

关键时相长势—环境和景观特征对河北省县级尺度冬小麦单产估算精度影响分析

朱炯^{1,2}, 杜鑫¹, 李强子¹, 张源¹, 王红岩¹, 赵云聪^{1,2}

1. 中国科学院空天信息创新研究院 遥感卫星应用国家工程实验室, 北京 100101;

2. 中国科学院大学, 北京 100049

摘要: 区域尺度上精准、快速的作物单产估算可以有效地为国家粮食安全相关政策的制定提供数据支撑。本文针对县级估产时相和特征类型选择问题, 基于遥感、气象和统计等多源数据, 通过不同时相和特征要素之间的组合分析来探索其对于县级尺度冬小麦单产估算的影响。特征要素主要考虑作物长势、环境(水分和光温条件)和农田景观3个类型; 时相主要考虑由冬小麦生长过程NDVI(Normalized Difference Vegetation Index)曲线特征提取的5个关键时段(P1—P5)。利用不同时相与类型特征的组合与统计单产构建随机森林回归模型, 根据精度评价结果分析各组合的优劣。2014年—2017年的数据用来建模, 2018年数据用来验证。对于单时相, P2、P3、P4的表现明显好于P1和P5; 多时相的准确度明显优于单时相, 其中P2、P4的组合效果最佳。对于不同类型的特征要素, 作物长势特征参量对估产精度的影响最大, 而水分影响和光温条件等环境因子的加入对估产准确性并没有明显提升, 农田景观参数的加入能够有效提升估产的准确性。在最优组合的基础上, 剔除冗余变量优选出5个重要的指标因子(PROP、NDVI_P2、B2_P2、ED、B1_P4), 并建立单产估算模型获取2018年河北省冬小麦县级尺度单产。结果表明, 平均相对误差(MRE)仅为2.85%, 决定系数(R^2)为0.83, 均方根误差(RMSE)为253.25 kg/ha, 归一化均方根误差(NRMSE)为4.09%。研究结果为全国县级冬小麦单产估算提供了新的思路和方法参考。

关键词: 遥感, 估产, 冬小麦, 统计数据, NDVI, 随机森林, 河北省

引用格式: 朱炯, 杜鑫, 李强子, 张源, 王红岩, 赵云聪. 2022. 关键时相长势—环境和景观特征对河北省县级尺度冬小麦单产估算精度影响分析. 遥感学报, 26(7): 1354–1367

Zhu J, Du X, Li Q Z, Zhang Y, Wang H Y and Zhao Y C. 2022. Analysis of the influence of key-phase growth-environment-landscape features on the accuracy of county-level winter wheat yield estimation in Hebei Province. National Remote Sensing Bulletin, 26(7): 1354–1367 [DOI: 10.11834/jrs.20221019]

1 引言

及时准确地估计农作物产量可以为国家制定粮食政策、调控市场价格提供信息支撑, 对发展农村经济和对外粮食贸易具有重要意义。此外, 精准的产量估算还可以为保险和市场决策提供有用的信息 (Lobell, 2013; Lobell 等, 2015)。更早、更准确地掌握作物产量信息, 有利于在粮食贸易中占据主动, 获得更好的经济效益。

传统的作物估产方法主要基于农艺学和气象

模型。通过抽样调查的方式建立产量与主要农艺、气象因子之间的回归模型 (吴炳方, 2004)。这种方法虽然精度高, 但时间和人力成本较高, 而且容易受到外界因素的干扰, 在大区域作物产量估算时尤为明显。此外, 有研究利用各种统计指标和数学模型来进行统计产量估算 (赵稷伟 等, 2016)。此类方法精度较高, 但缺乏理论支持, 在产量波动较大的年份估算精度较差, 并且估产模型需输入大量当年的统计数据, 模型估算结果滞后性严重。

收稿日期: 2021-01-11; 修订日期: 2021-12-14

基金项目: 国家重点研发计划(编号: 2016YFD0300603-5); 高分辨率对地观测系统重大专项(编号: 20-Y30F10-9001-20122, 11-Y20A16-9001-17/18); 广西科技重大专项(编号: 桂科AA18118048-2); 国家自然科学基金(编号: 41701486)

第一作者简介: 朱炯, 研究方向为农业遥感估产。E-mail: zhujiong@aircas.ac.cn

通信作者简介: 李强子, 研究方向为农业遥感。E-mail: liqz@aircas.ac.cn

1972年,美国成功发射了第一颗地球资源卫星以来,遥感技术发展迅速,并广泛应用于农情监测的各个领域,例如农作物分布图制作,农业灾害监测和作物长势评估及产量估算等(Doraiswamy, 2004; Nilsson, 1995; Pachavo和Murwira, 2014; Tatsumi等, 2015; West等, 2019; Yang等, 2015)。遥感技术以其宏观性、客观性和及时性等优势,在作物估产中显示出巨大的潜力。自全球小麦估产项目LACIE (Large Area Crop Inventory Experiment) (Erickson, 1984)启动以来,遥感估产技术经过近40年的探索,无论是从数据质量和多样性还是从机理模型研究上都取得了长足的进步。早期的研究探索建立原始光谱特征或植被指数和产量之间的统计回归模型(Rasmussen, 1992; Tennakoon等, 1992; Thenkabail等, 1994; Wiegand等, 1989)。研究结果表明产量与作物生长旺季的NDVI (Normalized Difference Vegetation Index)具有明显的相关性(Lewis等, 1998; Maselli等, 2000),但NDVI的饱和特性会导致产量估算结果存在偏差。此外,此类经验模型的时空泛化能力较差,在不同区域或年份应用时往往需要重新构建模型。事实上,遥感数据直接反映了地表植被的综合状况,而不仅是与果实(产量)有关的器官部分。因此,首先估算作物干物质含量,然后根据其与作物果实部分之间的比例关系(收获指数)来计算产量成为可能。基于该思想,有研究利用遥感数据进行光合有效辐射吸收比FAPAR (the Fraction of Absorbed Photosynthetically Active Radiation)估算,并将其纳入作物光能利用效率LUE (Light Use Efficiency)模型,用于产量估算(Lobell等, 2003; Xin等, 2013; 纪甫江等, 2020; 刘真真等, 2017)。此类方法具有作物生理学基础,所需参数较少,估算结果精度高,得到了广泛的应用和发展。然而,作物产量的形成是土壤、气候、作物品种、田间管理等多种因素综合作用的结果,综合考量上述因素对估算作物产量至关重要。作物生长模型能够更准确地描述作物生长发育在作物—土壤—大气系统中的动态过程。随着作物系统建模技术的日趋成熟(Holzworth等, 2015),综合遥感数据的空间连续性与作物生长模型的时间连续性进行作物产量估算是当前研究的热点。数据同化是当前的主流方法,它使用基于遥感数据计算所得的同化变量来

校准、强制和更新作物生长模型,以获得更真实的产量结果(Dorigo等, 2007; Jin等, 2018; 程志强和蒙继华, 2015; 黄健熙等, 2015)。目前主要的同化算法有四维变分同化4D-Var (Four-Dimensional Variational Assimilation)、集合卡尔曼滤波EnKF (Ensemble Kalman Filter)、粒子滤波PF (Particle Filter)等(De Wit和Van Diepen, 2007; Ines等, 2013; 姜志伟等, 2012; 王鹏新等, 2016)。除此之外,有研究利用作物生长模型对区域内作物单产形成过程进行模拟,生成大量伪数据。利用大量的伪数据,训练得到不同时相组合的单产估算模型,并将遥感数据输入模型中进行区域单产估算。单产估算结果与地面调查结果的相关性在0.38—0.76 (Lobell等, 2015),表明了该方法具有可行性,但目前有关研究较少,需要进一步挖掘其应用价值。

虽然利用作物生长模型或者光能利用率模型进行估产机理性更强,但在进行县市级的范围估产应用时面临许多困难,如作物生长模型本地化困难、输入参数过多,LUE模型的实际光能利用率、收获指数难以确定等。此外这些模型常用于像元或者地块尺度,在大区域估算时效率较低且验证困难,同时需要精细尺度的作物分布图支持。鉴于上述问题,对于县市级的单产估算,相比于复杂的机理模型,抽样统计调查和基于遥感气象等多源数据的经验模型其可靠性与可操作性更好,仍然是当前或未来一定时间内的主流方案。产量是作物品种特性、生长环境以及农田管理等因素共同作用的结果,在大区域范围,难以对所有因素进行定量观测,现有技术手段能够获取的信息主要包括光温(如日照时数、平均气温等),水分(如降水量、土壤湿度等)和作物的分布以及生长状态等(如冠层高度、叶面积指数、植被指数等)。许多研究已经利用统计单产与光谱反射率、植被指数或温度,降水等参数进行了回归分析,取得了不错的成效(Balaghi等, 2008; Kogan等, 2013; 朱婉雪等, 2018),但在多源特征筛选与分析方面还没有进行系统的研究。此外,有研究表明农作物生长过程中农田破碎程度和农业集约度也会对产量形成造成影响(葛玉娟等, 2020)。在县级尺度上,可以计算出作物的景观格局指标来表征这些信息,现有的研究缺乏对该指标的利用。同时,目前研究较少从产量形成过程

