

# 基于 TM 和 PLS 的冬小麦籽粒蛋白质含量预测

谭昌伟<sup>1</sup>, 王纪华<sup>2</sup>, 黄文江<sup>2</sup>, 王君婵<sup>1</sup>, 朱新开<sup>1</sup>, 郭文善<sup>1\*</sup>

(1. 扬州大学江苏省作物遗传生理重点实验室/农业部长江中下游作物生理生态与栽培重点开放实验室, 扬州 225009;

2. 国家农业信息化工程技术研究中心, 北京 100089)

**摘要:** 为进一步提高遥感预测冬小麦籽粒蛋白质含量精度, 分析了卫星遥感变量与冬小麦籽粒蛋白质含量间的定量关系, 运用偏最小二乘法构建了遥感预测籽粒蛋白质含量模型, 制作了冬小麦籽粒蛋白质含量空间等级分布图, 结果表明, 该模型的最佳主成分数为 5, 且归一化植被指数、冠层结构不敏感色素指数、比值植被指数、氮反射指数和植被衰减指数为预测籽粒蛋白质含量的敏感变量; 籽粒蛋白质含量预测的均方根误差为 0.307%, 决定系数为 0.642, 为提高遥感预测小麦品质的精度提供了一种有效途径, 有利于大面积应用和推广。

**关键词:** 模型, 预测控制系统, 蛋白质, Landsat TM, 冬小麦

doi: 10.3969/j.issn.1002-6819.2011.03.070

中图分类号: S127

文献标志码: A

文章编号: 1002-6819(2011)-03-0388-05

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 CSAE, 2011, 27(3): 388—392. (in Chinese with English abstract)

## 0 引言

籽粒蛋白质含量 (grain protein content, GPC) 是衡量小麦品质的关键指标, 目前的检测方法成本高、效率低、样点代表性差, 难以及时掌握小麦“面状”品质信息, 然而, 遥感技术凭借快速、准确、范围广的优势为提前获得小麦品质状况提供了一种有效途径。多年来, 农业遥感主要集中在作物长势监测以及产量估算等方面, 并形成一套比较完整的技术体系<sup>[1-3]</sup>, 但关于作物品质遥感预测研究报道较少, 主要是利用地面光谱对作物品质进行预测<sup>[4-6]</sup>。近年来, 不少学者基于空间卫星遥感平台预测作物品质进行了研究<sup>[7-9]</sup>。Badri 等<sup>[10]</sup>发现在小麦开花前两周从 TM 提取的植被指数与 GPC 关系较显著; Zhao 等<sup>[11]</sup>利用 TM 和 ETM 数据预测小麦 GPC, 效果较好; Reyniers 等<sup>[12]</sup>认为利用 KODAK-CIR 和 CropScan 数据在小麦收获前一个月预测冬小麦 GPC 是可行的。中国内外有关基于偏最小二乘 (partial least squares, PLS) 法进行农情遥感监测已有许多报道, 主要集中在作物病虫害及长势方面<sup>[13-19]</sup>。

有些关于小麦品质遥感预测的研究, 仍基于传统算法, 精度受到影响。本研究以卫星遥感变量与小麦 GPC

间的定量关系为基础, 采用 TM 数据和 PLS 算法, 构建及验证冬小麦 GPC 遥感预测模型, 旨在进一步提高小麦品质遥感预测精度。

## 1 材料与方法

### 1.1 试验设计与数据获取

试验于 2008 年在江苏省泰兴、姜堰、仪征、兴化、大丰、赣榆、灌云、宿迁 8 县进行, 每县设置采样点 15~20 个, 共计 163 个采样点, 每个采样点均采用美国 Trimble 公司生产的 Juno ST 手持式 GPS 定位, 调查冬小麦品种类型、生育期、群体长势及与病虫害。冬小麦品种为中弱筋类型, 主要有扬麦 13、扬麦 15、扬麦 16 和扬辐麦 2, 这些品种在试验县均有分布。取样时期为拔节期、抽穗期、开花期和成熟期, 具体取样时间根据卫星过境周期和实验区实际情况确定, 一般在试验地块中间, 选长势均匀的植株, 取 4 行 50 cm, 同时用 GPS 定位。成熟期取样后及时在实验室用凯氏定氮法测定 GPC。遥感数据为 TM 2008-05-02 (开花期)。2009 年在江苏省泰兴、姜堰、仪征、兴化、大丰 5 县进行, 每县设置采样点 15~20 个, 共计 82 个采样点。遥感数据为 TM 2009-04-26 (开花期) 影像。

