

doi:10.6041/j.issn.1000-1298.2022.01.022

融合多源时空数据的冬小麦产量预测模型研究

王来刚^{1,2} 郑国清^{1,2} 郭燕^{1,2} 贺佳^{1,2} 程永政^{1,2}

(1. 河南省农业科学院农业经济与信息研究所, 郑州 450002; 2. 农作物种植监测与预警河南省工程实验室, 郑州 450002)

摘要: 为提高大尺度冬小麦产量预测精度,以 2005—2019 年河南省遥感数据、气象数据、土壤含水率等多源时空数据为特征变量,分析其与小麦单产的相关性,并基于随机森林算法对特征变量进行了重要性分析,构建了融合多源时空数据的冬小麦产量预测模型。结果表明:增强型植被指数(Enhanced vegetation index,EVI)、日光诱导叶绿素荧光(Solar-induced chlorophyll fluorescence,SIF)与高程为小麦产量预测的重要因子,与小麦产量呈高度正相关,对小麦产量预测的重要性指标均超过 0.45,远大于土壤含水率、降水量、最高温度、最低温度等因子;基于随机森林算法构建的小麦不同生长阶段产量预测模型中,以 10 月—次年 5 月和 10 月—次年 4 月为特征变量的产量预测模型精度较高, R^2 分别为 0.85 和 0.84, RMSE 分别为 821.55、832.01 kg/hm²,在空间尺度上,豫西和豫南丘陵山地模型预测相对误差高于平原地区。该研究结果可为大尺度作物产量预测提供参考。

关键词: 冬小麦; 产量预测; 遥感; 多源时空数据; 随机森林算法

中图分类号: S127 文献标识码: A 文章编号: 1000-1298(2022)01-0198-07 OSID: 

Prediction of Winter Wheat Yield Based on Fusing Multi-source Spatio-temporal Data

WANG Laigang^{1,2} ZHENG Guoqing^{1,2} GUO Yan^{1,2} HE Jia^{1,2} CHENG Yongzheng^{1,2}

(1. Institution of Agricultural Economy and Information, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China
2. Henan Engineering Laboratory of Crop Planting Monitoring and Warning, Zhengzhou 450002, China)

Abstract: In order to improve the prediction accuracy of winter wheat yield in large scale region, taking remote sensing data, meteorological data, soil moisture data of Henan Province from 2005 to 2019 as characteristic variables, the correlation between them and wheat yield was analyzed. The importance of characteristic variables was analyzed based on random forest algorithm. And a wheat yield prediction model was established by means of fusing multi-source spatio-temporal data. The results showed that enhanced vegetation index (EVI), solar-induced chlorophyll fluorescence (SIF) and elevation was an important factor for remote sensing estimation of wheat yield, which was highly positively correlated with wheat yield. The importance of EVI, SIF and elevation to wheat yield exceeded 0.45, far greater than soil moisture, rainfall, maximum temperature, minimum temperature and other factors. The yield prediction model based on random forest algorithm and constructed with the wheat growth stage from October to next May and October to next April as the characteristic variables had higher accuracy, coefficient of determination (R^2) were 0.85 and 0.84, and respectively, the root mean square error (RMSE) were 821.55 kg/hm² and 832.01 kg/hm². The prediction relative errors in hills and mountains of western and southern Henan was higher than that in plain areas. The research results provided a reference for large-scale crop yield.

Key words: winter wheat; yield prediction; remote sensing; multi-source spatio-temporal data; random forest algorithm

收稿日期: 2021-10-13 修回日期: 2021-11-15
基金项目: 国家重点研发计划项目(2018YFD0300702)、河南省重大科技专项(171100110600)和河南省农业科学院创新团队项目(2021TD11)
作者简介: 王来刚(1979—),男,副研究员,主要从事农业遥感研究,E-mail: 26585273@qq.com
通信作者: 郑国清(1964—),男,研究员,主要从事农业信息技术研究,E-mail: zgqzx@hnagri.org.cn

0 引言

及时准确掌握区域农作物产量信息,能够为粮食生产宏观调控、经济政策制定和农作物保险提供决策支持,对服务国家粮食安全战略具有重要意义。小麦是我国粮食作物之一,也是我国重要的口粮作物,对粮食安全和居民生活都具有重要意义。

冬小麦产量形成是一个复杂的过程,中间涉及非常多的生理生化过程,气象条件、土壤条件、地理环境、物候信息等都是其产量估算需要考虑的因素,产量可看成是一段时期内多个影响因子相互叠加的结果。遥感技术在农业中已广泛应用,可以从遥感数据中提取出各种相关信息用于产量预测。特别是植被指数,如归一化差异植被指数(Normalized difference vegetation index, NDVI),已被广泛使用^[1-3]。其他指数如增强植被指数(Enhanced vegetation index, EVI)^[4]、作物水分胁迫指数^[5]、绿色植被指数、土壤调节植被指数^[6]等,也被用于作物产量预测。气象数据(如降水量、空气温度等)^[7-9]和土壤条件数据(如土壤含水率和温度)^[10],在产量预测中经常被用作作物生长环境指标。日光诱导叶绿素荧光遥感数据(SIF)为植物进行光吸收后叶绿素 a 释放出的微弱电磁辐射信号^[11],与光合速率紧密耦合。近年来,SIF 数据日益增多,为作物估产提供了一条新的有效途径^[12-14]。基于遥感数据的作物产量预测模型主要有 2 种:作物模拟模型和经验统计模型。尽管作物模拟模型精确地模拟了作物生长的物理过程,但需要精细的土壤和气象数据,这阻碍了它们的大规模应用。相比之下,经验统计模型简单,需要较少的输入数据,因此被广泛用作基于过程的模型的常见替代方案。上述相关研究多集中在利用植被指数和气象数据预测作物产量,将 SIF 数据作为产量预测因子的研究相对较少。