出发综合系统地考虑时间序列特征对作物估产的影响,在县级尺度上,研究大多采用优选单一时段进行模型估算(Bolton和Friedl,2013;任建强等,2015;王飞龙等,2020),虽然该方式简单易行,但丢失的其他时间特征可能会降低估产的精度。

本研究将现有易于获取的农情特征分为作物长势特征、环境特征、农田景观特征3类,综合评估了各类别特征以及产量形成的关键时相对冬小麦产量估算的影响,旨在创建一个基于遥感和气象数据的作物单产估算模型,实现县级尺度冬小麦产量高精度估算。研究主要围绕以下3个问题:(1)分析作物估产的关键时期;(2)评估环境特征的对作物估产的影响;(3)分析农田景观特征对作物估产的促进作用。最终在最优时相和特征的基础上进行变量筛选分析并构建最佳模型,实现县级尺度冬小麦单产的估算。

2 实验区与数据

2.1 实验区概况

本研究在河北省开展实验,河北省冬小麦种植区域集中于该省东南部的平原地区,全省平原面积为81459 km²,占总面积的43%。实验区地势平缓,海拔均不到100 m。河北省是中国主要的粮食产区(全国3大小麦集中产区之一),是中国重要的传统农业大省。其种植制度以冬小麦—夏玉米轮作体系为主,省内冬小麦播种面积以及产量分别占全国的9.87%和10.87%(张敬宇等,2015),在保障全国粮食产量安全中具有重要的地位。河北省东南部平原属于暖温带季风气候,四季分明。年平均气温和年降水量由南向北随纬度增加而递减,年平均气温在12—13℃,全年无霜期为200—210 d。年日照时数为1900—2300 h,多年平均降水量为490—550 mm。降水年内分布不均,年内降水量主要集中在6—9月,超过全年降水量的70%(赵晗等,2020)。冬小麦主要生育期可概括为10月中下旬播种、11月上旬进入分蘖期、11月下旬—12月上中旬进入越冬期、次年2月下旬—3月上旬进入返青期、3月中旬—4月中下旬进入起身拔节期、4月下旬—5月上旬进入孕穗抽穗期、5月上中旬开花、5月中下旬灌浆、5月底—6月上旬成熟。

本实验研究对象为县级尺度,搜集实验区内2014年—2018年连续种植冬小麦的县市,并考虑

各县内遥感与气象等数据的获取情况,确保县内有时间连续的数据,最终选取93个县市作为本文的研究对象。实验区具体情况如图1所示。

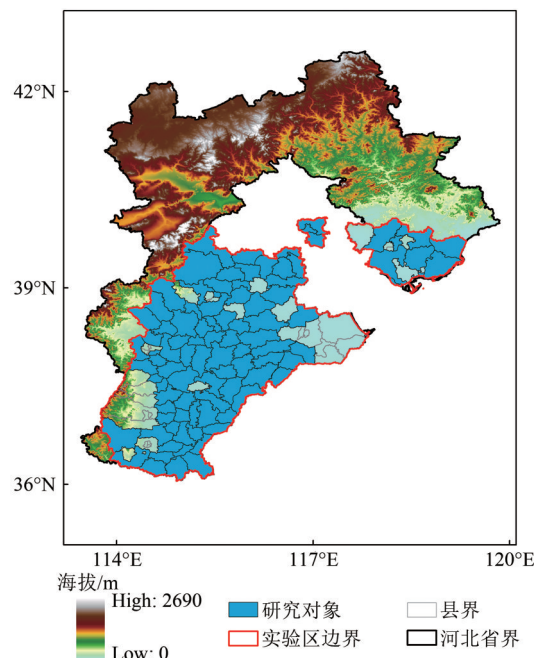


图1 实验区概况

Fig. 1 Overview of the study area

2.2 遥感数据

针对县级尺度的研究,本文选择中低分辨率的MODIS陆地标准产品。数据时间范围覆盖2014年—2018年河北冬小麦从返青期到成熟期的生长过程,即当年第49天开始至第177天。为充分探究不同类型特征与产量的关系,本文选择6个数据集共16个特征参与估产,包括归一化植被指数(NDVI)、增强型植被指数EVI(Enhanced Vegetation Index)、白天地表温度LST_DAY(Day-time Land Surface Temperature)、夜晚地表温度LST_NIGHT(Night-time Land Surface Temperature)、潜在蒸散PET(Potential Evapotranspiration)、实际蒸散ET(Evapotranspiration)、叶面积指数LAI(Leaf Area Index)、光合辐射有效吸收比(FPAR)、总初级生产力GPP(Gross Primary Productivity)以及地表反射率SR(Surface Reflectance),其中地表反射率共7个波段,详细的特征要素信息见表1。此外,以MOD13A2数据集为标准,将其他数据集重采样到1KM,并生成16 d合成产品。最后利用冬小麦分布数据进行掩膜。遥感数据按照各县内有效像元均值来代表县内整体水平。

表1 MODIS产品要素信息
Table 1 MODIS product information

数据集	变量类型	空间分辨率/m	时间分辨率/d	单位
MOD13A2.006	NDVI,EVI	1000	16	—
MOD11A2.006	LST_DAY,LST_NIGHT	1000	8	Kelvin
MOD16A2.006	ET,PET	500	8	kg/m ² /8 day
MOD15A2H.006	LAI,FPAR	500	8	m ² /m ² , Percent
MOD17A2H.006	GPP	500	8	kg C/m ²
MOD09A1.006	Surface Reflectance,SR	500	8	—
	Band 1 (620—670 nm),B1			—
	Band 2 (841—876 nm),B2			—
	Band 3 (459—479 nm),B3			—
	Band 4 (545—565 nm),B4			—
	Band 5 (1230—1250 nm),B5			—
	Band 6 (1628—1652 nm),B6			—
	Band 7 (2105—2155 nm),B7			—

2.3 气象数据

本文使用的日值气象数据包括降水量、日照时数、平均气温、最高气温、最低气温、平均水汽压共6个气象要素（国家气象科学数据中心，<http://data.cma.cn/>[2021-01-11]）。数据时间范围覆盖2014年—2018年每年第49天开始至第193天。为准确表征每个县的具体气象情况，直接使用县内气象站点数据来表征各县的气象水平。当县内存在多个站点时，选择距离冬小麦主产区最近的气象站点。以MOD13A2时间为标准对气象数据进行16 d合成，其中降水量取累积值，其他参量取平均值。

2.4 统计数据

本文所用统计数据来源于中国县域统计年鉴（县市卷）（国家统计局农村社会经济调查司，2020），包括冬小麦单产，单位为kg/ha；冬小麦播种面积，单位为ha；耕地面积，单位为ha。其中2014年—2017年数据用来训练，2018年数据用来验证。

2.5 冬小麦分布图

2014年—2018年的冬小麦布图以每年冬小麦播种期、生长期的Landsat卫星图像为主要影像数据源，根据河北省冬小麦的物候和光谱特征，采

用随机森林算法对冬小麦空间分布进行提取。参考Google Earth高分辨率卫星影像，目视选取训练样本与验证样本。最终获得30 m冬小麦空间分布结果（图2），总体精度均在90%以上。在此基础上，筛选出冬小麦种植面积超过80%的公里网格数据用来对MODIS产品进行掩膜。

2.6 农田景观数据

本文所使用的农田景观特征具体包括：冬小麦的边缘密度ED（Edge Density）、冬小麦播种面积（AREA）和冬小麦占耕地面积比（PROP）。其中边缘密度用来表征目标区域冬小麦田块分布的破碎程度，单位为m/ha，具体计算如式（1）所示；冬小麦占耕地面积比利用统计数据中冬小麦播种面积除以耕地面积计算得到，用来表征冬小麦种植规模。

$$ED = \frac{\sum_{k=1}^m e_{ik}}{A} (10000) \tag{1}$$

式中， e_{ik} 是斑块类型为*i*的景观的边缘总长度，单位为m；*A*是景观总面积，单位为m²；乘以10000转换为ha。

3 研究方法

本研究基于遥感数据、气象数据和统计数据，

搜集能表征作物农情信息的数据，并将其划分为作物长势特征、环境特征（包括水分特征与光谱特征）以及农田景观特征3个类型（表2）。同时利用NDVI曲线划分出5个关键时相（P1—P5），然后综合分析不同时相不同类型的特征参量进行单产估算建模，通过对预测结果的精度分析来探讨不同类型和时相特征对县级估产的影响。最终，选出最佳的特征组合，并在此基础上进行特征优选，剔除冗余变量，构建最优模型进行县级冬小麦单产估算。具体技术流程如图3所示。

