旬 VTCI。

将冬小麦的主要生育期划分为 4 个生育时期, 结合关中平原冬小麦越冬后的生长情况, 将返青期定为 3 月上旬至 3 月中旬, 拔节期为 3 月下旬至 4 月中旬, 抽穗—灌浆期为 4 月下旬至 5 月上旬, 乳熟期为 5 月中旬至 5 月下旬^[17]。计算每个生育期内各旬 VTCI 的平均值作为该生育期的 VTCI 值, 计算研究区 2011—2016 年 5 市各生育期的 VTCI。

1.4.2 LAI 计算

采用 2011—2016 年冬小麦主要生育期内的叶面积指数产品 (MCD15A3) 计算 LAI, 这一产品每 4 d 合成 1 次, 时间分辨率较高, 空间分辨率为 500 m, 有利于农作物长势监测。同样对影像进行投影转换、裁剪等预处理, 利用 S-G 滤波器消除云和大

气等噪声，改善数据质量。然后采用取最大值的方法逐像素地生成旬 LAI，取各生育期内所包含的多旬 LAI 的均值分别作为该生育期 LAI，实现 LAI、VTCI 时间分辨率的统一。

1.4.3 SVR 模型构建

①支持向量回归

支持向量回归的基本思想是采用一个非线性映射把数据映射到高维特征空间，然后在高维特征空间构造回归估计函数，再映回到原空间。通过定义适当的核函数 $k(x, x_i)$ 来实现非线性变换，无需得到确切的映射函数 $\Phi(x)$ ，给计算带来了极大方便^[18]。回归方程为：

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) k(x, x_i) + b \tag{4}$$

式(4)中， $a_i - a_i^*$ 是各支持向量的系数。

②格式转换及归一化

通过运行 FormatDataLibsvm.xls，将数据格式转化为 LIBSVM 规定的输入格式。为避免数据之间的量级差别，对 VTCI、LAI、单产数据分别采用归一化处理，归一化公式为：

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

式(5)中， X_{\max} 和 X_{\min} 分别为数据中的最大值和最小值，经过归一化消除了模型输入由于量纲和单位不同造成的影响，使样本数据更加适应回归建模与分析。

③核函数选择

支持向量回归包括线性和非线性，在线性回归的基础上引入核函数，得到非线性回归。支持向量回归将高维空间内积运算简化成了从低维空间进行输入的核函数计算^[19-20]，选择的核函数的类型决定了支持向量回归的很多特性，从某种程度上直接影响着回归模型的拟合能力和预测精度。因此，核函数的选择是用支持向量回归预测粮食产量的一个关键步骤。

常用的核函数有多项式核函数、径向基 (RBF) 核函数 (即高斯核函数)、Sigmoid 核函数等，选用不同的核函数可以构造出不同的支持向量回归模型。

许多实验研究表明^[21-22]，当缺少先验知识时，选用径向基核函数训练建模的效果较好，所得模型的总体性能较高，而且径向基核函数只含一个未知参数 σ ，易于进行优化，所以该研究选择的核函数为径向基核函数。公式为：

$$K(x, x_i) = \exp(- \|x - x_i\|^2 / 2\sigma^2) \tag{6}$$

④参数寻优

在用支持向量回归解决实际问题时，参数的选择对回归模型的性能影响很大。应用 RBF 核函数时，惩罚因子 C 和参数 σ 与回归模型的学习性能息息相关。惩罚因子 C 用于控制模型对误差范围以外的样本的惩罚程度，也可以说是对离群点的重视程度，惩罚因

2019年12月

子过高会造成过拟合。径向基核函数的参数 σ 反映了训练样本数据的范围或分布, 确定局部邻域的宽度。

该研究选用 LIBSVM 中的 ε -SVR 方法, 用网格搜索和 n-fold 交叉验证的方法进行参数寻优, 确定 C 和 σ 值。当有不同的 C 和 σ 都对应最高的精度时, 把参数 C 最小的那组 C 和 σ 作为最佳参数, 避免惩罚参数 C 太高造成过学习状态。

1.4.4 模型精度评价

该文选取决定系数 (Coefficient of Determination, R^2)、均方根误差 (Root Mean Square Error, RMSE)、平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 作为模型拟合程度的评价指标, 具体公式为:

$$R^2 = \left[\frac{\sum_{i=1}^n (Y_i - \bar{Y})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (E_i - \bar{E})^2}} \right]^2 \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - E_i)^2} \quad (8)$$

$$MAPE = \sum_{i=1}^n \left| \frac{Y_i - E_i}{Y_i} \right| \times \frac{100}{n} \quad (9)$$