随机森林回归模型是一种基于机器学习的统计方法,它能处理高维度的数据并且估算结果具有较高的准确率。杨北萍等^[15]将气象数据和遥感数据作为输入变量,基于随机森林回归方法预测了水稻产量。王鹏新等^[16]通过随机森林回归算法获取玉米主要生育时期各个特征变量的权重,构建了加权特征变量与玉米单产间的回归模型。程千等^[17]采用偏最小二乘回归、支持向量机回归和随机森林回归 3 种机器学习算法进行冬小麦产量估测,其中随机森林回归模型估测精度最高。本文以河南省为研究区域,拟通过探索分析多年小麦生长季的遥感数据、气象数据、土壤含水率数据、地形数据与小麦单

产的相关性,利用随机森林回归模型融合多源时空数据构建冬小麦产量预测模型,以期提高大尺度冬小麦产量预测精度。

1 材料与方法

1.1 研究区域概况

本文将河南省作为研究区域。河南省位于我国中部的黄河中下游地区,地处我国第二阶梯向第三阶梯的过渡地带,北、西、南三面的太行山、伏牛山、桐柏山、大别山沿省界呈半环形分布;中、东部为黄淮海冲积平原;西南部为南阳盆地。全省耕地面积 $8.1103 \times 10^6 \text{ hm}^2$,占中国总耕地面积的 6.24%,是我国小麦的最适生态区,小麦是河南省最主要的粮食作物,全省小麦播种面积、总产量均居全国首位。

1.2 数据获取与处理

研究需要的数据包括 2005—2019 年河南省小麦生长季的遥感数据、气象数据、土壤含水率数据、数字高程和小麦单产数据,所有数据利用小麦种植空间分布数据进行掩膜,并叠加河南省县级行政边界,经栅格均值计算,得各县区的变量均值。

1.2.1 遥感数据

卫星遥感参数分别选择增强型植被指数(EVI)和日光诱导叶绿素荧光(SIF)来反映小麦长势和光合作用。EVI 是一个广泛用于监测作物长势和预测产量的植被指数^[18],与 NDVI 相比,EVI 对较高的冠层叶面积指数敏感,受大气气溶胶影响较小。本文使用了美国国家航空航天局(NASA)MODIS EVI 产品(MOD13A1, Collection 6),该产品的空间分辨率为 500 m,时间分辨率为 16 d。近年来,随着技术的发展,SIF 遥感数据不断增多、空间分辨率不断提高,使得利用 SIF 遥感数据直接预测作物产量成为可能^[19-20]。SIF 数据是从全球空间连续日光诱导叶绿素荧光数据集产品(CSIF)获得(DOI:10.6084/m9.figshare.6387494),其时间分辨率为 4 d 和全球空间分辨率为 0.05°。应用最大值合成技术生成冬小麦生长季(10 月至次年 5 月)月时间尺度的 EVI 和 SIF。

1.2.2 气象数据

从 Terra Climate 数据集获得小麦整个生育期每月的降水量、最高温度、最低温度数据作为影响产量形成的特征变量。这些数据集对 1958—2019 年间的全球地表具有高空间分辨率(约 4 km),可用于区域作物产量预测^[21-22]。

1.2.3 土壤含水率数据

土壤含水率数据来自国家青藏高原科学数据中心的土壤水分数据集(DOI: 10.5281/zenodo.

4738556),时间分辨率为月,空间分辨率为0.05°。

1.2.4 数字高程数据

来源于美国国家航空航天局(NASA)ASTERG-DEM,栅格尺寸为30 m×30 m。

1.2.5 冬小麦单产数据

河南省县级冬小麦单产数据来源于《河南省统计年鉴》,2005—2019年间具有有效小麦单产数据的县市105个。

1.3 研究方法

1.3.1 随机森林回归模型

随机森林是由多棵分类回归树(Classification and regression tree,CART)构成的组合分类模型^[23]。该研究将各年份各县市的小麦产量因子特征变量数据和产量数据进行集成,共同构成随机森林的样本数据集,通过自助法(Bootstrap)从原始样本集采样得到构建*N*棵树所需的*N*个子集,每次未被抽到的数据称为袋外数据(Out-of-bag,OOB),用来进行内部误差估计和变量重要性评价;生成每棵树时,从规模为*M*的特征变量集中随机选择*m*个变量(*m*<*M*),对于回归模型,采用均方差作为节点分裂标准,递归执行选取最优分枝的操作。由于随机森林采用样本和特征的双重随机抽样构建决策树,因此即使不对决策树进行剪枝操作也不会出现传统CART决策树过拟合的现象^[14]。最后将这些树的结果取平均值即为目标变量的预测值。在随机森林算法中,根据试验结果实时优化决策树数目*N*和创建分枝所需特征变量个数*m*这2个参数。本文基于随机森林算法的冬小麦产量预测技术流程如图1所示。在对数据预处理的基础上,深入分析特征变量与冬小麦产量时空相关性,并基于随机森林算法进行特征变量重要性分析,建立融合多源时空数据冬小麦产量预测模型,最后从时间上和空间上对模型预测精度进行评价分析。