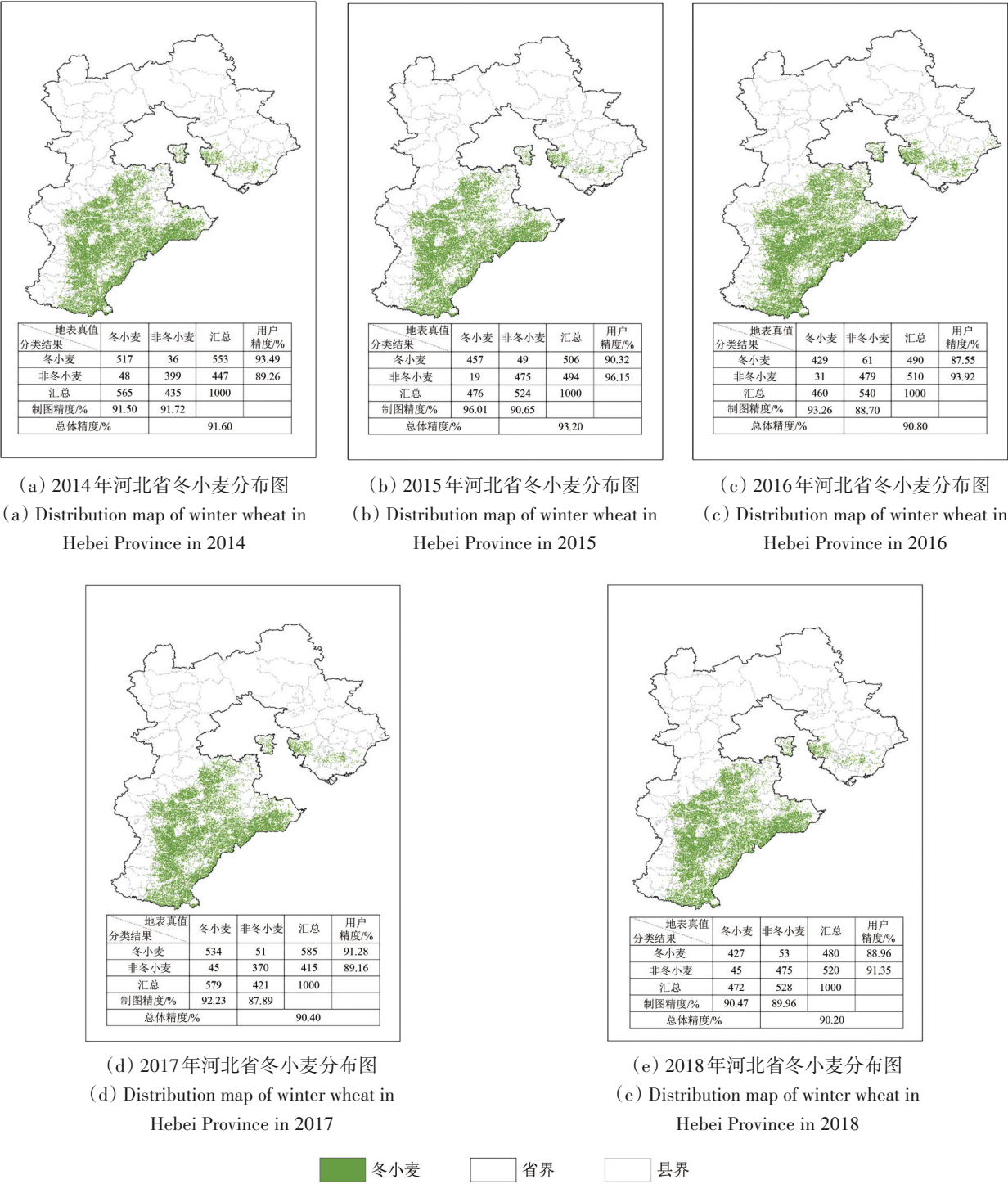


图2 河北省冬小麦种植分布图
Fig. 2 Distribution map of wheat in Hebei Province

表2 变量类型划分
Table 2 Classification of independent variables

变量类型		数据
作物长势特征(C)		NDVI, EVI, LAI, FPAR, GPP, SR (Band1—Band7)
环境特征 (ENV)	水分参量(M)	ET, PET, 降水量, 平均水汽压
	光温参量(T)	LST_DAY, LST_NIGHT, 日照时数, 平均气温, 最高气温, 最低气温
农田景观特征(LAND)		ED, AREA, PROP

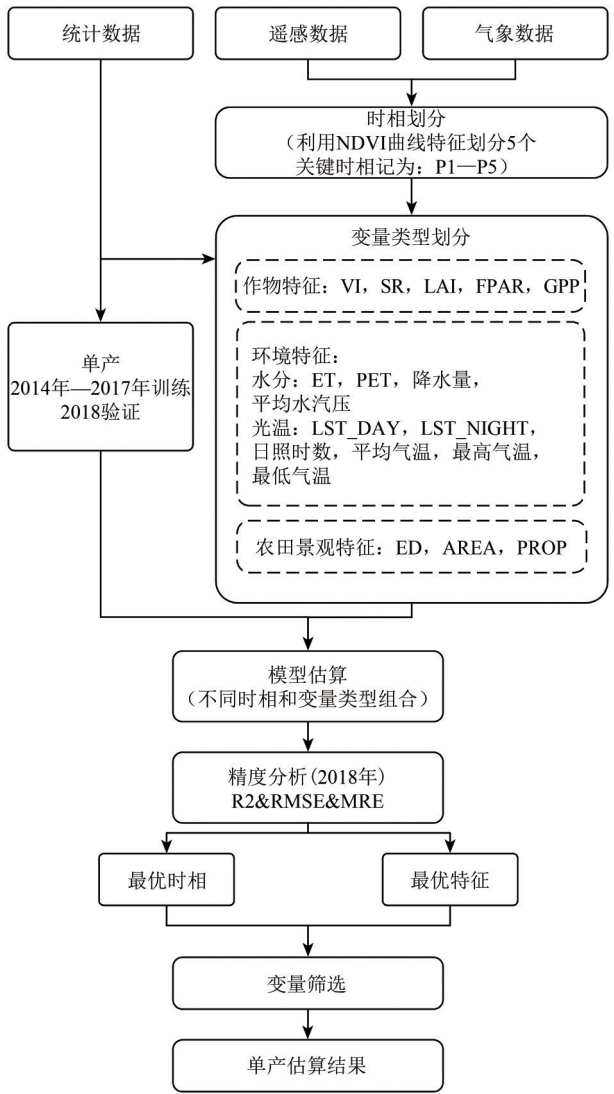


图3 县级产量估算的方法概述
Fig. 3 Overview of methods for estimating county-level wheat yield

3.1 时相划分

为了探究时相特征对作物估产的影响, 获取各县不同年份冬小麦的准确生育期是研究所必需的。在大范围生育期观测数据难以准确获取的情况下, NDVI曲线特征由于能够很好的反应冬小麦

生长状况, 已经被广泛用于划分冬小麦的生育期(李加林和刘闯, 2005; 钱永兰等, 2012; 权文婷等, 2015)。为了避免播种期时间差异的影响, 本文通过各县冬小麦生长季 MODIS 16 d 合成的 NDVI 的曲线特征提取出 5 个关键时相(图4), 记为 P1—P5(式(1))。该曲线表示 2018 年河北省景县的冬小麦 NDVI 变化, 时间范围从 2018-02-18 到 2018-06-26。P1—P5 依次为 2018 年第 81, 97, 113, 129, 145 天。这 5 个关键生长时段涵盖了冬小麦从返青到成熟的生长过程, 也是对产量影响最关键的时段(徐建文等, 2014)。其中 P3 为冬小麦生长旺盛期, 在 NDVI 曲线上表现为峰值点。P1、P2 分别是峰值点向左两个点位, P4、P5 分别是峰值点向右两个点位。