### 1.2 影像预处理

采用 ENVI 对卫星影像进行预处理。首先利用江苏地区 1:100 000 地形图对卫星影像进行几何粗校正, 然后再利用地面实测的 GPS 控制点对卫星影像进行几何精校正, 确保几何校正精度优于 1 个像元。大气校正和反射率转换是采用经验线性法转换进行<sup>[7]</sup>。

### 1.3 卫星遥感变量

依据作物光谱特性和结合光谱指数的物理意义, 选

收稿日期: 2010-10-28 修订日期: 2011-03-10

基金项目: 国家自然科学基金资助项目 (40801122); 江苏高校优势学科建设工程资助项目; 公益性行业 (农业) 科研专项经费资助项目 (200803037)  
作者简介: 谭昌伟 (1980—), 男, 博士, 研究方向为农业遥感与信息技术应用。扬州 扬州大学江苏省作物遗传生理重点实验室, 225009。

Email: tanwei010@126.com

\*通信作者: 郭文善 (1961—), 男, 教授, 博士, 研究方向为作物栽培生理与信息农业。扬州 扬州大学江苏省作物遗传生理重点实验室, 225009。

Email: guows@yzu.edu.cn

用 7 个 TM 波段和 12 种常用光谱植被指数（表 1），作为 PLS 分析构建遥感预测模型的入选自变量。

表 1 遥感植被指数的计算公式  
Table 1 Formulas of remote sensing vegetation indices

植被指数	计算公式
归一化植被指数 <i>NDVI</i>	$NDVI=(B4-B3)/(B4+B3)$
土壤调整植被指数 <i>SAVI</i>	$SAVI=(B4-B3)/(B4+B3+0.5) (1+0.5)$
调整土壤亮度植被指数 <i>OSAVI</i>	$OSAVI=(1+0.16)(B4-B3)/(B4+B3+0.16)$
作物氮反射指数 <i>NRI</i>	$NRI=(B2-B3)/(B2+B3)$
绿色归一化植被指数 <i>GNDVI</i>	$GNDVI=(B4-B2)/(B4+B2)$
冠层结构不敏感色素指数 <i>SIPI</i>	$SIPI=(B4-B1)/(B4+B1)$
植被衰减指数 <i>PSRI</i>	$PSRI=(B3-B1)/B4$
湿度指数 <i>WI</i>	$WI=(B7-B5)/(B7+B5)$
差值植被指数 <i>DVI</i>	$DVI= B4-B3$
比值植被指数 <i>RVI</i>	$RVI= B4/B3$
归一化差异水体指数 <i>NDWI</i>	$NDWI=(B4-B5)/(B4+B5)$
归一化差异水体指数 <i>NDWI2</i>	$NDWI2=(B4-B7)/(B4+B7)$

注：B1，B2，B3，B4，B5，B6，B7 分别代表各波段的反射率。

利用 ENVI 和 ArcGIS 提取对应 GPS 定位采样点光谱波段 DN（digital number）值，反演为反射率，结合表 1 提供的植被指数算法计算出卫星遥感变量。

1.4 偏最小二乘回归法

PLS 最早应用于化学计量学领域，是一种新型的多元分析方法，适用性较广，它集中了主成分回归分析（princlple component regression, PCR）、线性回归分析（linear regression, LR）以及多元回归分析（multiple linear regression, MLR）的特点，可以有效地解决许多用普通多元回归难以解决的问题，尤其当变量多且存在多重相关性时，能有效分解和筛选出对因变量解释性最强的综合变量，建立的模型比普通回归分析更可靠。PLS 先提取称之为成分的新变量作为自变量，建立因变量与自变量间的线性组合关系，其系数由 PLS 计算确定，然后构建因变量的回归方程。

利用 PLS 算法建立模型，主成分数增加会提高模型精度，但主成分过多又会造成过度拟合现象，误差增加，因此，确定 PLS 模型的最佳主成分数非常关键。本研究通过交互验证法（cross validation）来计算预测残差平方和（predicive residual error sum of square, PRESS），PRESS 值越小则模型的预测能力越强，因此，依据 PRESS 的最

小值可确定最佳主成分数。PRESS 可表达为

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

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

PLS 算法和 PRESS 的具体做法详见文献[20]。PLS 和 PRESS 处理过程均由自编的 MATLAB 程序完成。

1.5 模型的评价

分别用建模集和验证集的样本，通过绘制 GPC 预测值与实测值间的 1：1 关系图进行模型评价，评价指标为决定系数  $R^2$  和均方根误差  $RMSE$ 。

