

运用 PLS 算法由 HJ-1A/1B 遥感影像估测区域小麦实际单产

谭昌伟, 罗 明, 杨 昕, 马 昌, 周 健, 杜 颖, 王雅楠

(扬州大学江苏省作物遗传生理国家重点实验室培育点/粮食作物现代产业技术协同创新中心, 扬州 225009)

摘 要: 为进一步提高遥感估产精度, 显示国产影像在农业估产中的应用效果。该研究以 2010—2013 年 HJ-1A/1B 影像为遥感数据, 分析了卫星遥感变量与小麦实际单产的定量关系, 运用偏最小二乘回归算法构建及验证了以实际单产为目标的多变量遥感估产模型, 并制作了小麦实际单产空间等级分布图。研究表明: 实际单产与所选用的大多数遥感变量间关系密切, 且多数遥感变量两两间具有严重的多重相关关系; 实际单产偏最小二乘回归模型的最佳主成分为 5, 且植被衰减指数、绿色归一化植被指数、调整土壤亮度的植被指数、比值植被指数和归一化植被指数为实际单产遥感估测的敏感变量; 建模集和验证集实际单产估测模型的决定系数分别为 0.74 和 0.70, 均方根误差分别为 $754.05 \text{ kg} \cdot \text{hm}^{-2}$ 和 $748.20 \text{ kg} \cdot \text{hm}^{-2}$, 相对误差分别为 11.5% 和 8.88%, 且估测精度比线性回归算法分别提高 20% 以上和 40% 以上, 比主成分分析算法分别提高 18% 以上和 30% 以上, 说明偏最小二乘回归算法模型估测区域实际单产的效果要明显好于线性回归和主成分分析算法, 该模型应用结果与小麦实际单产区域分布情况相符合, 为提高区域小麦实际单产的遥感估测精度提供了一种途径。

关键词: 遥感; 算法; 回归分析; 产量估测; 偏最小二乘法; HJ-1A/1B; 小麦

doi: 10.11975/j.issn.1002-6819.2015.15.022

中图分类号: S127; TP79

文献标志码: A

文章编号: 1002-6819(2015)-15-0161-06

谭昌伟, 罗 明, 杨 昕, 马 昌, 周 健, 杜 颖, 王雅楠. 运用 PLS 算法由 HJ-1A/1B 遥感影像估测区域小麦实际单产[J]. 农业工程学报, 2015, 31(15): 161—166. doi: 10.11975/j.issn.1002-6819.2015.15.022 <http://www.tcsae.org>
Tan Changwei, Luo Ming, Yang Xin, Ma Chang, Zhou Jian, Du Ying, Wang Ya'nan. Remote sensing estimation of wheat practical yield on regional scale using partial least squares regression algorithm based on HJ-1A/1B images[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(15): 161—166. (in Chinese with English abstract)
doi: 10.11975/j.issn.1002-6819.2015.15.022 <http://www.tcsae.org>

0 引 言

准确地估测作物产量, 对于确保国家粮食安全, 指导和调控宏观种植业结构调整, 提高涉农企业与农民的经营管理水平均具有重要意义。随着遥感数据在空间上、时间上和光谱上的分辨率提高及成本上的极大降低, 当前遥感已被广泛地应用于多种作物估产中, 成为遥感与农业交叉的研究重点。

目前作物估产的方法很多, 如作物产量气象预报、人工抽样调查、统计模拟模型、遥感估产等^[1-6]。其中, 作物产量气象预报比较适宜于小范围作物估产, 但对于大范围作物因不同麦区田间气象差异较大导致估产精度偏低; 人工抽样调查测产精度较高, 但耗时耗力、成本高, 不适于在大范围内进行; 统计模拟模型是结合历史资料构建数学关系模型, 对未来变化趋势做出估测的一种非主观方法, 但是可能会因为气候的变化导致结果的偏差; 遥感能够大面积、快速、动态和多尺度获取农田丰富信息, 在大面积作物长势监测、作物估产、农情监测与预报、农业资源调查等方面具有巨大的应用潜力^[7-12]。国内外有关作物遥感估产研究已做了大量工作^[13-17], 但所用

的卫星遥感资料时效较低, 模型模拟效果的稳定性有待检验, 所使用的卫星数据源主要是 MODIS、NOAA/AVHRR、TM 等国外低空间分辨率数据, 以及 IRS、P6、ETM、Quickbird、SPOT、IKONOS、ALOS 等国外高空间分辨率数据, 但后者数据源多价格高。因此, 农情遥感中推广应用国产卫星影像数据意义重大。中国于 2008 年 9 月 6 日成功发射具有自主知识产权的 HJ-1A/1B 卫星, 该卫星传感器的空间分辨率为 30 m, 时间分辨率为 2 d, 成为农情遥感业务化运行较为理想的数据源。