1.3.2 特征变量重要性评价

随机森林算法是众多决策树并行式集成学习的方法,除了可对数据集进行分类和回归外,还可以进行变量重要性分析、奇异值检测等^[24],能够解释若干自变量对因变量的作用。通过模型内部重要性评价结果,分析不同影响因子对小麦产量的影响程度。这种方法是直接分析评价每种特征变量对模型预测准确率的影响,基本思想是重新排列特征变量的顺序,观测模型准确率的降低程度。对于不重要的特征变量,这种方法对模型准确率的影响很小,但是对于重要特征变量却会极大降低模型的准确率。具体过程为:

(1)对于随机森林中的每一棵决策树,使用相

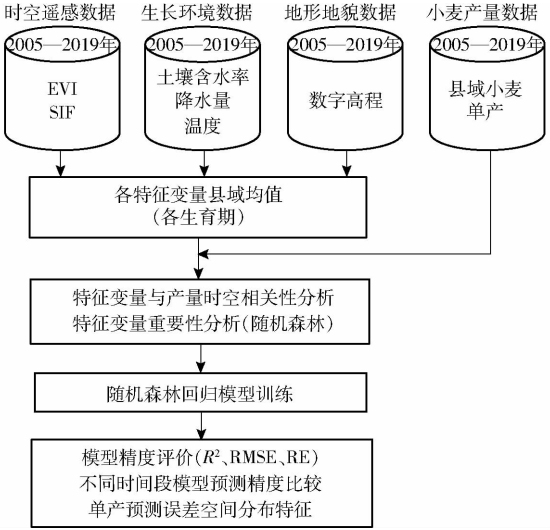


图1 基于随机森林算法的冬小麦产量预测流程图
Fig.1 Flow chart of winter wheat yield prediction based on random forest algorithm

应的袋外数据(OOB)来计算它的袋外数据误差,记为 e_{OOB1} 。

(2)随机地对OOB数据所有样本的特征变量加入噪声干扰,随机改变样本在特征变量处的值,再次计算它的袋外数据误差,记为 e_{OOB2} 。

(3)假设随机森林中有 N_{tree} 棵树,则特征变量的重要性指标为 $\sum(e_{OOB2}-e_{OOB1})/N_{tree}$ 。之所以可以用这个表达式作为相应特征的重要性度量值,是因为:若给某个特征随机加入噪声之后,袋外的准确率大幅度降低,则说明这个特征对于样本的分类结果影响很大,即重要程度比较高。

1.3.3 模型精度评价

采用决定系数(Coefficient of determination, R^2)、均方根误差(Root mean square error, RMSE)和相对误差(Relative error, RE)3个指标评价模型精度。其中 R^2 评价预测模型拟合能力, RMSE和RE评价预测值和实测值离散程度。

2 结果与分析

2.1 特征变量与冬小麦产量时空相关性分析

为研究特征变量与小麦产量之间的关系,分别从时间和空间上进行了各特征变量数据与产量之间相关性分析,见图2,箱线图是特征变量与产量相关性的时间模式,相关性的空间模式基于具有最高相关系数的月份,即方框图中的红点。在时间相关性模式上,遥感参数EVI和SIF与产量都呈正相关,其中4月相关性达到最高峰,相关系数可达0.7左右,随后由于小麦冠层衰老和籽粒形成而开始下降,使得遥感参数与产量之间的相关性降低。与遥感参数不同,土壤含水率与产量之间的最大相关系数出现