式 (7) ~ (9) 中, i 表示第 i 个样本点数据, Y_i 为第 i 个实际产量, 单位为 kg/hm^2 ; E_i 为根据模型拟合的冬小麦产量估算值, 单位为 kg/hm^2 ; \bar{Y} 为实际产量的均值, 单位为 kg/hm^2 ; \bar{E} 为模型估测产量的均值, 单位为 kg/hm^2 。

2 结果与分析

2.1 回归结果及误差分析

研究以 2011—2015 年研究区的数据作为训练集, 采用 SVR 算法构建冬小麦估产模型, 图 3 为 2011—2015 年研究区各市的实际产量和回归产量, 可以看出预测产量和对应的实际产量波动的趋势大致相符。以实际产量和估测产量分别为横纵坐标绘制散点图 (图 4), 添加趋势线和 45° 参考线, 趋势线对比参考线偏离较小, 可见模型对测试集的拟合效果较好。计算模型对训练集样本的预测精度, 回归模型具有较高的决定系数 $R^2=0.88$, 估测产量与实际产量间的 RMSE 为 $336.39 \text{ kg}/\text{hm}^2$, MAPE 为 6.12%。

对训练集的拟合学习说明了 SVR 构建的模型能够以较小的误差拟合历史数据。采用构建好的模型预测 2016 年 5 市冬小麦单产, 表 2 为预测结果及误差, 可见 5 市数据共同构建的模型进行单产估测时有一定误差, 其中面积较小的铜川市的相对误差较小, 仅有 1.28%。分别对 2011—2015 年 5 市冬小麦产量数据进行建模预测 (表 3), 预测结果的 RMSE 平均值为 $232.63 \text{ kg}/\text{hm}^2$, 其中宝鸡市、渭南市的 MAPE 低至 2.98%、3.37%, 5 市 MAPE 的平均值为 5.74%, 可见模型精度较高。

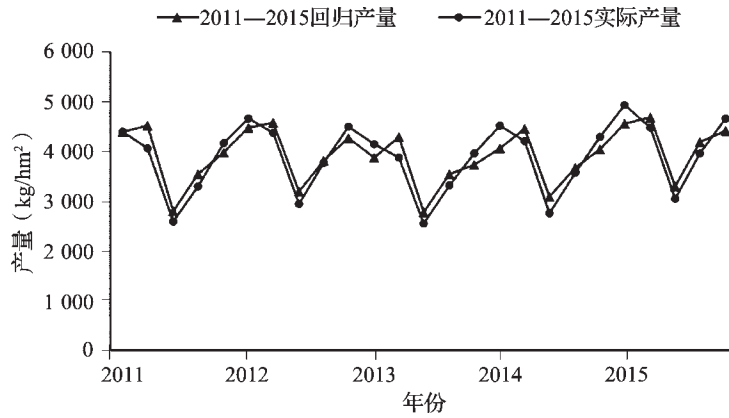


图 3 2011—2015 年 5 市冬小麦实际产量与回归产量对比
Fig.3 Comparison of actual yield and regression yield from 2011 to 2015

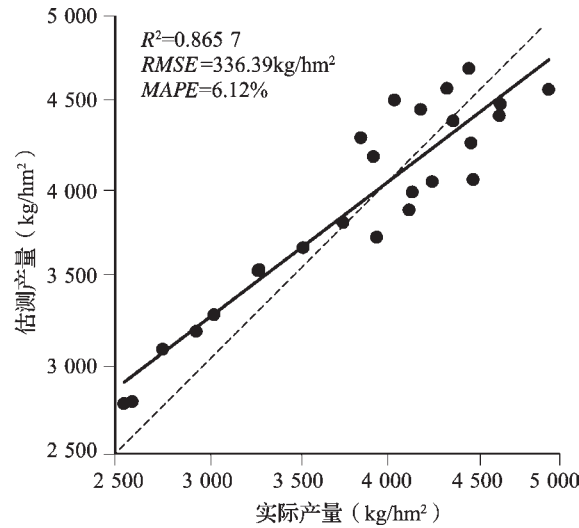


图 4 训练集拟合情况
Fig.4 Fitting effect of training set

表 2 2016 年 5 市冬小麦回归预测结果及误差
Table 2 Regression prediction results and errors

地区	实际产量 (kg/hm ²)	预测产量 (kg/hm ²)	绝对误差 (kg/hm ²)	相对误差 (%)
西安市	4 856	4 346.00	-510.00	10.50
宝鸡市	4 410	4 673.45	263.45	5.97
铜川市	2 950	2 987.63	37.63	1.28
渭南市	3 829	3 917.53	88.53	2.31
咸阳市	4 532	4 055.51	-476.49	10.51