本研究以 HJ-1A/1B 影像为遥感数据源, 结合偏最小二乘回归 (partial least squares regression, PLS) 算法, 研究田间采样点获取的实际单产与开花期遥感变量间的相关关系, 得到小麦实际单产的敏感遥感变量, 以利于后来构建实际单产遥感估测模型, 实现区域小麦实际单产遥感估测, 以期由政府及其他有关部门及时了解不同生态区域小麦产量丰欠变化趋势, 制定小麦粮食贸易和宏观调控政策提供参考。

1 材料与方法

1.1 试验设计与数据获取

试验 1: 2010 年在江苏省泰兴、姜堰、仪征、兴化、大丰 5 县进行, 每县选 4~7 个镇, 每镇分散设置采样点 3~5 个, 共计 93 个随机采样点, 每个采样点位置均采用 Juno ST 手持式 GPS 定位获取地理信息, 调查时期为开花

收稿日期: 2015-04-03 修订日期: 2015-06-25

基金项目: 国家自然科学基金资助项目 (41271415); 江苏高校优势学科建设工程资助项目 (PAPD)

作者简介: 谭昌伟, 男, 副教授, 博士, 硕士生导师, 扬州 扬州大学农学院, 225009. Email: tanwei010@126.com

期, 小麦品种为中弱筋类型, 主要有扬麦 13、扬麦 15、扬麦 16 等。成熟期用 GPS 定位, 以各县农技推广中心提供实打实收单产作为实际单产。卫星数据为 HJ-1A/1B (开花期) 过境影像。试验 2: 2011 年共 82 个采样点, 卫星数据为 HJ-1A/1B (开花期) 过境影像, 其他同试验 1。试验 3: 2012 年共 80 个采样点, 卫星数据为 HJ-1A/1B (开花期) 过境影像, 其他同试验 1。试验 4: 2013 年共 80 个采样点, 卫星数据为 HJ-1A/1B (开花期) 过境影像, 其他同试验 1。经田间调查, 试验 1~4 小麦同期长势无显著差异。

1.2 影像预处理

采用 ENVI 对 HJ-1A/1B 影像进行预处理。首先利用江苏地区 1:10 万地形图对 HJ-1A/1B 影像进行几何粗校正, 然后再利用地面实测的 GPS 控制点对 HJ-1A/1B 影像进行几何精校正。采用经验线性法转换进行大气校正和反射率转换^[14]。

1.3 影像定标

在研究区选择代表性水体作为低反射定标物, 空旷水泥路面作为高反射定标物, 采用手工定标方法对 HJ-1A/1B 影像进行定标, 利用绝对定标系数将 DN 值图像转换为辐亮度图像的公式为:

$$L = DN/a + L_0 \quad (1)$$

式中: L 为辐亮度, 转换后辐亮度单位为 $W/m^2/sr/\mu m$; a 为绝对定标系数增益; L_0 为偏移量。

1.4 卫星遥感变量

依据作物光谱特性和国内外研究文献, 结合光谱指数的物理意义, 本研究选用 4 个 HJ-1A/1B 波段和 9 种常用光谱植被指数 (表 1) 作为 PLS 分析构建遥感估产模型的入选自变量。利用 ENVI 和 ArcGIS 提取对应 GPS 定位采样点光谱反射率, 结合表 1 提供的遥感植被指数算法, 在 Excel 中计算出卫星遥感变量。

表 1 遥感植被指数的计算公式^[18]

Table 1 Formulas of remote sensing vegetation indices

植被指数 Vegetation index	计算公式 Calculation formulas
归一化植被指数 Normalized difference vegetation index (NDVI)	$NDVI = (B4 - B3) / (B4 + B3)$
土壤调整植被指数 Soil adjusted vegetation index (SAVI)	$SAVI = (B4 - B3) / (B4 + B3 + 0.5) \times (1 + 0.5)$
调整土壤亮度的植被指数 Optimal soil adjusted vegetation index (OSAVI)	$OSAVI = (1 + 0.16) \times (B4 - B3) / (B4 + B3 + 0.16)$
作物氮反应指数 Nitrogen reflectance index (NRI)	$NRI = (B2 - B3) / (B2 + B3)$
绿色归一化植被指数 Green normalized difference vegetation index (GNDVI)	$GNDVI = (B4 - B2) / (B4 + B2)$
结构加强色素植被指数 Structure intensive pigment index (SIPI)	$SIPI = (B4 - B1) / (B4 + B1)$
植被衰减指数 Plant senescence reflectance index (PSRI)	$PSRI = (B3 - B1) / B4$
差值植被指数 Difference vegetation index (DVI)	$DVI = B4 - B3$
比值植被指数 Ratio vegetation index (RVI)	$RVI = B4 / B3$

注: $B1, B2, B3, B4$ 分别代表蓝光、绿光、红光、近红外光波段的光谱反射率。下同。

Note: $B1, B2, B3$ and $B4$ denoted spectrum reflectance at blue, green, red and near infrared bands, respectively. The same as below.