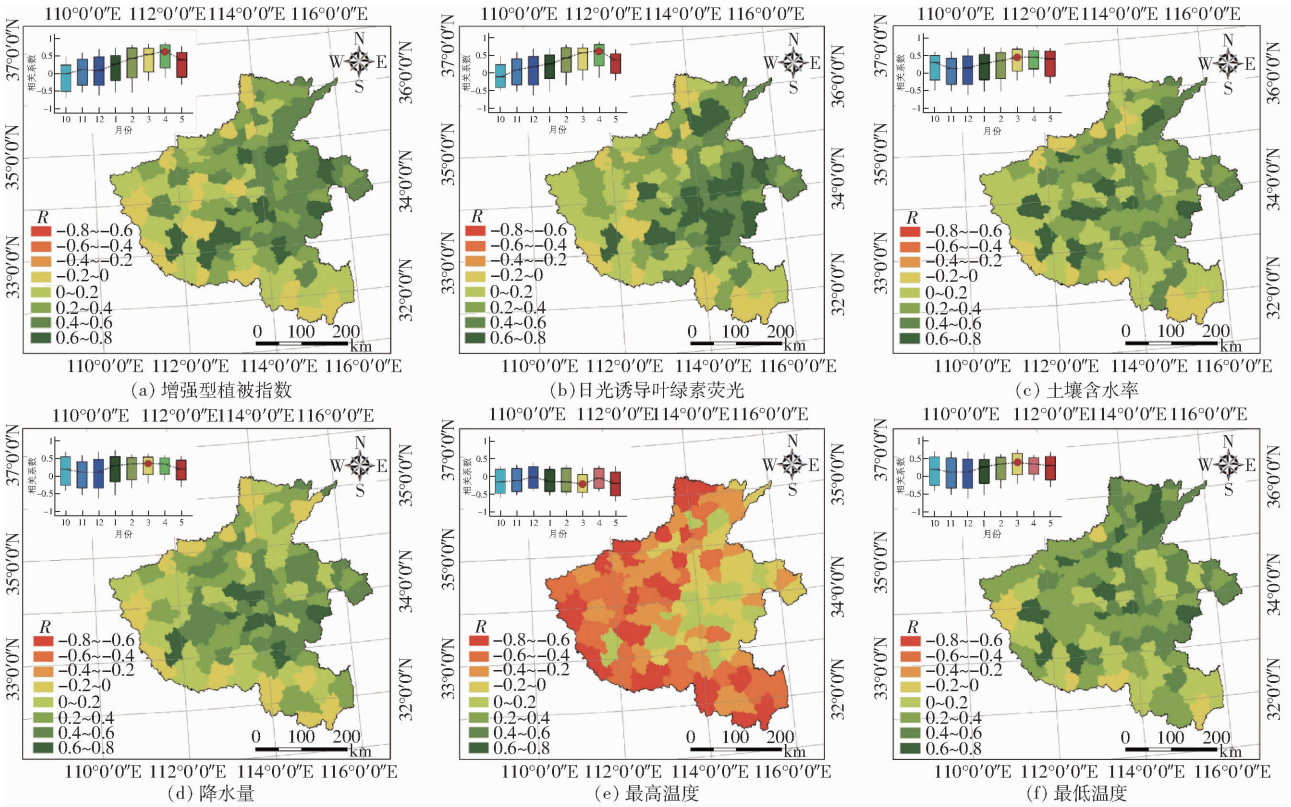


图 2 特征变量与冬小麦产量的时空相关性

Fig. 2 Spatiotemporal correlations between characteristic variable and wheat yield

时期并不突出,主要在 10 月和次年 3、4 月相关性比较大,但整体相关性低于遥感参数。气象变量中,各生长阶段降水量和最低温度与小麦气象波动产量为正相关,降水量与产量之间的相关性最大的月份与土壤含水率基本相同,最低温度与产量相关性最大的月份为 3 月。3 月小麦主要处于拔节-孕穗阶段,最低温度是影响小麦产量的重要因素。最高温度与产量主要呈负相关性,但相关系数较小。

选择特征变量与小麦产量相关性最高的月份,分析特征变量与小麦产量相关性的空间分布特征。遥感参数 EVI 和 SIF 在豫北平原和豫东平原与产量相关性较高,相关系数最高达 0.8。豫西豫南山地丘陵地区和部分种植模式混杂的县区相关性较低,相关系数大部分在 0~0.2 之间,这些差异可能是由于豫西、豫南山地丘陵地区地块破碎,EVI 和 SIF 混合像元较多,遥感参数难以真实反映小麦长势情况。土壤含水率与产量相关性在空间分布上,无明显特征。降水量在灌溉条件较差的丘陵地区与产量相关性较高,豫北地区相关性略低,可能由于豫北地区灌溉条件较好,小麦生长对降水量依赖性不高。最高温度在全省与小麦产量主要呈负相关,在山地丘陵地区相关性较高。最低温度在豫北地区相关性较高,豫北地区冬小麦易遭受晚霜冻害,最低温度是影响产量的一项重要制约因素。

2.2 特征变量重要性分析

冬小麦特征变量可以分为动态特征变量(遥感和气象变量等)和静态特征变量(高程)。利用 2005—2019 年的冬小麦特征动态变量数据与产量分别进行随机森林的 OOB 重要性分析,按照模型准确率降低的程度对特征变量由大到小排序(表 1)。4 月 EVI 和 SIF、2 月降水量排名位居前 3 位。开花期是冬小麦体内新陈代谢最旺盛的生长时期,正是小麦产量形成的关键时期。河南省冬小麦的开花期主要集中在 4 月,EVI 和 SIF 与冬小麦叶面积指数、生物量等苗情指标具有较强的相关性,因此在各特征变量中重要性更高,重要性指标在 0.4 左右。2 月是河南省冬小麦返青期,是促进苗情转化升级的关键时期,降水量是一个重要的影响因素。重要性指标第 4~6 位分别为 10 月土壤含水率、4 月最低温度和 2 月 EVI。10 月是冬小麦播种出苗时期,土壤墒情是出苗品质的关键因子。4 月初小麦处于孕穗期,很容易受到低温冻害,特别是豫北地区。2 月 EVI 代表了小麦生长前期的总体苗情,也是影响产量的重要因子。

对 EVI、SIF、土壤含水率、降水量、最高温度、最低温度和高程 7 大类特征变量进行基于随机森林的袋外 OOB 重要性分析(图 3)。结果表明,EVI、SIF 和高程对小麦产量的重要性远大于其他因素,重要