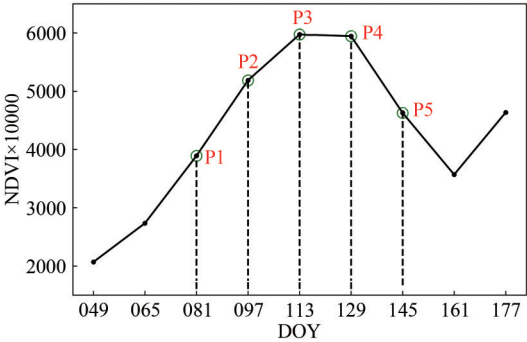


图4 冬小麦关键生长时段划分示意图
Fig. 4 Division of key growth periods of winter wheat

$$P_k = i + 16 \cdot (k - 3) \tag{2}$$

式中, $k=1, 2, 3, 4, 5$ 。 i 是县级 NDVI 曲线在第 65 天至第 145 天内 NDVI 最大值所对应的 DOY。

3.2 试验设计

本文针对研究目标的 3 个问题, 分别设计试验, 其中涉及的特征组合信息见表 3。

试验(a) 基于作物长势特征和环境特征筛选分析冬小麦单产估算的关键时相组合。首先分析了单个时相的估产精度, 在此基础上对双时相、三时相以及全部时相组合进行精度分析来探讨最优时相组合。

试验(b) 在作物长势特征的基础上, 逐步加入水分(M)、光温(T)和全部环境特征(ENV), 研究不同环境特征对估产精度的影响程度。

试验(c) 进一步加入边缘密度(ED)、冬小麦播种面积(AREA)、冬小麦占耕地面积比(PROP)等农田景观特征, 深入分析不同农田景观特征加入对估产精度的影响。

表3 不同特征组合

Table 3 Data combination of random forest input

试验	组合	数据
(a)	P1	P1时相的作物长势特征和环境特征
	P2	P2时相的作物长势特征和环境特征
	P3	P3时相的作物长势特征和环境特征
	P4	P4时相的作物长势特征和环境特征
	P5	P5时相的作物长势特征和环境特征
	P23	P2,P3时相的作物长势特征和环境特征
	P24	P2,P4时相的作物长势特征和环境特征
	P34	P3,P4时相的作物长势特征和环境特征
	P234	P2,P3,P4时相的作物长势特征和环境特征
	C_ENV	所有时相上作物长势特征和环境特征
(b)	C	所有时相上作物长势特征
	C_M	所有时相上作物长势特征和水分特征
	C_T	所有时相上作物长势特征和光温特征
	C_ENV	所有时相上作物长势特征和环境特征
(c)	C	所有时相上作物长势特征
	C_ED	所有时相上作物长势特征和边缘密度
	C_AREA	所有时相上作物长势特征和冬小麦播种面积
	C_PROP	所有时相上作物长势特征和冬小麦占耕地面积比
	C_LAND	所有时相上作物长势特征和农田景观特征

本研究基于不同类型特征以及时相的组合构建随机森林回归模型,实现县级尺度的单产估算。随机森林(Breiman, 2001)是一种有监督的集成学习算法,在分类或回归问题上被广泛使用,是目前最有效的模型之一。随机森林回归算法的学习和预测过程可以简述为以下3步:

(1) 样本集的选择。在样本集合:

$$D = \{x_{i1}, x_{i2}, \dots, x_{im}, y_i\} (i \in [1, N])$$

中利用自助法(Bootstrap算法)重复进行有放回的随机抽样,抽取 K 组子样本集作为训练样本,用来构建 K 棵分类回归树CART(Classification And Regression Tree)。

(2) CART的生成。在每一轮生成CART的过程中,从 M 个特征中随机选择其中的 m 个特征($m < M$)组成一个新的特征集:

$$d_j = \{x_{i1}, x_{i2}, \dots, x_{im}, y_i\} (i \in [1, n])$$

通过使用新的特征集来生成CART。使用均方差MSE(Mean Squared Error, 式(3))作为节点分裂准则来确定最优分类特征,递归地进行每个节点的分支过程,最终生成 K 个CART组成森林。

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (3)$$

式中, n 是样本个数, f_i 是CART输出结果, y_i 是真实值。

(3) 模型的预测。由于生成的 K 个CART之间是相互独立的,其重要性是相等的,因此无需考虑它们的权值,使用所有单株CART输出的平均结果来作为最终的预测结果。(Everingham等, 2016)。

随机森林回归算法的主要优势有以下几点:所有子树以并行计算的方式运行,速度较快;可以处理大量特征,并给出特征重要性排序结果;有较强的抗噪能力,对变量间共线性不敏感;可以利用袋外样本对模型自身进行精度评价,无需进行交叉验证或利用单独的测试样本进行验证。先前的研究已经使用随机森林估算了包括小麦、玉米和马铃薯(Jeong等, 2016; 刘峻明等, 2019; 王鹏新等, 2019)在内的多种农作物的产量。

3.3 精度验证

本研究使用决定系数 R^2 (Coefficient of Determination),均方根误差RMSE(Root Mean Squared Error),平均相对误差MRE(Mean Relative Error)和归一化均方根误差NRMSE(Normalized Root Mean Squared Error)4个指标来评价每个组合的模型预测精度,从而分析各特征组合对估产的影响,计算公式如式(4)~式(7)。

$$R^2 = 1 - \frac{\sum_i (E_i - O_i)^2}{\sum_i (E_i - \bar{O}_i)^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (E_i - O_i)^2}{n}} \quad (5)$$

$$MRE = \sum_{i=1}^n \frac{|E_i - O_i|}{n \times O_i} \times 100\% \quad (6)$$

$$NRMSE = \frac{RMSE}{\bar{O}_i} \times 100\% \quad (7)$$

式中, E_i 表示第 i 个县的冬小麦单产估算值, O_i 表示测试数据集中第 i 个县的冬小麦单产的统计值, $\bar{O}_i$ 表示单产统计值的平均值, n 表示样本数量。

4 结果分析

4.1 县级冬小麦单产估算的关键时相筛选分析

从不同时相组合的分析结果(图5试验(a)),

表4中试验(a))来看,对于县级尺度冬小麦单产的估算,P1—P5贡献度不同,但P2、P3、P4时相的估产精度要明显高于P1、P5两个时相,而P2、P3、P4之间在精度上没有明显差异。其中P1、P5时相的 R^2 均小于0.5,且RMSE在410 kg/ha以上,NRSME大于6.5%;P2、P3、P4的 R^2 相近,在0.66—0.68,RMSE在335—345 kg/ha,NRMSE介于5.4%—5.5%。以河北省冬小麦物候期来看,P2、P3、P4主要处在起身拔节期到抽穗灌浆期,这与王瑞峥等(王瑞峥等,2018)的分析结果一致。

多时相的组合结果精度均高于单个时相的结果,其中P23的结果提升最小,P34结果有明显提升,P24的结果提升最大且与P234和所有时相参与的结果(C_ENV)无明显差异, R^2 达到0.7以上,且RMSE小于300 kg/ha,MRE为3.26%,NRMSE等于4.74%。综上,P24的时相组合相对最优,再增加其他时相特征对估产精度提升的帮助不明显。

4.2 环境特征对县级冬小麦单产估算影响分析

在作物长势特征基础上加入环境特征之后发现,基于遥感或气象数据获取的环境特征参量,对于县级冬小麦单产估算精度提升的贡献非常有限。从图5试验(b)和表4试验(b)的结果可以看出,仅利用作物长势特征参量进行估产, R^2 达到0.7以上且RMSE小于300 kg/ha。环境特征的加入对估产结果精度提升帮助不大,其中光温特征几乎不提供额外的精度提升,而水分特征的加入仅使RMSE减少了7 kg/ha。遥感获取的作物长势特征在一定程度上包含了环境影响的综合表现,所以气象环境特征对于本研究的估产建模所提供的更多是冗余信息。

4.3 农田景观特征对县级冬小麦单产估算影响分析

从图5试验(c)和表4试验(c)的结果可以看出,农田景观特征的加入对提升县级冬小麦估产精度效果明显,在加入3种景观特征(C LAND)后 R^2 达到0.84,RMSE为247.79 kg/ha,平均相对误差仅为2.74%,NRMSE为4%。作用最大的是PROP参数,与仅用作物长势特征结果(C)相比RMSE减少了34 kg/ha,且 R^2 达到0.82;其次

是ED参数,使RMSE减少了21 kg/ha, R^2 达到0.79;相对而言,AREA参数的加入对估产精度影响不太明显。

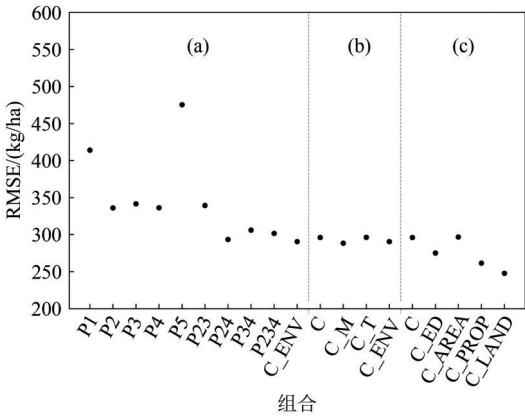


图5 不同组合随机森林验证结果
Fig. 5 Validation results of random forest with different combinations