1.5 偏最小二乘回归 PLS

PLS 是一种适用性较广的新型多元分析方法, 能解决利用许多普通多元回归算法无法解决的问题, 尤其当变量多且具有多重相关关系时, 可有效筛选对因变量解释性最强的综合变量, 建立的模型比普通回归模型更可靠。PLS 法先提取称之为成分的新变量作为自变量, 然后建立自变量与因变量之间的线性关系, 利用 PLS 计算出系数, 然后建立因变量的回归模型。即回归模型为:

$$y = a_{0m} + a_{1m} \cdot x_1 + \cdots + a_{pm} \cdot x_p \quad (m = 1, 2, \cdots p) \quad (2)$$

式中: y 实际单产, kg/hm^2 ; x_1, \cdots, x_p 分别为线性组合自变量, 即遥感变量, $a_{0m}, a_{1m}, \cdots, a_{pm}$ 为 PLS 计算出的模型系数。

由 PLS 算法建立模型时, 增加主成分数能提高模型精度, 但主成分过多将产生“过拟合”现象, 精度降低, 因此, 确定最佳主成分数是建立 PLS 模型的关键。本研究利用交互验证法计算出估测残差平方和 (predictive residual error sum of square, PRESS), PRESS 值越小, 表示模型的估测精度越高, 因此, 基于 PRESS 最小值确定最佳主成分数。PRESS 计算公式为:

$$PRESS = \sum_{i=1}^k (y_i - y_{i,-i})^2 \quad (3)$$

式中: $y_i, y_{i,-i}$ 分别为第 i 个样本对应的实测值和剔除第 i 个样本的估计值; k 为验证循环次数。

PLS 和 PRESS 的原理和步骤, 见文献[19]。PLS 和 PRESS 过程均由自编的 MATLAB 程序完成。

1.6 模型的评价

分别用建模集和验证集的样本, 通过绘制实际单产估测值与实测值间的 1:1 散点图进行模型评价, 评价指标为决定系数 (determination coefficient, R^2)^[20]、均方根误差 (root mean square error, RMSE)^[20]和相对误差 (relative error, RE), 其中 R^2 越大, 表明模型越好, RMSE 和 RE 越小, 表明模型的估测精度越高。

2 结果与分析

2.1 产量分布

整理试验 1~4 中测定的实际单产数据, 按实际单产数值的大小排序, 为提高估测模型的稳定性, 在保证实际单产最大值和最小值置于建模样本集的前提下, 将 335 个实际单产数值样本按 3:2 的比例, 随机分成建模集和验证集。由表 2 可知, 实际单产建模集和验证集样本的变幅、平均值、标准差和标准误差均相近, 具有较好的一致性。

表 2 建模集和验证集样本实际单产分布

Table 2 Distribution of wheat practical yield in modeling and testing sample datasets

样本集 Sample datasets	样本数 Sample number	变幅 Amplitude/ (kg·hm ⁻²)	平均值 Average value/ (kg·hm ⁻²)	标准差 Standard deviation/ (kg·hm ⁻²)	标准误差 Standard error/ (kg·hm ⁻²)
建模集 Modeling datasets	201	2299.2~ 9663.9	5610.9	1311.9	124.8
验证集 Testing datasets	134	2806.05~ 9565.95	5306.1	1371.15	118.65

表 3 遥感变量与实际单产间相关系数 (n=201)

Table 3 Correlation between remote sensing variables and wheat practical yield

	实际单产 Practical yield	B1	B2	B3	B4	NDVI	SAVI	OSAVI	NRI	GNDVI	SIPI	PSRI	DVI	RVI
实际单产 Practical yield	1.00													
B1	0.48**	1.00												
B2	0.44**	0.99**	1.00											
B3	0.36**	0.99**	0.99**	1.00										
B4	0.32**	0.95**	0.97**	0.95**	1.00									
NDVI	0.52**	0.83**	0.87**	0.82**	0.96**	1.00								
SAVI	0.52**	0.83**	0.87**	0.82**	0.96**	1.00**	1.00							
OSAVI	0.52**	0.83**	0.87**	0.82**	0.96**	1.00**	1.00	1.00						
NRI	0.24*	0.92**	0.91**	0.95**	0.87**	0.75**	0.75**	0.75**	1.00					
GNDVI	0.68**	0.91**	0.94**	0.92**	0.98**	0.97**	0.97**	0.97**	0.88**	1.00				
SIPI	0.61**	0.92**	0.95**	0.93**	0.98**	0.96**	0.96**	0.96**	0.88**	0.99**	1.00			
PSRI	0.73**	0.92**	0.93**	0.94**	0.91**	0.84**	0.84**	0.84**	0.91**	0.91**	0.96**	1.00		
DVI	0.38**	0.86**	0.90**	0.85**	0.98**	0.99**	0.99**	0.99**	0.77**	0.97**	0.96**	0.84**	1.00	
RVI	0.18	0.70**	0.75**	0.69**	0.88**	0.97**	0.97**	0.97**	0.61**	0.90**	0.88**	0.72**	0.96**	1.00

注：*, ** 分别表示在 5% 和 1% 水平下显著。

Note: *, ** indicated significant at 5% and 1% level respectively.