表 1 动态特征变量重要性指标

Tab.1 Importance of dynamic feature variables								
特征 变量	月份							
	10	11	12	1	2	3	4	5
EVI	0	0.02	0.05	0.05	0.25	0.32	0.37	0.15
SIF	0	0.02	0.06	0.06	0.22	0.30	0.41	0.12
降水量	0.21	0.05	0.08	0.09	0.34	0.25	0.25	0.28
最高温度	0.16	0.06	0.04	0.06	0.13	0.18	0.24	0.10
最低温度	0.16	0.09	0.06	0.09	0.11	0.15	0.30	0.09
土壤含水率	0.25	0.10	0.10	0.11	0.16	0.30	0.25	0.09

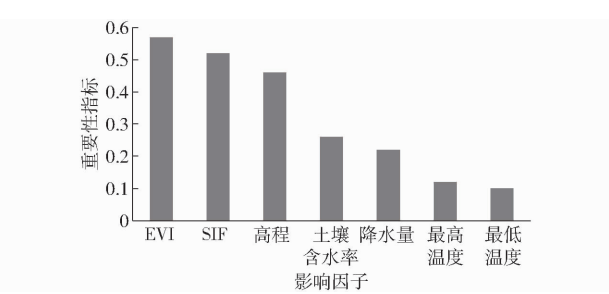


图 3 影响因子重要性统计图

Fig.3 Importance of impact factors

性指标均超过 0.45,说明除了遥感参数外,地形对产量的影响也非常大。其次是土壤含水率和降水量,说明含水率是影响小麦产量的重要环境因子。最后是最高温度和最低温度,重要性指标只有 0.10 左右。综上,描述作物生长状况 EVI 和 SIF 对产量预

测的贡献大于描述作物生长环境的气象和土壤因子。

2.3 不同时间段模型预测精度比较

为了深入研究不同时间段的冬小麦单产预测精度,并结合单产预测实际工作需求,分别输入 10 月一次年 2 月、10 月一次年 3 月、10 月一次年 4 月、10 月一次年 5 月的特征变量,对小麦产量进行随机森林建模,结果见图 4。由图 4 可知,基于小麦整个生长阶段 10 月一次年 5 月和 10 月一次年 4 月特征变量的估产精度较高, R^2 分别为 0.85 和 0.84, RMSE 均最小,2 个时间段模型预测精度基本相同。河南省冬小麦抽穗期和开花期主要集中在 4 月,开花期是冬小麦体内新陈代谢最旺盛的生长时期,正是小麦产量形成的关键时期,温度和降水量与小麦产量密切相关^[25]。仅有小麦生长前期的 10 月一次年 2 月和 10 月一次年 3 月特征变量的估产精度相对偏低, RMSE 均大于 900 kg/hm²。因此,4 月底或 5 月初是冬小麦产量预测的最佳时间。

2.4 单产预测误差空间分布特征

输入冬小麦各生长阶段的 EVI、SIF、土壤含水率、气象要素和高程等全部变量,以小麦产量为目标变量构建随机森林模型,预测 2018、2019 年河南省各县冬小麦产量,图 5 为冬小麦产量预测相对误差与数字高程叠加空间分布情况。整体上,2 年模型