表4 随机森林分析结果
Table 4 Random forest analysis results

试验	组合	R^2	RMSE/(kg/ha)	NRMSE/%	MRE/%
(a)	P1	0.44	413.92	6.68	5.52
	P2	0.68	336.01	5.42	4.15
	P3	0.66	341.51	5.51	4.24
	P4	0.68	336.32	5.43	4.15
	P5	0.3	475.19	7.67	6.56
	P23	0.67	339.27	5.48	4.2
	P24	0.74	293.43	4.74	3.26
	P34	0.71	305.95	4.94	3.31
	P234	0.72	301.64	4.87	3.24
	C_ENV	0.74	290.44	4.69	3.28
(b)	C	0.73	295.95	4.78	3.25
	C_M	0.75	288.21	4.65	3.24
	C_T	0.73	296.18	4.78	3.29
	C_ENV	0.74	290.44	4.69	3.28
(c)	C	0.73	295.95	4.78	3.25
	C_ED	0.79	274.99	4.44	3.05
	C_AREA	0.76	296.8	4.79	3.21
	C_PROP	0.82	261.33	4.22	2.8
	C LAND	0.84	247.79	4	2.74

4.4 基于最优特征组合的县级冬小麦单产估算结果

图5和表4显示了P24时相组合以及作物长势特征和农田景观特征对县级冬小麦单产估算最为

重要, 考虑到高度相关的数据可能会导致变量重要性评估不准确, 数据冗余也会降低模型的运算效率, 本研究进一步对P24时期的作物长势特征参量和农田景观特征参量进行优选分析。首先, 基于相关性分析, 将相关系数大于0.8的两个变量视为有强共线性, 并剔除一个(图6)。最终剩余13个特征变量, 分别为ED、AREA、PROP、NDVI_P2、LAI_P2、B1_P2、B2_P2、B3_P2、B5_P2、B6_P2、

B1_P4、B2_P4、B5_P4。然后, 对这13个变量利用随机森林重要性排序由高到低逐个累加进行结果分析。实验结果表明: 在第5个变量加入之后, 估产模型的准确性趋于稳定(图7)。因此, 利用前5个变量(PROP、NDVI_P2、B2_P2、ED、B1_P4)基于随机森林回归进行估产建模, 最终估算结果(图8)的 R^2 达到0.83, MRE为2.85%, RMSE为253.25 kg/ha, NRMSE为4.09%。

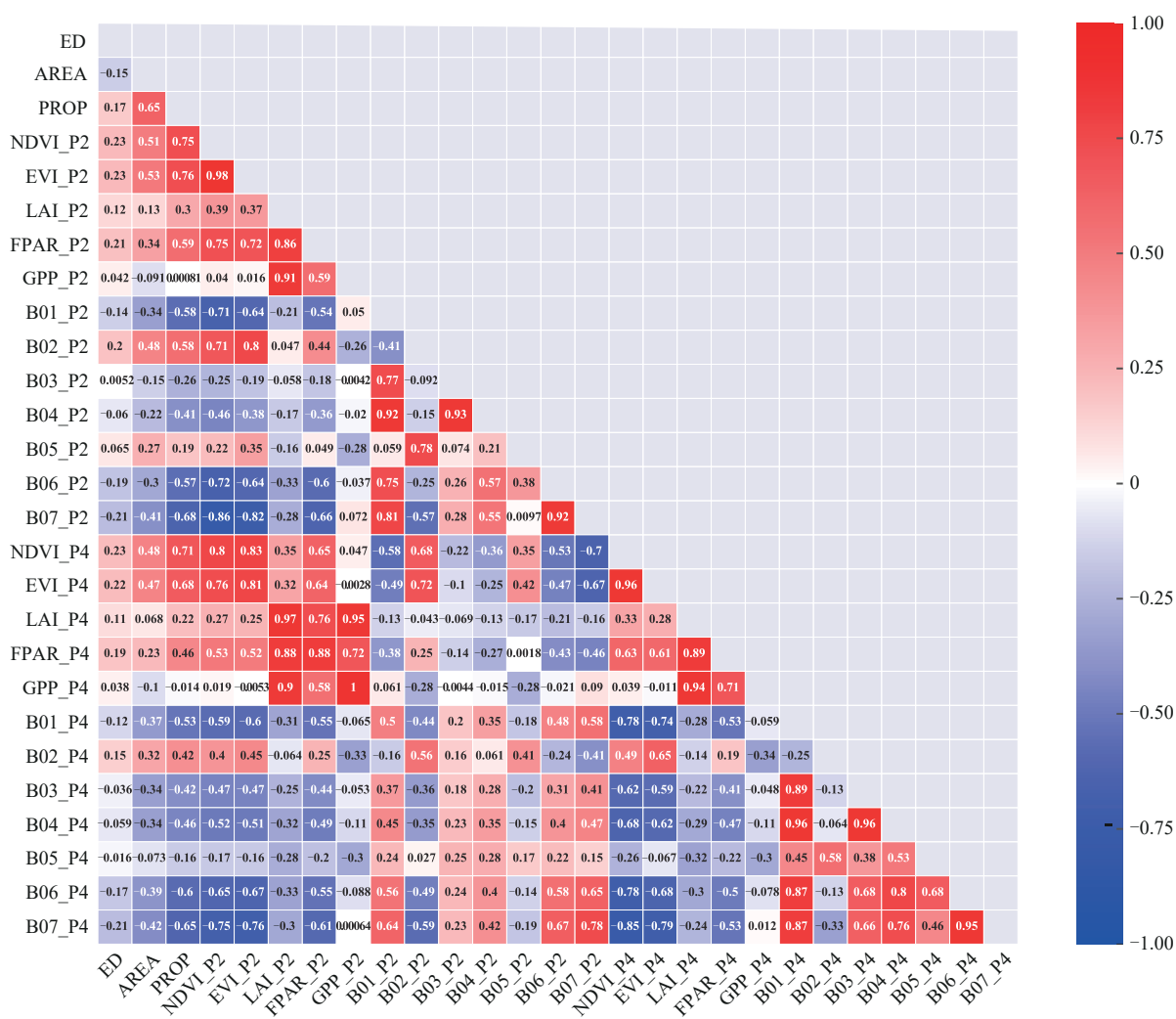


图6 相关系数矩阵热力图

Fig. 6 Heat map of correlation coefficient matrix

由2018年河北省县级尺度冬小麦单产估算结果及相对误差分布图(图9)可以看出, 河北省冬小麦单产中南部较高(6700 kg/ha左右), 而东北部较低(6000 kg/ha以下)。大部分县的估产精度高于0.95, 共80个, 估产精度高于0.9的共90个县, 仅有3个县的相对误差绝对值大于10%, 最低估产精度为0.86。这表明本文所构建的县级尺度冬小麦单产估算模型精度较高, 能够满足统计部

门需求, 具备为地方政府相关政策决策的制订提供数据支撑的能力。经统计, 有35个县的相对误差大于0, 而58个县的相对误差为负值, 这表明模型对冬小麦产量估算出现了一定程度的高估, 且模型在高产区表现优于低产区, 这可能与样本分布不均有关, 后续研究需要进一步分析低产区误差成因, 改进模型在低产区的表现。

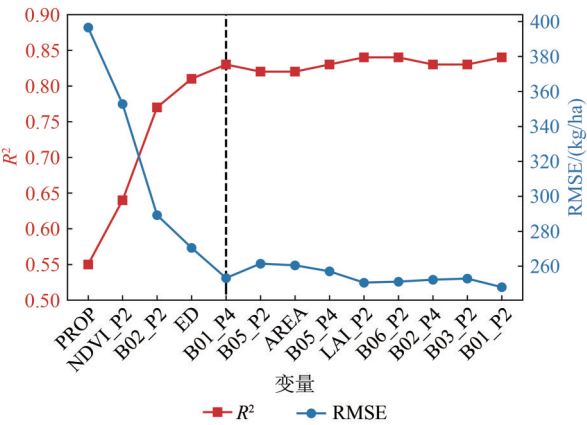


图7 随机森林变量累加分析结果图
Fig. 7 Cumulative analysis results of random forest variables

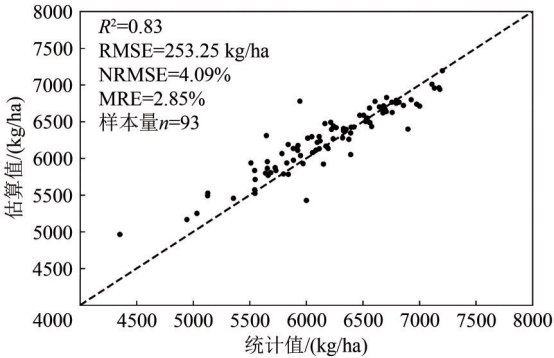


图8 最优模型在验证数据集的观测值和估计产量之间的线性回归
Fig. 8 Linear regression between observed and estimated yield for the validation data set from the optimal model