2.3 构建 PLS 模型

2.3.1 最佳主成分数目确定

图 1 为 PRESS 随主成分数的变化。起始时随着主成分数增加，实际单产 PRESS 值都呈较大幅度地降低，由此表明由于主成分数目较少，模型拟合极其不充分，即出现缺失拟合现象，直至实际单产模型主成分数为 5 时，PRESS 值达到最小，为 24.17；之后，随着主成分数逐渐增加，PRESS 值陡然增加，直至趋于缓平饱和，说明因主成分数过多而出现“过拟合”现象。因此，选择 PRESS 值最小时对应的主成分数作为 PLS 模型的最佳主成分数是合理的，即该实际单产 PLS 模型的最佳主成分数为 5。

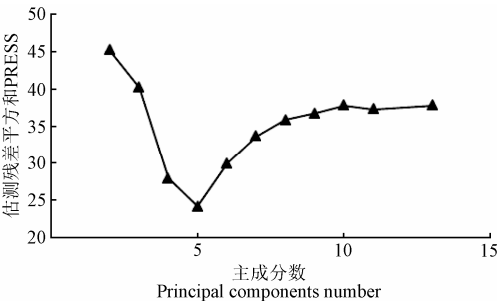


图 1 PRESS 随主成分数的变化

Fig.1 PRESS changes with the principal components

2.3.2 PLS 模型建立及评价

基于 PLS 算法，以主成分数为 5 的 5 个植被指数，即 PSRI、GNDVI、OSAVI、RVI 和 NDVI 为自变量，以

2.2 遥感变量与实际单产间的关系定量分析

建模集中 201 个样本的实际单产与遥感变量间关系定量分析表明（表 3），实际单产与大多数遥感变量间具有显著或极显著关系，其中实际单产与 PSRI 关系最密切，其次与 GNDVI，相关系数分别为 0.73 和 0.68，大多数植被指数与实际单产间的相关性明显好于单波段；多数遥感变量两两间具有严重的多重相关关系，其相关系数几乎介于 0.80 与 1.00 之间，尤其 B1~B4 两两间相关系数介于 0.95 与 0.99 之间，且多数植被指数两两间相关系数在 0.90 以上，说明相对于传统统计算法和普通最小二乘法，运用 PLS 算法建立模型的区域估测精度可能更高。

实际单产为因变量，使用 201 个建模集样本及其对应的 HJ-1A/1B 影像，构建实际单产估测模型为：

$$y=1706.1\text{GNDVI}-1964.7\text{PSRI}+1764.45\text{OSAVI}+1478.1\text{RVI}+1374.75\text{NDVI}+2133.9。$$

建好实际单产估测模型后，利用本模型估测实际单产，将实际单产估测值与实测值绘成 1:1 散点图，统计出最优直线回归方程及其 R^2 、RMSE 和 RE，由图 2 可知，建模集和验证集的实际单产估测值与实测值间的 R^2 均大于 0.7，RMSE 分别为 $754.05 \text{ kg} \cdot \text{hm}^{-2}$ 和 $748.2 \text{ kg} \cdot \text{hm}^{-2}$ ，RE 分别为 11.5% 和 8.88%，说明利用该 PLS 模型能较好地估测区域小麦实际单产。

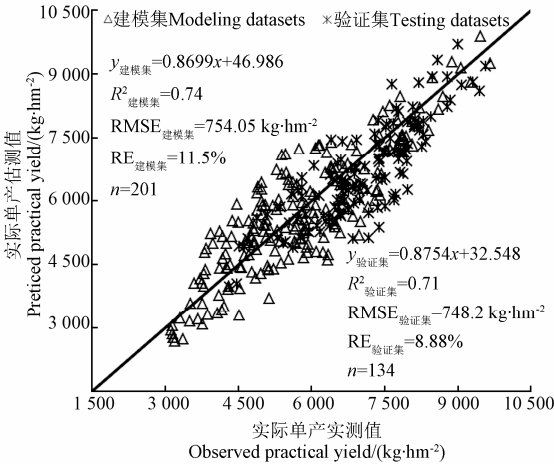


图 2 实际单产估测模型评价

Fig.2 Evaluation of practical estimation yield model

2.3.3 与传统算法比较

为了与传统算法进行比较,采用线性回归(linear regression, LR)和主成分分析(principle components analysis, PCA)算法,利用建模集样本建立实际单产估测模型,依据实际单产估测值与实测值的相关系数、RMSE 和 RE 评价模型。表 4 为 PLS、LR 和 PCA 算法模型估测结果,比较

发现:样本数相同,实际单产 PLS 模型的估测值与实测值相关系数大于 LR 和 PCA 模型, RMSE 和 RE 都小于 LR 和 PCA 模型,说明 PLS 模型估测区域实际单产的效果要好于 LR 和 PCA 模型,其中实际单产建模集和验证集估测精度分别比 LR 算法提高 22.46%和 41.21%,分别比 PCA 算法提高 18.57%和 32.56%。