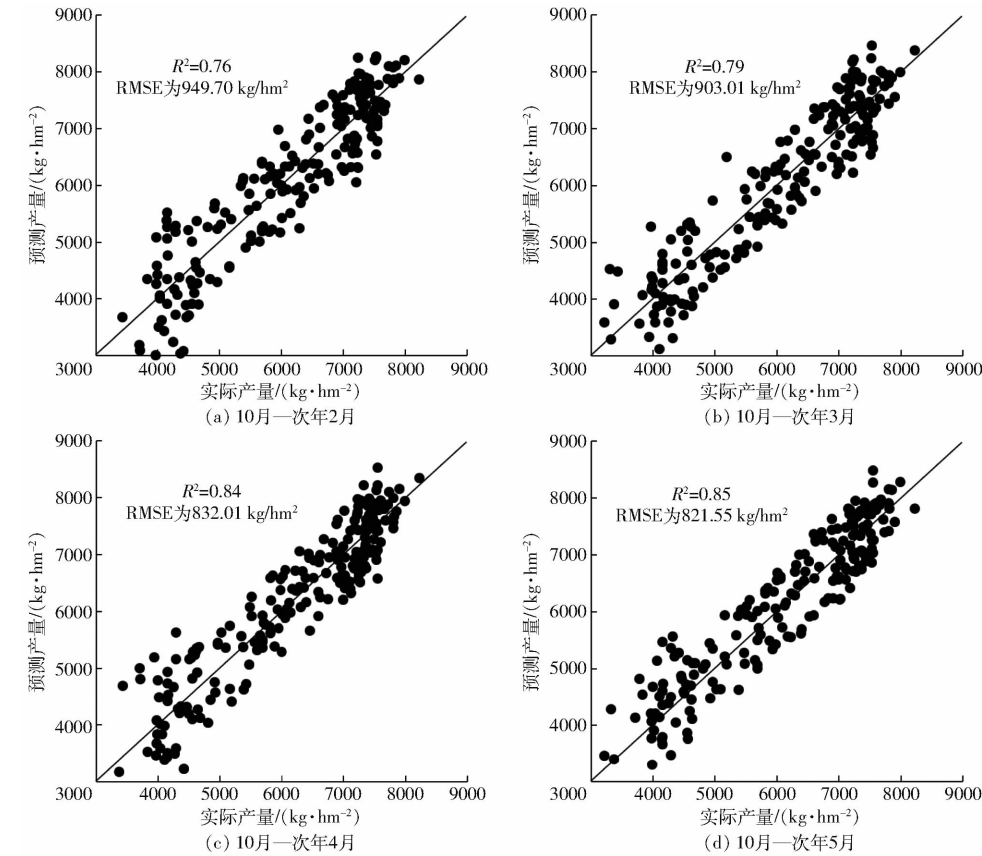


图 4 不同时间段模型预测结果对比

Fig.4 Comparisons of prediction results of different periods

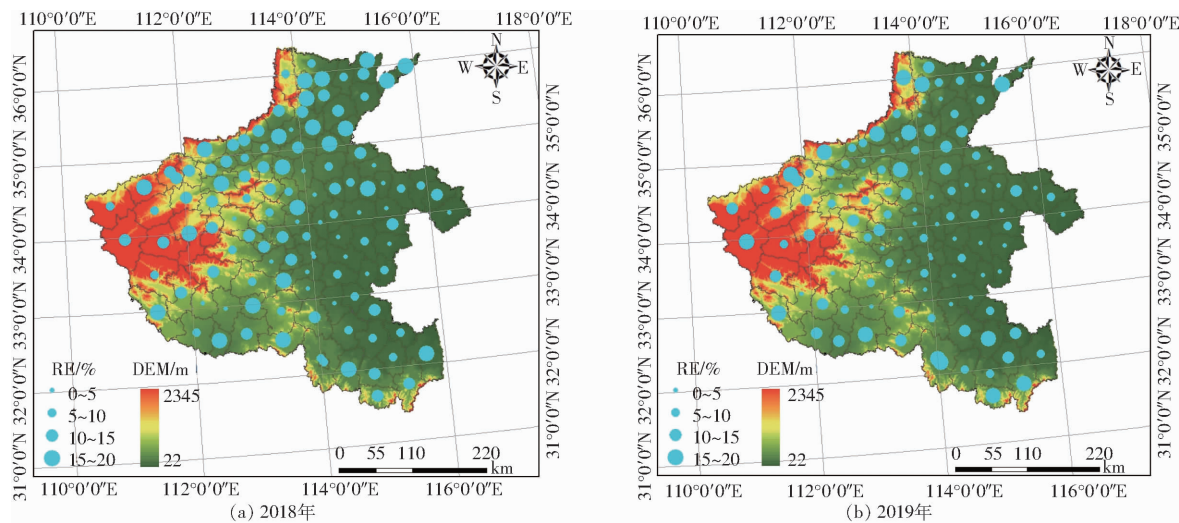


图 5 河南省冬小麦产量预测相对误差空间分布
Fig.5 Relative error spatial distributions of winter wheat yield prediction in Henan Province

预测相对误差空间分布基本相似,产量预测相对误差均在 20% 以内,平原地区模型预测相对误差大部分在 10% 以内,而豫西和豫南丘陵山地模型预测相对误差高于平原地区,这可能由于丘陵山地区域地块破碎,而模型输入数据 EVI、SIF 等数据空间分辨率较低,混合像元现象比较严重,很难准确反映小麦实际长势情况。另外,信阳地区和开封部分县小麦种植地块零星破碎,预测精度也相对偏低。2 年数据对比可以看出,2019 年预测相对误差低于 2018 年,特别在豫北地区,2018 年预测相对误差明显较大。这可能由于 2018 年 4 月初河南豫北地区发生一次较重的晚霜冻害,不同小麦品种抗冻害能力差别较大,产量损失程度更是不一致,而该预测模型无品种特征变量。

3 讨论

研究表明,SIF 与小麦产量呈正相关,这一结果得到了已有研究的支持^[26-27]。研究区内,SIF 和 EVI 与小麦产量的相关性相近,EVI 对绿叶结构、叶绿素含量和生物量的变化比较敏感,而 SIF 有着直接指示植被光合作用的巨大潜力,2 种数据可能共同反映一些与地上作物生物量相关的信息。本文研究的小麦产量预测模型输入变量重点考虑了气象条件、植被生长状态、地形地貌等,但在实际生产过程中小麦产量会受到多种环境因素的影响,以及各种人为因素造成的产量波动,这些都会影响最终的预测精度。气象条件中虽然用到了最高温度和最低温度等特征变量,但显然预测模型对灾害信息反应不敏感,在对 2018 年冬小麦产量预测中,受到晚霜冻害的区域小麦预测精度偏低。根据实际调查,小麦冻害除了与最低温度有关之外,与品种、种植密度和

土壤特性都有明显的相关性。因此,未来的研究中可以增加灾害分布图层作为输入变量,从而提高模型预测精度。

随机森林算法在预测产量上具有很大的潜力,但并不意味着该算法一定是预测产量的最优算法,后期研究将对比深度神经网络、卷积神经网络等深度学习算法在作物预测产量的优缺点,解决算法对于极端天气等突发事件的适应性^[28-29]。这项研究受到了一些不确定性的困扰,其中一个问题是 SIF 和 EVI 的低信噪比和粗糙的时空分辨率^[30-31],可能无法真实反映小麦长势空间特征的细节。特别是在山地丘陵地块破碎的地区,产量预测精度都有不同程度下降。因此,如果要提高遥感数据在作物产量预测中的作用,时空连续高分辨率数据是非常重要的。随着卫星遥感技术发展,将实现具有更高空间和时间分辨率的 EVI 和 SIF。例如,欧空局设计的 FLEX (Fluorescence explorer),提供全球尺度上的植被叶绿素荧光测量卫星数据,目前其空间分辨率设定为 300 m × 300 m,轨道幅宽为 150 km,非常有利于小尺度精准化的作物估产研究^[32]。