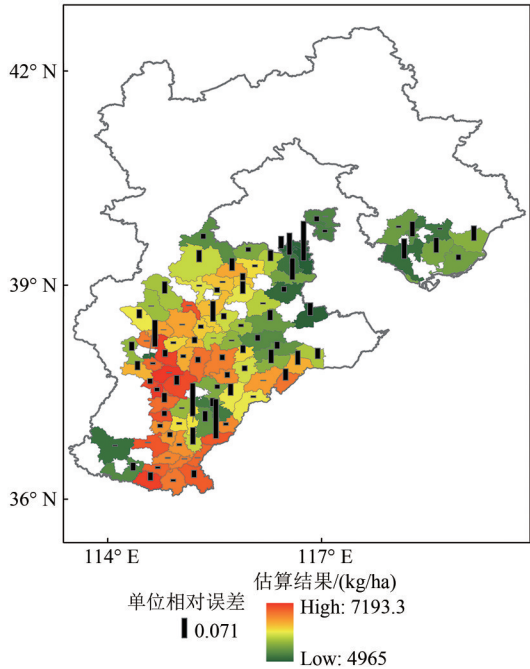


图9 2018年河北省县级冬小麦单产估算结果及相对误差分布图
Fig. 9 Estimated yield and relative error map of winter wheat at county-level in Hebei province in 2018

5 讨论

(1) 关键时相/生育期对于农作物单产估算的影响是显而易见的，不同区域农作物的生育期受气候地形、农田管理措施等因素的综合影响，存在明显的差异。因此，直接用某个特定时间或时间段数据来估算不同区域的农作物单产在实际应用中往往是难以满足要求的。本研究通过冬小麦NDVI过程曲线快速提取关键生育期时相，并在此基础上，确定了冬小麦单产形成过程中的关键生育期时相：拔节期和抽穗期，即P2、P4时相。灌浆—成熟期特征（P5）在最优模型中并未体现，其可能原因是：河北省冬小麦在灌浆—成熟阶段的冠层光谱变化较小，相关信息难以被遥感数据捕捉；本研究采用的16 d合成的数据产品不利于对冬小麦灌浆—成熟期的快速生长变化信息进行准确获取。所以，后续对于冬小麦单产精准估算需要更加深入的开展，应该重点开展更高时频的遥感数据的引入，或者多源遥感数据协同、时间重建等方面的研究。

(2) 本研究对比分析了作物长势和环境特征变量对于县级冬小麦单产估算精度的影响，并从中发现，相对于像元、地块等精细尺度的农作物单产估算，对于县级尺度以上的农作物单产估算而言，直接反映农作物生长状态的长势特征变量明显体现了更大的优势，其主要原因可能在于两个方面：遥感监测获取的全覆盖大范围的平均状况相对于单个气象站点代表的区域状况更加能够准确反映目标区域的宏观状态；遥感监测获取的农作物生长状态在一定程度上也包含了气象环境条件的影响，即其本身也是一种农作物对于现实气象环境条件的响应特征。

(3) 本研究在一定程度上表明了农田景观特征对于农作物单产估算的潜力，在作物集中连片种植区单产水平较高且估产精度更好。由于数据获取原因，本文只选择了3个较为直接的参数来分析农田景观特征对于县级尺度农作物单产估算的影响，以提升模型的区域泛化能力。鉴于本研究实验区域内冬小麦种植相对集中连片，对于农田地块相对破碎区域的农作物单产估算中农田景观特征的影响作用后续需要进一步深入分析。

6 结论

本文在中低分辨率遥感影像、气象资料和县

级统计数据的支持下,探究不同时相和特征的组合对河北省县级冬小麦单产估算的影响,并基于最优特征组合构建随机森林模型实现了河北省冬小麦主产区产量水平的快速准确估算。具体研究结论如下:

(1) 针对冬小麦区域尺度单产估算传统方法中的时相固定或单一问题,本文对各县基于MODIS 16 d合成NDVI过程曲线提取了冬小麦产量形成过程中5个关键时相(记为P1—P5),排除了各县生育期时间不一致的影响。结果表明,P2、P4组合的效果最佳,与前人研究中NDVI上升期、峰值期和冬小麦拔节期到灌浆期表现出良好的一致性。

(2) 针对传统冬小麦单产估算经验模型中的特征要素选择不系统问题,本文通过将现有农情数据划分为作物长势因素、环境因素、农田景观因素,系统的探究了各类别因素对冬小麦县级单产估算的影响,发现与单独使用作物长势因素(RMSE 295.95 kg/ha)相比,将作物长势因素与农田景观因素相结合(RMSE 247.79 kg/ha)可提供更准确的产量估计。这表明除了传统的NDVI指数外,农田景观特征参量对于估产精度的影响也不容忽视,其中,冬小麦占耕地面积比(PROP)和边缘密度(ED)尤其需要重视。

下一步研究内容考虑在现有的3个农田景观指标基础上进行扩充,充分挖掘农田景观特征在区域冬小麦单产估算的潜力。同时细化时相,更好的对应冬小麦物候节律。此外还应向其他与河北省地理、气候、农业种植条件差别较大的区域进行拓展实验,如在干旱半干旱地区进一步分析验证环境因素的影响,在云贵高原等地块破碎地区进一步分析农田景观因素的影响,探究本研究方法的区域适用性。

参考文献(References)

Balaghi R, Tychon B, Eerens H and Jlibene M. 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, 10(4): 438-452 [DOI: 10.1016/j.jag.2006.12.001]

Bolton D K and Friedl M A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173: 74-84 [DOI: 10.1016/j.

agrformet.2013.01.007]

Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5-32 [DOI: 10.1023/A:1010933404324]

Cheng Z Q and Meng J H. 2015. Research advances and perspectives on crop yield estimation models. *Chinese Journal of Eco-Agriculture*, 23(4): 402-415 (程志强, 蒙继华. 2015. 作物单产估算模型研究进展与展望. *中国生态农业学报*, 23(4): 402-415) [DOI: 10.13930/j.cnki.cjea.141218]

De Wit A J W and Van Diepen C A. 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1/2): 38-56 [DOI: 10.1016/j.agrformet.2007.05.004]

Doraiswamy P. 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment*, 92(4): 548-559 [DOI: 10.1016/j.rse.2004.05.017]

Dorigo W A, Zurita-Milla R, De Wit A J W, Brazile J, Singh R and Schaepman M E. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2): 165-193 [DOI: 10.1016/j.jag.2006.05.003]

Erickson J. 1984. The LACIE experiment in satellite aided monitoring of global crop production//Woodwell G M, ed. *SCOPE23: The Role of Terrestrial Vegetation in the Global Carbon Cycle*. New York: John Wiley and Sons

Everingham Y, Sexton J, Skocaj D and Inman-Bamber G. 2016. Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, 36(2): 27 [DOI: 10.1007/s13593-016-0364-z]

Ge Y J, Zhao Y L and Ren H Y. 2020. The effect of farmland fragmentation on the intensity of farmland in different utilization ways. *Advances in Earth Science*, 35(2): 180-188 (葛玉娟, 赵宇鸾, 任红玉. 2020. 山区耕地细碎化对不同利用方式农地集约度的影响. *地球科学进展*, 35(2): 180-188) [DOI: 10.11867/j.issn.1001-8166.2020.011]

Holzworth D P, Snow V, Janssen S, Athanasiadis I N, Donatelli M, Hoogenboom G, White J W and Thorburn P. 2015. Agricultural production systems modelling and software: current status and future prospects. *Environmental Modelling and Software*, 72: 276-286 [DOI: 10.1016/j.envsoft.2014.12.013]