表 4 PLS、LR 和 PCA 算法模型估测结果比较

Table 4 Comparison of predicted results with PLS, LR and PCA

算法 Algorithm	主成分数 Principal components number	样本数 Sample number		相关系数 Correlation coefficient		均方根误差 RMSE/(kg·hm ⁻²)		相对误差 RE/%	
		建模集 Modeling datasets	验证集 Testing datasets	建模集 Modeling datasets	验证集 Testing datasets	建模集 Modeling datasets	验证集 Testing datasets	建模集 Modeling datasets	验证集 Testing datasets
PLS	5	201	134	0.861	0.834	754.05	748.2	11.5	8.88
PCA	6	201	134	0.762	0.747	1048.95	1098.15	25.36	31.26
LR	0	201	134	0.714	0.677	1119.45	1197.6	27.73	35.47

基于上述 PLS 模型,利用 2013 年小麦开花期 HJ-1A/1B 影像,依据表 1 计算公式,分别生成敏感卫星遥感变量数值图,一一进行求算,经二值化掩膜,利用研究区 GPS 定位样点数据,采用监督分类法提取小麦种植面积,然后实地抽样校正,以保证小麦种植面积提取精度,从而得到小麦种植空间分布图(图 3a),叠加包含研究区域的江苏行政区划矢量数据,制作江苏中部地区小麦实际单产空间分布遥感估测图(图 3b),实际单

产分布主要为 $\geq 3\,750 \sim 5\,250 \text{ kg} \cdot \text{hm}^{-2}$,其中兴化及其周边麦区主要为 $3\,750 \sim 4\,500 \text{ kg} \cdot \text{hm}^{-2}$,兴化以北麦区主要为 $\geq 5\,250 \sim 6\,000 \text{ kg} \cdot \text{hm}^{-2}$,淮安以北麦区频现大于 $6\,000 \text{ kg} \cdot \text{hm}^{-2}$,以南麦区极少出现高于 $6\,000 \text{ kg} \cdot \text{hm}^{-2}$,沿江麦区主要为 $\geq 3\,750 \sim 5\,250 \text{ kg} \cdot \text{hm}^{-2}$,尤其长江以南麦区主要为 $\geq 3\,750 \sim 4\,500 \text{ kg} \cdot \text{hm}^{-2}$ 。该结果与收获期外业调查及当地农技部门提供的分布情况是一致的,进一步说明利用 PLS 模型估测区域小麦实际单产是可靠的。

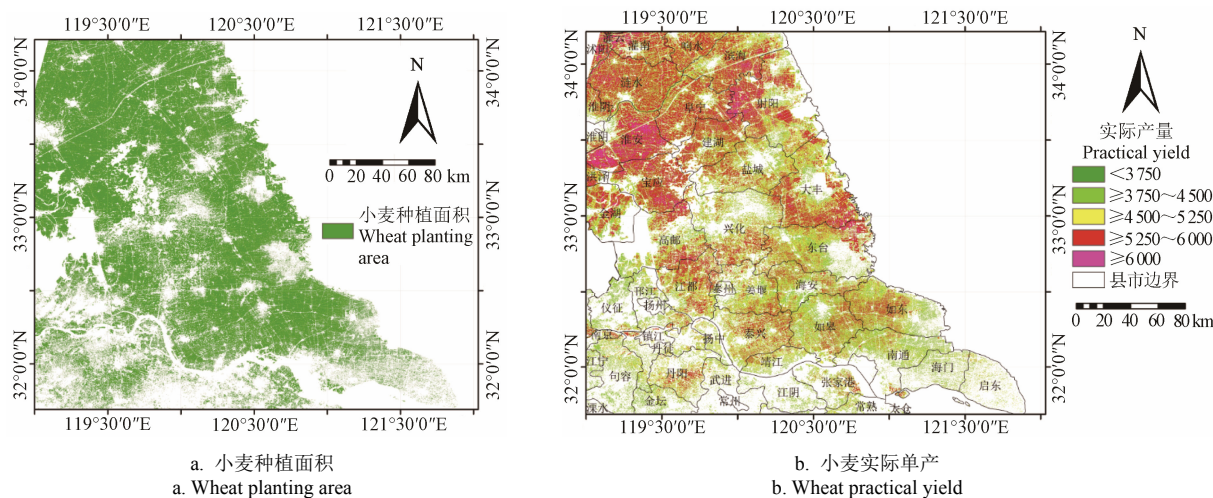


图 3 江苏中部地区小麦种植面积和小麦实际单产空间分布
Fig.3 Spatial distribution level of wheat planting area and wheat practical yield in central Jiangsu region