4 结论

(1) 以遥感数据、气象数据、土壤含水率数据和地形数据为特征变量,分析了 2005—2019 年小麦产量与特征变量的相关性时空特征,基于随机森林算法对特征变量进行了重要性分析,并建立了不同时间段的小麦产量预测模型。

(2) SIF 是指示植被光合变化的有效探针,与小麦产量呈高度正相关,在小麦产量预测特征变量重要性分析中,与 EVI 起着同等重要作用,是小麦产量预测的一项重要因子。

(3)在7大类特征变量中,EVI,SIF和高程对小麦产量的重要性远大于其他因素,重要性指标都超过0.45,其次是土壤含水率和降水量,最后是最高温度和最低温度。高程为静态特征变量,在空间有差异,年度间无变化。

(4)基于随机森林算法以小麦生长阶段10月—

次年5月和10月一次年4月为特征变量构建的产量预测模型精度较高, R^2 分别为0.85和0.84, RMSE分别为821.55、832.01 kg/hm²。在产量预测误差空间分布特征上,全省相对误差均在20%以内,平原地区模型预测相对误差大部分在10%以内,而豫西和豫南丘陵山地模型预测相对误差高于平原地区。

参 考 文 献

[1] SHRESTHA R, DI L, YU E G, et al. Regression based corn yield assessment using MODIS based daily NDVI in Iowa state[C] //2016 Fifth International Conference on Agro-Geoinformatics. IEEE, 2016.

[2] RANJAY S, LIPING D, EUGENE G Y, et al. Regression model to estimate flood impact on corn yield using MODIS NDVI and USDA cropland data layer[J]. Journal of Integrative Agriculture, 2017,16(2):398–407.

[3] LOFTON J, TUBANA B S, KANKE Y, et al. Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index[J]. Sensors, 2012, 12(6):7529–7547.

[4] JINRU X, BAOFENG S. Significant remote sensing vegetation indices: a review of developments and applications[J]. Journal of Sensors, 2017, 2017:1353691.

[5] GHAEMI A, MOAZED H, RAFIE RAFIEE M, et al. Determining CWSI to estimate eggplant evapotranspiration and yield under greenhouse and outdoor conditions[J]. Iran Agricultural Research, 2016, 34(2):49–60.

[6] PANDA S S, AMES D P, SURANJAN P. Application of vegetation indices for agricultural crop yield prediction using neural network techniques[J]. Remote Sensing, 2010, 2(3):673–696.

[7] MATHIEU J A, AIRES F. Using neural network classifier approach for statistically forecasting extreme corn yield losses in eastern United States[J]. Earth and Space Science, 2018, 5(10):622–639.

[8] FUNK C, PETERSON P, LANDSFELD M, et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes[J]. Scientific Data, 2015, 2(1):2539–2558.

[9] CRANE-DROESCH A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture[J]. Environmental Research Letters, 2018, 13(11):1–12.

[10] POURMOHAMMADALI B, HOSSEINIFARD S J, SALEHI M H, et al. Effects of soil properties, water quality and management practices on pistachio yield in Rafsanjan region, southeast of Iran[J]. Agricultural Water Management, 2018, 213:894–902.

[11] BOCCA F F, RODRIGUES L. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling[J]. Computers and Electronics in Agriculture, 2016, 128:67–76.

[12] 高云. 基于OCO-2日光诱导叶绿素荧光遥感数据的区域作物估产研究[D]. 南京: 南京大学, 2020.

GAO Yun. The ability of OCO-2 sun-induced chlorophyll fluorescence data to estimate regional crop yield[D]. Nanjing: Nanjing University, 2020. (in Chinese)

[13] JC A, ZHAO Z A, FT B, et al. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches[J]. Agricultural and Forest Meteorology, 2021,297:108275.

[14] ZHANG L, ZHANG Z, LUO Y, et al. Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches[J]. Remote Sensing, 2019, 12(1):1–21.

[15] 杨北萍, 陈圣波, 于海洋, 等. 基于随机森林回归方法的水稻产量遥感估算[J]. 中国农业大学学报, 2020, 25(6):26–34.

YANG Beiping, CHEN Shengbo, YU Haiyang, et al. Remote sensing estimation of rice yield based on random forest regression method[J]. Journal of China Agricultural University, 2020, 25(6):26–34. (in Chinese)

[16] 王鹏新, 齐璇, 李俐, 等. 基于随机森林回归的玉米单产估测[J]. 农业机械学报, 2019, 50(7):237–245.

WANG Pengxin, QI Xuan, LI Li, et al. Estimation of maize yield based on random forest regression[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7):237–245. (in Chinese)

[17] 程千, 徐洪刚, 曹引波, 等. 基于无人机多时相植被指数的冬小麦产量估测[J]. 农业机械学报, 2021, 52(3):160–167.

CHENG Qian, XU Honggang, CAO Yinbo, et al. Grain yield prediction of winter wheat using multi-temporal UAV based on multispectral vegetation index[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3):160–167. (in Chinese)

[18] CAI Y, GUAN K, LOBELL D, et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches[J]. Agricultural and Forest Meteorology, 2019, 274(5):144–159.