Huang J X, Li X L, Liu D Y, Ma H Y, Tian L Y and Su W. 2015. Comparison of winter wheat yield estimation by sequential assimilation of different spatio-temporal resolution remotely sensed LAI datasets. *Transactions of the Chinese Society for Agricultural Machinery*, 46(1): 240-248 (黄健熙, 李昕璐, 刘帝佑, 马鸿元, 田丽燕, 苏伟. 2015. 顺序同化不同时空分辨率LAI的冬小麦估产对比研究. *农业机械学报*, 46(1): 240-248) [DOI: 10.6041/j.issn.1000-1298.2015.01.034]

Ines A V M, Das N N, Hansen J W and Njoku E G. 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Envi-*

- ronment, 138: 149-164 [DOI: 10.1016/j.rse.2013.07.018]
- Jeong J H, Resop J P, Mueller N D, Fleisher D H, Yun K, Butler E E, Timlin D J, Shim K M, Gerber J S, Reddy V R and Kim S H. 2016. Random forests for global and regional crop yield predictions. *PLoS One*, 11(6): e0156571 [DOI: 10.1371/journal.pone.0156571]
- Ji F J, Meng J H and Fang H T. 2020. Study on soybean yield estimation using the CoupledCASA and WOFOST model. *Remote Sensing Technology and Application*, 35(2): 406-415 (纪甫江, 蒙继华, 方慧婷. 2020. 基于CASA-WOFOST耦合模型的大豆单产遥感估算研究. *遥感技术与应用*, 35(2): 406-415) [DOI: 10.11873/j.issn.1004-0323.2020.2.0406]
- Jiang Z W, Chen Z X, Ren J Q and Zhou Q B. 2012. Estimation of crop yield using CERES-Wheat model based on particle filter data assimilation method. *Transactions of the Chinese Society of Agricultural Engineering*, 28(14): 138-146 (姜志伟, 陈仲新, 任建强, 周清波. 2012. 粒子滤波同化方法在CERES-Wheat作物模型估产中的应用. *农业工程学报*, 28(14): 138-146) [DOI: 10.3969/j.issn.1002-6819.2012.14.022]
- Jin X L, Kumar L, Li Z H, Feng H K, Xu X G, Yang G J and Wang J H. 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92: 141-152 [DOI: 10.1016/j.eja.2017.11.002]
- Kogan F, Kussul N, Adamenko T, Skakun S, Kravchenko O, Kryvobok O, Shelestov A, Kolotii A, Kussul O and Lavrenyuk A. 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *International Journal of Applied Earth Observation and Geoinformation*, 23: 192-203 [DOI: 10.1016/j.jag.2013.01.002]
- Lewis J E, Rowland J and Nadeau A. 1998. Estimating maize production in Kenya using NDVI: some statistical considerations. *International Journal of Remote Sensing*, 19(13): 2609-2617 [DOI: 10.1080/014311698214677]
- Li J L and Liu C. 2005. Research on NDVI seasonal dynamics of Agri-ecosystem on south Coast of Hangzhouwan bay using MODIS data. *Geography and Geo-Information Science*, 21(3): 30-34 (李加林, 刘闯. 2005. 基于MODIS的杭州湾南岸农业生态系统NDVI季节变化特征研究. *地理与地理信息科学*, 21(3): 30-34) [DOI: 10.3969/j.issn.1672-0504.2005.03.008]
- Liu J M, He X T, Wang P X and Huang J X. 2019. Early prediction of winter wheat yield with long time series meteorological data and random forest method. *Transactions of the Chinese Society of Agricultural Engineering*, 35(6): 158-166 (刘峻明, 和晓彤, 王鹏新, 黄健熙. 2019. 长时间序列气象数据结合随机森林法早期预测冬小麦产量. *农业工程学报*, 35(6): 158-166) [DOI: 10.11975/j.issn.1002-6819.2019.06.019]
- Liu Z Z, Zhang X W, Chen Y S, Zhang C C, Qin F and Zeng H W. 2017. Remote sensing estimation of biomass in winter wheat based on CASA model at region scale. *Transactions of the Chinese Society of Agricultural Engineering*, 33(4): 225-233 (刘真真, 张喜旺, 陈云生, 张传才, 秦奋, 曾红伟. 2017. 基于CASA模型的区域冬小麦生物量遥感估算. *农业工程学报*, 33(4): 225-233) [DOI: 10.11975/j.issn.1002-6819.2017.04.031]
- Lobell D B. 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143: 56-64 [DOI: 10.1016/j.fcr.2012.08.008]
- Lobell D B, Asner G P, Ortiz-Monasterio J I and Benning T L. 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems and Environment*, 94(2): 205-220 [DOI: 10.1016/S0167-8809(02)00021-X]
- Lobell D B, Thau D, Seifert C, Engle E and Little B. 2015. A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, 164: 324-333 [DOI: 10.1016/j.rse.2015.04.021]
- Maselli F, Romanelli S, Bottai L and Maracchi G. 2000. Processing of GAC NDVI data for yield forecasting in the Sahelian region. *International Journal of Remote Sensing*, 21(18): 3509-3523 [DOI: 10.1080/014311600750037525]
- Nilsson H E. 1995. Remote sensing and image analysis in plant pathology. *Canadian Journal of Plant Pathology*, 17(2): 154-166 [DOI: 10.1080/07060669509500707]
- Pachavo G and Murwira A. 2014. Remote sensing net primary productivity (NPP) estimation with the aid of GIS modelled shortwave radiation (SWR) in a Southern African Savanna. *International Journal of Applied Earth Observation and Geoinformation*, 30: 217-226 [DOI: 10.1016/j.jag.2014.02.007]
- Qian Y L, Hou Y Y, Yan H, Mao L X, Wu M X and He Y B. 2012. Global crop growth condition monitoring and yield trend prediction with remote sensing. *Transactions of the Chinese Society of Agricultural Engineering*, 28(13): 166-171, 298 (钱永兰, 侯英雨, 延昊, 毛留喜, 吴门新, 何延波. 2012. 基于遥感的国外作物长势监测与产量趋势估计. *农业工程学报*, 28(13): 166-171, 298) [DOI: 10.3969/j.issn.1002-6819.2012.13.027]
- Quan W T, Zhou H, Li H M, Yi W J and Li H L. 2015. Remote Recognition and Growth Monitoring of Winter Wheat in Key Stages Based on S-G Filter in Guanzhong Region. *Chinese Journal of Agrometeorology*, 36(1): 93-99 (权文婷, 周辉, 李红梅, 伊万娟, 李化龙. 2015. 基于S-G滤波的陕西关中地区冬小麦生育期遥感识别和长势监测. *中国农业气象*, 36(1): 93-99) [DOI: 10.3969/j.issn.1000-6362.2015.01.012]
- Rasmussen M S. 1992. Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR. *International Journal of Remote Sensing*, 13(18): 3431-3442 [DOI: 10.1080/01431169208904132]
- Ren J Q, Chen Z X, Zhou Q B, Liu J and Tang H J. 2015. MODIS vegetation index data used for estimating corn yield in USA. *Journal of Remote Sensing*, 19(4): 568-577 (任建强, 陈仲新, 周清波, 刘佳, 唐华俊. 2015. MODIS植被指数的美国玉米单产遥感估测. *遥感学报*, 19(4): 568-577) [DOI: 10.11834/jrs.20154146]
- Rural Social and Economic Investigation Department of National Bu-