3 讨论

随着 HJ-1A/1B 卫星投入使用,所获数据质量不断提高,且免费提供给用户使用,为区域性作物遥感估产提供了便捷的数据平台。本研究试验区地处江苏沿江、沿海地带,麦田整体较为零碎,种植结构布局复杂,所选用的 HJ-1A/1B 影像时间分辨率为 2 d,单景影像幅宽为 750 km,能满足区域性小麦估产需求。虽然年度之间田间物候存在一定的差异,但能保证所选用遥感数据处于同一物候期。

本研究将遥感估测的实际单产进行不同等级空间分布,并赋予地理空间信息,以此表达小麦实际单产的空

间实际分布状况和趋势,说明实时的作物遥感估产,可以在空间上定性和定量地分析作物实际单产状况,不但可以分级量化显示作物实际单产分布状况,而且可以分区统计作物实际单产不同等级所占的比重,从而能够向大田区域生产提供及时可靠的实际单产信息。

笔者研究了基于 TM 和 PLS 的小麦籽粒蛋白质含量预测,结果表明,PLS 模型对小麦籽粒蛋白质含量预测能力要好于 LR 和 PCA 模型^[21]。本研究结果也证明了这一结论,说明本研究结果具有一定的可靠性。但本研究仍存在不足,主要表现为:1) 在建模思路,仅选择了一个生长期数据,使得研究结果具有较大的局限性,比如开花期的数据不一定能获得最好的预测效果,其他时

期数据的预测效果可能会更好, 另外, 不同时期的数据可能最优分析建模方法也不同, 这方面将来非常有必要深入探究; 2) 所用实际单产样本主要以 $2\ 800 \sim 8\ 750\ \text{kg}/\text{hm}^2$, $>8\ 750$ 和 $<2\ 800\ \text{kg}/\text{hm}^2$ 的样本偏少, 且缺少 $>9\ 663.9$ 和 $<2\ 299.2\ \text{kg}/\text{hm}^2$ 的样本, 若增加建模样本的变幅, 将会进一步优化 PLS 模型和扩大其应用范围, 建立的小麦实际单产遥感估测模型, 精度更可靠, 所取得结果是仅基于江苏试验区 HJ-1A/1B 数据建立的, 因此, 本模型是否适用于其他遥感传感器数据以及生态地区, 有待进一步研究; 3) 本研究是基于 HJ-1A/1B 影像进行的, 由于其空间分辨率偏低, 常常导致“同物异谱”、“同谱异物”、“过拟合”等状况, 从而降低遥感估产精度, 以致遥感估产结果仍不能满足精准农业生产需求, 以后将多源、多时相遥感数据与作物生长模型进行耦合, 可能是解决这一问题的有效途径; 4) RVI 理论上值域变幅较大, 但在本研究中相对稳定, 可能原因是小麦该生长期长势均匀、变化幅度小, 对应的近红外和红光波段反射性能稳定, 至于建立的公式各参数的系数差异较小, 尤其是 RVI, 在此不能妄下结论, 其原因是否与数据源有关, 还是其他原因, 仍有待深入研究; 5) 本研究仅将 PLS 与 LR 和 PCA 进行对比, 应该增加包括人工神经网络、支持向量机、蚁群、粒子群等算法, 以确保结论的可靠性, 此外, 也未考虑小麦栽培因素影响, 实际上, 这些算法和因素对小麦遥感估产存在较大的影响, 也有待进一步研究。

4 结 论

利用 PLS 算法构建了以 PSRI、GNDVI、OSAVI、RVI 和 NDVI 指数为自变量的实际单产遥感估测模型, 经建模集和验证集评价, 实际单产估测模型的 R^2 分别为 0.74 和 0.70, RMSE 分别为 754.05 和 $748.20\ \text{kg}/\text{hm}^2$, RE 分别为 11.50% 和 8.88%, 且估测精度比线性回归算法分别提高 22.46% 和 41.21%, 比主成分分析算法分别提高 18.57% 和 32.56%, 说明 PLS 模型区域估测效果明显好于 LR 和 PCA。本模型应用结果与小麦区域生产实际相符合, 为提高区域性小麦遥感估产精度提供了一种有效途径。

致谢: 感谢江苏高校优势学科建设工程资助项目为本研究提供试验平台。

参 考 文 献

- [1] 杨武德, 宋艳曦, 宋晓彦, 等. 基于 3S 和实测相结合的冬小麦估产研究[J]. 农业工程学报, 2009, 25(2): 131—135.
Yang Wude, Song Yantun, Song Xiaoyan, et al. Winter wheat yield estimating based on 3S integration and field measurement[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2009, 25(2): 131—135. (in Chinese with English abstract)
- [2] 黄敬峰, 杨忠恩, 王人潮, 等. 基于 GIS 的水稻遥感估产模型研究[J]. 遥感技术与应用, 2002, 17(3): 125—128.
Huang Jingfeng, Yang Zhongen, Wang Renchao, et al. The rice production forecasting models using NOAA/AVHRR data based on GIS[J]. Remote Sensing Technology and Application, 2002, 17(3): 125—128. (in Chinese with English abstract)