表3 5市分别构建回归模型的预测结果及精度

地区	实际产量 (kg/hm ²)	预测产量 (kg/hm ²)	MAPE (%)	RMSE (kg/hm ²)
西安市	4 856	4 524.94	6.82	331.06
宝鸡市	4 410	4 278.50	2.98	131.50
铜川市	2 950	2 703.49	8.36	246.51
渭南市	3 829	3 699.89	3.37	129.11
咸阳市	4 532	4 207.02	7.17	324.98
平均值	—	—	5.74	232.63

2.2 参数敏感性分析

通过观察不同参数对预测精度的影响,判断对预测精度提高帮助最大的参数。图5、图6给出了C和 σ 值变化时3个误差评价指标的相应变化。结果显示,发生同等程度的变化时,核参数 σ 引起的MAPE、RMSE、 R^2 的改变都比惩罚因子C引起的改变更大。因此,核参数 σ 比惩罚因子C更重要。对2个参数进行耦合分析,发现该研究中当惩罚因子C为2、核参数 σ 为0.354时模型精度最高。

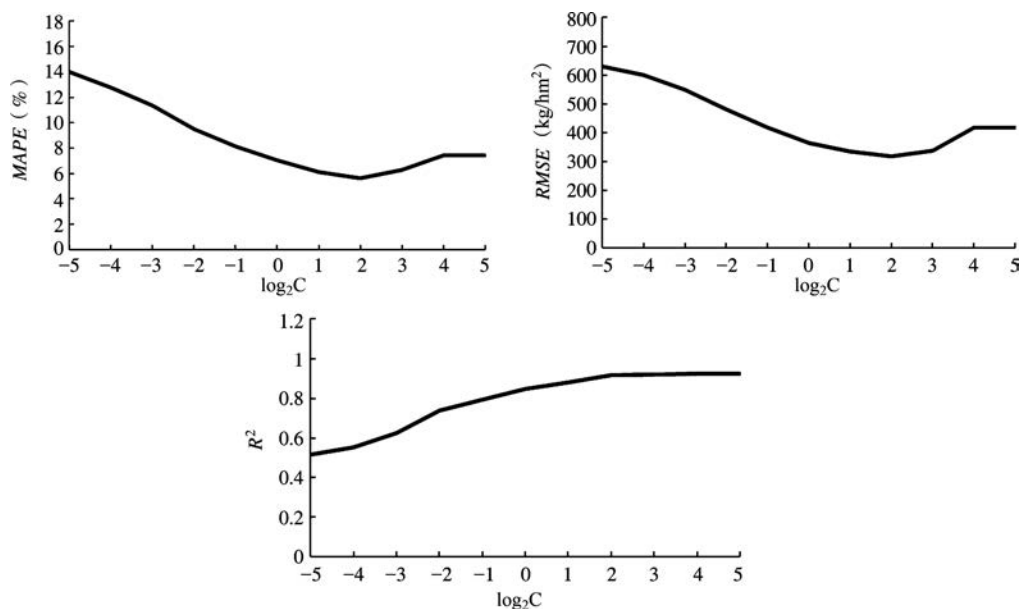


图5 不同参数C条件下的回归模型精度变化

Fig.5 The accuracy change of regression model under different parameter C

3 结论与讨论

3.1 结论

该研究计算了2种与植被生长状况息息相关的参数——VTCI、LAI,将冬小麦各个生育期的VTCI和LAI作为输入变量,划分训练集和测试集,用支持向量回归算法对这些数据进行学习和训练,从而找到其内部隐含的函数关系。寻找最优的模型参数,建立冬小麦各生育期的VTCI、LAI与冬小麦产量之间的支持向量回归估产模型。