- reau of Statistics. 2020. China Statistical Yearbook (County-Level)-2019. Beijing: China Statistics Press (国家统计局农村社会经济调查司). 2020. 中国县域统计年鉴(县市卷)-2019. 北京: 中国统计出版社)
- Tatsumi K, Yamashiki Y, Torres M A C and Taïpe C L R. 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, 115: 171-179 [DOI: 10.1016/j.compag.2015.05.001]
- Tennakoon S B, Murty V V N and Eiumnoh A. 1992. Estimation of cropped area and grain yield of rice using remote sensing data. *International Journal of Remote Sensing*, 13(3): 427-439 [DOI: 10.1080/01431169208904047]
- Thenkabail P S, Ward A D and Lyon J G. 1994. LANDSAT-5 thematic mapper models of soybean and corn crop characteristics. *International Journal of Remote Sensing*, 15(1): 49-61 [DOI: 10.1080/01431169408954050]
- Wang F L, Wang F M, Hu J H, Xie L L and Xie J K. 2020. Estimating and Mapping Rice Yield Using UAV-Hyperspectral Imager based Relative Spectral Variates. *Remote Sensing Technology and Application*, 35(2): 458-468 (王飞龙, 王福民, 胡景辉, 谢莉莉, 谢京凯. 2020. 基于相对光谱变量的无人机遥感水稻估产及产量制图. *遥感技术与应用*, 35(2): 458-468) [DOI: 10.11873/j.issn.1004-0323.2020.2.0458]
- Wang P X, Qi X, Li L, Wang L and Xu L X. 2019. Estimation of maize yield based on random forest regression. *Transactions of the Chinese Society for Agricultural Machinery*, 50(7): 237-245 (王鹏新, 齐璇, 李俐, 王蕾, 许连香. 2019. 基于随机森林回归的玉米单产估测. *农业机械学报*, 50(7): 237-245) [DOI: 10.6041/j.issn.1000-1298.2019.07.026]
- Wang P X, Sun H T, Wang L, Xie Y, Zhang S Y and Li L. 2016. Winter wheat yield estimation based on 4D Variational assimilation method and remotely sensed vegetation temperature condition index. *Transactions of the Chinese Society for Agricultural Machinery*, 47(3): 263-271 (王鹏新, 孙辉涛, 王蕾, 解毅, 张树誉, 李俐. 2016. 基于4D-VAR和条件植被温度指数的冬小麦单产估测. *农业机械学报*, 47(3): 263-271) [DOI: 10.6041/j.issn.1000-1298.2016.03.037]
- Wang R Z, Jiang H, Jin J X and Cheng M. 2018. Respond of winter wheat phenology to climate change and its effect on yield in Huang-Huai-Hai area. *Jiangsu Agricultural Sciences*, 46(22): 71-75 (王瑞峥, 江洪, 金佳鑫, 程敏. 2018. 黄淮海地区冬小麦物候对气候变化的响应及对产量的影响. *江苏农业科学*, 46(22): 71-75) [DOI: 10.15889/j.issn.1002-1302.2018.22.015]
- West H, Quinn N and Horswell M. 2019. Remote sensing for drought monitoring and impact assessment: Progress, past challenges and future opportunities. *Remote Sensing of Environment*, 232: 111291 [DOI: 10.1016/j.rse.2019.111291]
- Wiegand C, Shibayama M, Yamagata Y and Akiyama T. 1989. Spectral observations for estimating the growth and yield of rice. *Japanese Journal of Crop Science*, 58(4): 673-683 [DOI: 10.1626/jcs.58.673]
- Wu B F. 2004. China crop watch system with remote sensing. *Journal of Remote Sensing*, 8(6): 481-497 (吴炳方. 2004. 中国农情遥感速报系统. *遥感学报*, 8(6): 481-497) [DOI: 10.11834/jrs.20040601]
- Xin Q C, Gong P, Yu C Q, Yu L, Broich M, Suyker A E and Myneni R B. 2013. A production efficiency model-based method for satellite estimates of corn and soybean yields in the Midwestern US. *Remote Sensing*, 5(11): 5926-5943 [DOI: 10.3390/rs5115926]
- Xu J W, Mei X R, Ju H, Li Y C, Liu Q and Yang J Y. 2014. Simulation of winter wheat yield in response to irrigation level at critical growing stages in the Huang-Huai-Hai plain. *Acta Agronomica Sinica*, 40(8): 1485-1492 (徐建文, 梅旭荣, 居辉, 李迎春, 刘勤, 杨建堂. 2014. 黄淮海地区冬小麦关键生育期不同灌溉水平对产量影响的模拟. *作物学报*, 40(8): 1485-1492) [DOI: 10.3724/SP.J.1006.2014.01485]
- Yang H, Chen E X, Li Z Y, Zhao C J, Yang G J, Pignatti S, Casa R and Zhao L. 2015. Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 34: 157-166 [DOI: 10.1016/j.jag.2014.08.010]
- Zhang J Y, Fu J, Yang K J, Wang Y F, Wang Z H and Wei J P. 2015. Effects of different plant patterns and planting density on yield and photosynthetic characteristics of maize. *Journal of Anhui Agricultural Sciences*, 43(23): 29-32, 93 (张敬宇, 付健, 杨克军, 王玉凤, 王智慧, 魏金鹏. 2015. 不同种植方式和密度对玉米产量及光合特性的影响. *安徽农业科学*, 43(23): 29-32, 93) [DOI: 10.13989/j.cnki.0517-6611.2015.23.013]
- Zhao H, Liu Y C, Wu D and Han Z Z. 2020. Analysis of irrigation water requirement characteristics for wheat-maize continuous cropping in the five counties of Hebei Province. *South-to-North Water Transfers and Water Science and Technology*, 18(4): 185-194 (赵晗, 刘玉春, 吴迪, 韩振中. 2020. 河北省5个县域冬小麦-夏玉米连作灌溉需水特性分析. *南水北调与水利科技(中英文)*, 18(4): 185-194) [DOI: 10.13476/j.cnki.nsbdqk.2020.0086]
- Zhao J W, Wang X P, Du X Y and Shang Z Y. 2016. Spatial structure and control factors of winter wheat production in Hebei province. *Chinese Journal of Eco-Agriculture*, 24(12): 1683-1692 (赵稷伟, 王锡平, 杜汛雨, 尚志云. 2016. 河北省冬小麦生产空间格局及其控制因素. *中国生态农业学报*, 24(12): 1683-1692) [DOI: 10.13930/j.cnki.cjea.160573]
- Zhu W X, Li S J, Zhang X B, Li Y and Sun Z G. 2018. Estimation of winter wheat yield using optimal vegetation indices from unmanned aerial vehicle remote sensing. *Transactions of the Chinese Society of Agricultural Engineering*, 34(11): 78-86 (朱婉雪, 李仕冀, 张旭博, 李洋, 孙志刚. 2018. 基于无人机遥感植被指数优选的田块尺度冬小麦估产. *农业工程学报*, 34(11): 78-86) [DOI: 10.11975/j.issn.1002-6819.2018.11.010]

Analysis of the influence of key-phase growth-environment-landscape features on the accuracy of county-level winter wheat yield estimation in Hebei Province

ZHU Jiong^{1,2}, DU Xin¹, LI Qiangzi¹, ZHANG Yuan¹, WANG Hongyan¹, ZHAO Yuncong^{1,2}

1. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Accurate and rapid regional-scale crop yield estimation can provide effective data support for the formulation of national food security policies. Compared with complex mechanism models, sampling statistical surveys and empirical models based on multi-source data have better reliability and operability for county-level or city-level yield estimation. Previous studies have proposed many factors related to winter wheat yield, but systematic research on the selection and analysis of multi-source factors is lacking. On the basis of remote sensing, meteorological, and statistical data, this study systematically explored the influence of key-phase growth-environment-landscape features on winter wheat yield estimation at the county-level and determined the best time phase and characteristic parameters.

The considered features included crop condition, environmental forcing (e.g., precipitation, light, and temperature), and farmland landscape. The key phases were the key periods of winter wheat yield formation (P1—P5), which were extracted from the NDVI curve of the crop growth process. Random Forest (RF) regression models were developed using different combinations of phases and features to simulate statistical wheat yield data and evaluate the importance of different combinations via an accuracy assessment. The performance of the models built from each layer combination was compared using the Mean Relative Error (MRE), Root Mean-Squared Error (RMSE), Normalized Root Mean-Squared Error (NRMSE), and coefficient of determination (R^2). Data on years 2014—2017 were used to build the models, and 2018 data were utilized for validation.

Results showed that P2, P3, and P4 resulted in higher accuracy than P1 and P5 in terms of single phases. The model accuracy using multi-phase features was higher than that of using single phases, and the combination of P2 and P4 was the best. Among all the features, crop growth features had the greatest impact on yield estimation accuracy, whereas the addition of environmental forcing factors (e.g., water, light, and temperature) did not significantly improve the accuracy. The addition of farmland landscape features could effectively improve the accuracy of yield estimation. Moreover, five important features (PROP, NDVI_P2, B2_P2, ED, and B1_P4) were selected, and a yield estimation model was established to obtain the county-level yield of winter wheat in Hebei Province. The MRE of wheat yield estimation at the county level in 2018 was as low as 2.85%, and the RMSE, NRMSE, and R^2 were 253.25 kg/ha, 4.09%, and 0.83, respectively.

Conclusion Multi-phase performance is better than single-phase performance. Combining crop growth features with farmland landscape features (RMSE of 247.79 kg/ha) provides more accurate estimates than using crop growth features alone (RMSE of 295.95 kg/ha). Furthermore, the RF model produces good yield estimation results. This study provides insights into and new methods for nationwide estimation of winter wheat yield at the county level.

Key words: remote sensing, yield estimation, winter wheat, statistical data, NDVI, random forest, Hebei Province

Supported by National key Research and Development Program of China (No. 2016YFD0300603-5); Key Program of High-Resolution Earth Observation System (No. 20-Y30F10-9001-20122, 11-Y20A16-9001-17/18); Guangxi Science and Technology Major Project (No. Guike AA18118048-2); National Natural Science Foundation of China (No. 41701486)