- [3] 吴炳方. 全国农情监测与估产的运行化遥感方法[J]. 地理学报, 2000, 55(1): 25—35.
Wu Bingfang. Operational remote sensing methods for agricultural statistics[J]. Journal of Geographical Sciences, 2000, 55(1): 25—35. (in Chinese with English abstract)
- [4] 李卫国, 王纪华, 赵春江, 等. 基于遥感信息和产量形成过程的小麦估产模型[J]. 麦类作物学报, 2007, 27(5): 904—907.
Li Weiguo, Wang Jihua, Zhao Chunjiang, et al. A model of estimating winter wheat yield based on TM image and yield formation[J]. Journal of Triticeae Crop, 2007, 27(5): 904—907. (in Chinese with English abstract)
- [5] 任建强, 陈仲新, 唐华俊. 基于 MODIS—NDVI 的区域冬小麦遥感估产——以山东省济宁市为例[J]. 应用生态学报, 2006, 17(12): 2371—2375.
Ren Jianqiang, Chen Zhongxin, Tang Huajun. Regional scale remote sensing-based yield estimation of winter wheat by using MODIS-NDVI data: A case study of Jining city in Shandong Province[J]. Chinese Journal of Application Ecology, 2006, 17(12): 2371—2375. (in Chinese with English abstract)
- [6] 王长耀, 林文鹏. 基于 MODIS EVI 的冬小麦产量遥感预测研究[J]. 农业工程学报, 2005, 21(10): 90—94.
Wang Changyao, Lin Wenpeng. Winter wheat yield estimation based on MODIS EVI[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2005, 21(10): 90—94. (in Chinese with English abstract)
- [7] 欧文浩, 苏伟, 薛文振, 等. 基于 HJ-1 卫星影像的三大农作物估产最佳时相选择[J]. 农业工程学报, 2010, 26(11): 176—182.
Ou Wenhao, Su Wei, Xue Wenzhen, et al. Selection of optimum phase for yield estimation of three major crops based on HJ-1 satellite images[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2010, 26(11): 176—182. (in Chinese with English abstract)
- [8] Reyniers M, Vrindts E, Baerdemaeker J D. Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat[J]. European Journal of Agronomy, 2006, 24(2): 87—94.
- [9] Zhang P, Anderson B, Tan B, et al. Potential monitoring of crop production using a satellite-based climate-variability impact index[J]. Agricultural and Forest Meteorology, 2005, 132(3): 344—358.
- [10] Domenikiotis C, Spiliotopoulos M, Tsiros E, et al. Early cotton yield assessment by the use of the NOAA/AVHRR derived Vegetation Condition Index (VCI) in Greece[J]. International Journal of Remote Sensing, 2004, 25(14): 2807—2819.
- [11] 任建强, 陈仲新, 周清波, 等. 基于叶面积指数反演的区域冬小麦单产遥感估测[J]. 应用生态学报, 2010, 21(11): 2883—2888.
Ren Jianqiang, Chen Zhongxin, Zhou Qingbo, et al. LAI-based regional winter wheat yield estimation by remote sensing[J]. Chinese Journal of Application Ecology, 2010, 21(11): 2883—2888. (in Chinese with English abstract)
- [12] 任建强, 刘杏认, 陈仲新, 等. 基于作物生物量估计的区域冬小麦单产预测[J]. 应用生态学报, 2009, 20(4): 872—878.
Ren Jianqiang, Liu Xinren, Chen Zhongxin, et al. Prediction of winter wheat yield based on crop biomass estimation at regional scale[J]. Chinese Journal of Application Ecology, 2009, 20(4): 872—878. (in Chinese with English abstract)
- [13] 朱再春, 陈联裙, 张锦水, 等. 基于信息扩散和关键期遥感数据的冬小麦估产模型[J]. 农业工程学报, 2011, 27(2): 187—193.
Zhu Zaichun, Chen Lianqun, Zhang Jinshui, et al. Winter wheat yield estimation model based on information diffusion and remote sensing data at major growth stages[J].

- Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2011, 27(2): 187—193. (in Chinese with English abstract)
- [14] 徐新刚, 王纪华, 黄文江, 等. 基于权重最优组合和多时相遥感的作物估产[J]. 农业工程学报, 2009, 25(9): 137—142. Xu Xingang, Wang Jihua, Huang Wenjiang, et al. Estimation of crop yield based on weight optimization combination and multi-temporal remote sensing data[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2009, 25(9): 137—142. (in Chinese with English abstract)
- [15] Hochheim K P, Barber D G. Spring wheat yield estimation for western Canada using NOAA NDVI data[J]. Canada Journal of Remote Sensing, 1998, 24(2): 17—27.
- [16] Nemani R R, Keeling C D, Hashimoto H, et al. Climate driven increases in global net primary production from 1982 to 1999[J]. Science, 2003, 300(5625): 1560—1563.
- [17] Singh R, Semwal D P, Rai A, et al. Small area estimation of crop yield using remote sensing satellite data[J]. International Journal of Remote Sensing, 2002, 23(1): 49—56.
- [18] 谭昌伟, 杨昕, 马昌, 等. 小麦花后 15d 主要苗情参数多光谱卫星遥感定量监测[J]. 麦类作物学报, 2015, 35(4): 569—576.
Tan Changwei, Yang Xin, Ma Chang, et al. Monitoring major seeding parameters of wheat at 15 days after anthesis using multi-spectral remote sensing based on HJ-CCD images[J]. Journal of Triticeae Crops, 2015, 35(4): 569—576. (in Chinese with English abstract)
- [19] 高惠璇. 两个多重相关变量组的统计分析(3): 偏最小二乘回归与 PLS 过程[J]. 数理统计与管理, 2002, 21(2): 58—64. Gao Huixuan. Statistical analyses for multiple correlation variables of two sets(3): Partial least squares regression and PLS Procedure[J]. Application of Statistics and Management, 2002, 21(2): 58—64. (in Chinese with English abstract)
- [20] 谭昌伟, 周清波, 齐腊, 等. 水稻氮素营养高光谱遥感诊断模型[J]. 应用生态学报, 2008, 19(6): 1261—1268. Tan Changwei, Zhou Qingbo, Qi La, et al. Hyperspectral remote sensing diagnosis models of rice plant nitrogen nutritional status[J]. Chinese Journal of Application Ecology, 2008, 19(6): 1261—1268. (in Chinese with English abstract)
- [21] 谭昌伟, 王纪华, 黄文江, 等. 基于 TM 和 PLS 的冬小麦籽粒蛋白质含量预测[J]. 农业工程学报, 2011, 27(3): 388—392. Tan Changwei, Wang Jihua, Huang Wenjiang, et al. Predicting grain protein content in winter wheat based on TM images and partial least squares regression[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2011, 27(3): 388—392. (in Chinese with English abstract)

Remote sensing estimation of wheat practical yield on regional scale using partial least squares regression algorithm based on HJ-1A/1B images

Tan Changwei, Luo Ming, Yang Xin, Ma Chang, Zhou Jian, Du Ying, Wang Ya'nan

(Jiangsu Key Laboratory of Crop Genetics and Physiology/Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou 225009, China)

Abstract: Estimation of crop yield by remote sensing is a key research and application field in agriculture, and such research can provide timely and reliable yield information for regional food production. In order to further improve the accuracy of estimating wheat yield by remote sensing, and demonstrate the application of satellite imaging products in agricultural production, we used HJ-1A/1B images on April 26th 2010, April 28th 2011 and 2012, May 2nd 2013 at wheat anthesis stage as remote sensing data. 335 samples of wheat yield were collected from agriculture production field and divided into modeling dataset and validation dataset on a ratio of 3:2. Based on the minimum value of predictive residual error sum of square (PRESS), the number required for principal component model was determined. The yield estimation model was assessed through determination coefficient (R^2), root mean square error (RMSE) and relative error (RE). This research was undertaken to make a systematic analysis on the quantitative relationship of satellite remote sensing variables to actual wheat yield. Depending on the partial least squares regression (PLS), the multivariable remote sensing estimation models and the space level distribution maps of actual wheat yield were constructed and verified by the modeling dataset and validation dataset, and the estimation effect of the PLS model was compared to linear regression (LR) and principal components analysis (PCA) algorithm models, respectively. The results of this research indicated that the majority of remote sensing variables were significantly ($P < 0.05$) related to practical yield, and there were significant ($P < 0.05$) multiple relationships among the majority of remote sensing variables. For the actual yield estimation model based on PLS, the number of the best principal components was 5. Plant senescence reflectance index (PSRI), green normalized difference vegetation index (GNDVI), optimal soil adjusted vegetation index (OSAVI), ratio vegetation index (RVI) and normalized difference vegetation index (NDVI) were identified as the sensitive remote sensing variables for estimating wheat yield. Through testing the actual yield estimation model based on PLS algorithm with modeling dataset and validation dataset, the R^2 were 0.74 and 0.71, respectively, and the RMSE were 754.05 kg/hm² and 748.2 kg/hm², respectively, the RE were 11.50% and 8.88%, respectively. The PLS model with selected sensitive variables performed better to estimate wheat yield. PLS algorithm models to estimate wheat yield obtained the higher accuracy by above 20% and above 40% than the LR algorithm models, by above 18% and above 30% than the PCA algorithm models for modeling dataset and validation dataset, respectively. Based on the above PLS model and HJ-1A/1B image on May 2nd, 2013, the wheat practical yield spatial distribution level was mapped in central Jiangsu region. The results of applying the PLS models were correspondent with the actual distribution of wheat yield. It was concluded that PLS algorithm can provide an effective way to improve the accuracy of estimating wheat yield on regional scale based on aerospace remote sensing, and can contribute to large-scale application of the research results.

Key words: remote sensing; algorithms; regression analysis; yield estimation; partial least squares method; HJ-1A/1B; wheat