[19] GUAN K, BERRY J A, ZHANG Y, et al. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence[J]. Global Change Biology, 2016, 22(2):716–726.

[20] GUANTER L, ZHANG Y, JUNG M, et al. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence[J]. Proceedings of the National Academy of Sciences, 2014, 111(14):1327–1333.

Transactions of the Chinese Society for Agricultural Machinery,2020,51(8):414–419. (in Chinese)

[18] 杜柳青,余永维. 基于序列深度学习的数控机床热误差建模与预测方法[J]. 机床与液压,2020,48(23):88–92.
DU Liuqing, YU Yongwei. Thermal error modeling and prediction for NC machine tools based on sequence deep learning[J]. Machine Tool and Hydraulics,2020,48(23):88–92. (in Chinese)

[19] ZUO Yunbo, WANG Hongjun, WU Guoxin, et al. Research on remote state monitoring and intelligent maintenance system of CNC machine tools[J]. The Journal of Engineering,2019(23):8671–8675.

[20] YI Z, YANG J, HUI J. Machine tool thermal error modeling and prediction by grey neural network[J]. The International Journal of Advanced Manufacturing Technology, 2012, 59(9):1065–1072.

[21] POERNOMO A, KANG D K. Biased Dropout and Crossmap Dropout: learning towards effective dropout regularization in convolutional neural network[J]. Neural Network, 2018,104:60–67.

[22] LI J, LUONG M T, JURAFSKY D. A hierarchical neural autoencoder for paragraphs and documents[J]. Computer Science, 2015(1):1106–1115.

[23] LIPTON Z C, BERKOWITZ J, ELKAN C. A critical review of recurrent neural networks for sequence learning[J]. Computer Science, 2015(1):104–112.

[24] RUBEN Z, ALICIA L D, JAVIER G D, et al. Language identification in short utterances using long short-term memory (LSTM) recurrent neural networks[J]. Plos One,2016,11(1):1–17.

[25] WEN Y, ZHANG K, LI Z, et al. A discriminative feature learning approach for deep face recognition [C] // European Conference on Computer Vision,2016:499–515.

(上接第 204 页)

[21] 林滢,邵怀勇. 基于随机森林算法的河南省冬小麦产量预测最佳时间窗和影响因子研究[J]. 麦类作物学报,2020,40(7):874–880.
LIN Ying, SHAO Huaiyong. Study on optimal time and influencing factors for winter wheat yield prediction in Henan based on random forest algorithm[J]. Journal of Triticeae Crops, 2020, 40(7):874–880. (in Chinese)

[22] CAO J, ZHANG Z, TAO F L, et al. Identifying the contributions of multi-source data for winter wheat yield prediction in China[J]. Remote Sensing, 2020, 12(5):750.

[23] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1):5–32.

[24] NICODEMUS K K. Letter to the editor: on the stability and ranking of predictors from random forest variable importance measures[J]. Briefings in Bioinformatics, 2011(4):369–373.

[25] 檀艳静. 1981—2016 年河南省冬小麦产量对气象因素变化的响应[J]. 气象与环境科学, 2018, 41(4):108–116.
TAN Yanjing. Response of winter wheat yield to the changes of meteorological factors in Henan Province during 1981—2016 [J]. Meteorological and Environmental Sciences, 2018, 41(4):108–116. (in Chinese)

[26] LIU L, GUAN L, LIU X. Directly estimating diurnal changes in GPP for C3 and C4 crops using far-red sun-induced chlorophyll fluorescence[J]. Agriculturaland Forest Meteorology, 2017, 232:1–9.

[27] CHEN X, MO X, ZHANG Y, et al. Drought detection and assessment with solar-induced chlorophyll fluorescence in summer maize growth period over North China Plain[J]. Ecological Indicators, 2019, 104(7):347–356.

[28] 周亮,慕号伟,马海蛟,等. 基于卷积神经网络的中国北方冬小麦遥感估产[J]. 农业工程学报, 2019, 35(15):119–128.
ZHOU Liang, MU Haowei, MA Haijiao, et al. Remote sensing estimation on yield of winter wheat in North China based on convolutional neural network[J]. Transactions of the CSAE, 2019, 35(15):119–128. (in Chinese)

[29] SUN J, DI L, SUN Z, et al. County-level soybean yield prediction using deep CNN–LSTM model[J]. Sensors, 2019, 19(20):4363.

[30] GUAN K, WU J, KIMBALL J S, et al. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields[J]. Remote Sensing of Environment, 2017, 199:333–349.

[31] 程琳琳,李玉虎,孙海元,等. 京津冀 MODIS 长时序增强型植被指数拟合重建方法适用性研究[J]. 农业工程学报, 2019, 35(11):148–158.
CHENG Linlin, LI Yuhu, SUN Haiyuan, et al. Applicability of fitting and reconstruction method of MODIS long-time enhanced vegetation index in Beijing–Tianjin–Hebei[J]. Transactions of the CSAE, 2019, 35(11):148–158. (in Chinese)

[32] DRUSCH M, MORENO J, BELLO U D, et al. The fluorescence explorer mission concept—ESA’s Earth Explorer 8[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017,55(3):1273–1284.