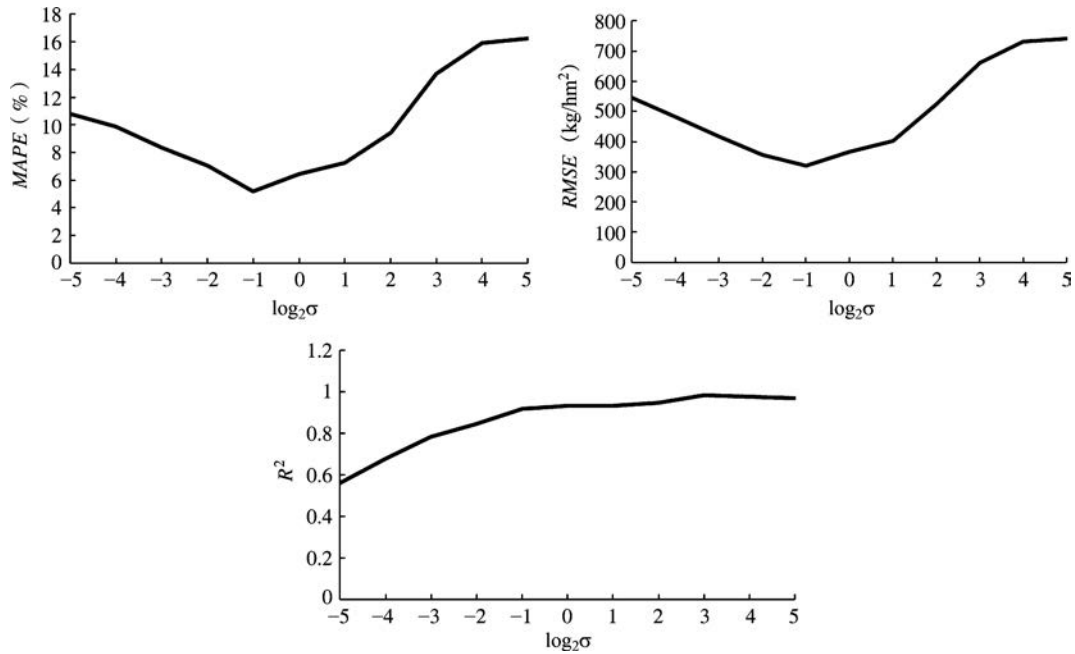


图 6 不同参数 σ 条件下的回归模型精度变化

Fig.6 The accuracy change of regression model under different parameter σ

支持向量回归模型较好地反映了输入的影响因素 VTCI、LAI 同冬小麦单产之间的复杂的非线性映射关系，建立的回归模型决定系数达到 0.88，模型拟合较为理想，此外，模型参数的选取对模型精度有明显的影响，径向基核参数 σ 比惩罚因子 C 对预测结果影响更大。

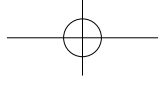
该文采用支持向量机方法对关中平原的冬小麦产量进行预测，在样本较少的情况下，建立的回归模型精度较高，能够推广到其他地区，也可以为其他作物的产量预测研究提供参考，该方法在粮食产量预测领域有良好的应用前景。

3.2 讨论

该文仅选用了 VTCI 和 LAI 2 个指标，只能在一定程度上描述作物本身的生长状态。实际上，影响粮食产量的外部因素众多，还包括很多非遥感信息，如自然灾害、劳动主体的积极性、农业科技水平、农业的基础设施等。该研究缺少足够的先验知识，模型设置过于理想化，并未充分考虑多层次的因素，若要得到更为科学合理的估产结果，仍需继续探索与粮食产量密切相关的指标。

该研究核函数的选定依照前人的经验，而无充分的理论支撑，因此核函数的选用是未来需要进一步研究的内容。随着新技术的不断涌现，支持向量回归和其他理论方法的结合值得深入发掘，例如参数寻优可采用粒子群优化算法^[23]、遗传算法，核函数可选用混合核函数^[24]等。

此外，该研究采用统计年鉴的产量数据进行验证，但产量数据不可避免地存在着差错或漏报^[25]，因此如何有效验证模型精度也是有待解决的问题之一。



2019年12月

参考文献

- [1] 何中虎, 庄巧生, 程顺和, 等. 中国小麦产业发展与科技进步. 农学学报, 2018, 8(1): 99~106.
- [2] 王新华, 鲁艳, 王锐, 等. 我国小麦进出口贸易发展现状、原因及对策. 农业经济, 2017(1): 114~116.
- [3] 赵春江. 农业遥感研究与应用进展. 农业机械学报, 2014, 45(12): 277~293.
- [4] 黎锐, 李存军, 徐新刚, 等. 基于支持向量回归(SVR)和多时相遥感数据的冬小麦估产. 农业工程学报, 2009, 25(7): 114~117.
- [5] Turner D P, Gower S T, Cohen W B, et al. Effects of spatial variability in light use efficiency on satellite-based NPP monitoring. *Remote Sensing of Environment*, 2002, 80(3): 397~405.
- [6] Groten S M E. NDVI-crop monitoring and early yield assessment of Burkina Faso. *International Journal of Remote Sensing*, 1993, 14(8): 1495~1515.
- [7] Cortes Corinna, Vapnik Vladimir. Support-Vector Networks. *Machine Learning*, 1995(20): 273~297.
- [8] Alberto González-Sánchez, Frausto-Solis J, Ojeda-Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research*, 2014, 12(2): 313~328.
- [9] Pandey A, Mishra A. Application of artificial neural networks in yield prediction of potato crop. *Russian Agricultural Sciences*, 2017, 43(3): 266~272.
- [10] 张东霞, 张继贤, 常帆, 等. 遥感技术在主要粮食作物估产中的应用. 测绘科学, 2014, 39(11): 95~98, 103.
- [11] 王莺, 巩垠熙. 遥感光谱技术在农作物估产中的应用研究进展. 中国农学通报, 2019, 35(3): 69~75.
- [12] Mkhabela M S, Bullock P, Raj S, et al. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agricultural and Forest Meteorology*, 2011, 151(3): 385~393.
- [13] Sekhar P S, Ames D P, Suranjan P. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2010(2): 673~696.
- [14] Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol(TIST)*, 2011, 2(3): 27.
- [15] 张树誉, 孙辉涛, 王鹏新, 等. 基于同化叶面积指数和条件植被温度指数的冬小麦单产估测. 干旱地区农业研究, 2017, 35(6): 266~271.
- [16] 王鹏新, 龚健雅, 李小文. 条件植被温度指数及其在干旱监测中的应用. 武汉大学学报(信息科学版), 2001(5): 412~418.
- [17] 黄弘, 王鹏新, 李俐. 关中平原小麦生育期 VTCI 加权估算及其与产量的相关性研究. 干旱地区农业研究, 2011, 29(6): 173~178.
- [18] Li J, Gao F, Guan X, et al. SVR kernel parameters selection based on steady-state genetic algorithm. *IEEE 2006 6th World Congress on Intelligent Control and Automation*, 2006: 4405.
- [19] 辛霄. 基于贝叶斯支持向量机的粮食产量预测的研究. 北京: 首都师范大学, 2012.
- [20] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines. *Machine Learning*, 2002, 46(1/3): 131~159.
- [21] Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 2001, 12(6): 783~789.
- [22] 杨柳, 刘艳芳. 将微粒群和支持向量机用于耕地驱动因子选择的研究. 武汉大学学报(信息科学版), 2010, 35(2): 248~251.
- [23] 赵桂芝, 赵华洋, 李理, 等. 基于混沌-SVM-PSO的粮食产量预测方法研究. 中国农机化学报, 2019, 40(1): 179~183.
- [24] Kumar A, Sminchisescu C. Support kernel machines for object recognition. *2007 IEEE International Conference on Computer Vision*, 2007: 1855~1862.
- [25] 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. 农业工程学报, 2008, 24(2): 290~298.

Study on yield prediction of winter wheat in Guanzhong Plain based on SVR

Zeng Yan¹, Wang Di^{1*}, Zhao Xiaojuan²

(1. Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences/Key Laboratory of Agricultural Remote Sensing, Ministry of Agricultural and Rural Affairs Beijing 100081, China;

2. Center of agriculture and husbandry remote sensing of Qinghai province, Xining 810008, China)

Abstract: [**Purpose**] Prediction of wheat yield can provide basis for relevant departments to formulate policies and economic plans, and play an important role in macro-control of grain. [**Method**] Support Vector Regression (SVR) method is used to estimate winter wheat yield. Five cities in Guanzhong Plain of Shaanxi Province (Xi'an City, Baoji City, Tongchuan City, Weinan City and Xianyang City) were selected as the study area. Vegetation Temperature Condition Index (VTCI) and Leaf Area Index (LAI) of four growth stages (green-turning stage, jointing stage, heading-filling stage and milk-ripening stage) of winter wheat in the study area from 2011 to 2016 were selected. VTCI, LAI and annual yield data were used as total samples which were divided into training set and experimental set. Based on the MATLAB platform and LIBSVM3.23 software package, the study established the prediction model of winter wheat yield in the study area, got the output prediction results and evaluated the accuracy of the model. [**Result**] The determinant coefficient of the regression model was 0.88. The average absolute percentage error was 6.12%. The root mean square error of the model was 336.39 kg/hm². [**Conclusion**] The support vector regression model fitting has high prediction accuracy and strong generalization ability. The important parameters in regression are penalty factor C and kernel parameter σ , in which kernel parameter σ has a greater impact on accuracy of model. The results show that it is feasible to use this regression model to predict winter wheat yield, and the support vector regression method has a good application prospect in the field of grain yield prediction.

Key words: winter wheat yield; support vector regression; yield estimation