2 结果与分析

2.1 GPC 分布

整理试验中测定的 GPC 数据，按籽粒样本中 GPC 的大小顺序排列，为增强所建预测模型的稳定性，在保证 GPC 最大值和最小值置于建模样本集的前提下，将 120 个样本按 3：1 的比例随机分成建模集和验证集。由表 2 可知，建模集和验证集 GPC 的变幅、平均值和标准差均相近，建模集和验证集样本具有较好的一致性。

表 2 建模集和验证集样本 GPC 分布  
Table 2 Distribution of grain protein content of wheat in calibration and validation datasets

样本集	样本数	变幅/%	平均值/%	标准差
建模集	90	9.76~13.56	11.96	0.99
验证集	30	9.81~13.29	11.36	0.92

2.2 遥感变量与 GPC 间的定量分析

建模集中 90 个样本的 GPC 与遥感变量间定量分析表明（表 3），GPC 与大多数遥感变量间都存在显著或极显著关系，其中与 *NDVI* 关系最密切，其次与 *NRI*，相关系数分别为 0.74 和 0.7，与宋晓宇等<sup>[7]</sup>研究结果一致，由于籽粒蛋白质来源于氮素运转和积累，因此也与 Serrano 等<sup>[2]</sup>利用遥感变量对叶片全氮含量的研究结果一致；多数植被指数与 GPC 间的相关性明显好于单波段；多数遥感变量两两间存在严重的多重相关性，特别是植被指数间更突出，其相关系数几乎介于 0.85 与 0.99 之间，说明用 PLS 算法建立模型比用传统统计算法和普通最小二乘法更为合理，可能会获得比较好的预测结果。

表 3 遥感变量与 GPC 相关系数  
Table 3 Correlation between remote sensing variables and grain protein content

	GPC	B1	B2	B3	B4	B5	B6	B7	NDVI	OSAVI	SAVI	SIPI	NDWI	NDWI2	PSRI	NRI	WI	RVI	TVI	NSI
B1	-0.42	1.00																		
B2	-0.35	0.97	1.00																	
B3	-0.53	0.89	0.96	1.00																
B4	0.52	-0.47	-0.63	-0.68	1.00															
B5	0.31	0.25	0.28	0.36	0.33	1.00														
B6	-0.39	0.66	0.38	0.28	0.06	0.28	1.00													
B7	-0.14	0.60	0.68	0.77	-0.18	0.83	0.34	1.00												
NDVI	0.74	-0.68	-0.84	-0.86	0.95	0.23	-0.02	-0.28	1.00											
OSAVI	0.60	-0.68	-0.84	-0.86	0.95	0.23	-0.02	-0.28	0.93	1.00										

接上页

	GPC	B1	B2	B3	B4	B5	B6	B7	NDVI	OSAVI	SAVI	SIPI	NDWI	NDWI2	PSRI	NRI	WI	RVI	TVI	NSI
SAVI	0.56	-0.68	-0.84	-0.86	0.95	0.23	-0.02	-0.28	0.92	0.95	1.00									
SIPI	0.67	-0.68	-0.74	-0.74	0.96	0.36	-0.05	-0.16	0.98	0.98	0.98	1.00								
NDWI	-0.27	-0.65	-0.75	-0.87	0.58	-0.56	-0.17	-0.86	0.94	0.89	0.98	0.92	1.00							
NDWI2	0.39	-0.36	-0.80	-0.91	0.57	-0.54	-0.24	-0.89	0.95	0.90	0.97	0.93	0.98	1.00						
PSRI	0.50	-0.66	-0.29	-0.20	0.71	0.66	0.37	0.29	0.83	0.92	0.98	0.90	0.95	0.94	1.00					
NRI	0.70	0.65	-0.85	-0.90	0.61	-0.36	-0.10	-0.71	0.72	0.72	0.72	0.89	0.89	0.85	0.96	1.00				
WI	-0.49	-0.66	0.79	0.89	-0.56	0.46	0.29	0.85	-0.86	-0.86	-0.87	-0.82	-0.90	-0.96	0.88	0.87	1.00			
RVI	0.59	0.63	-0.81	-0.82	0.92	0.30	0.58	-0.20	0.99	0.99	0.99	0.98	0.93	0.94	0.80	0.85	-0.82	1.00		
TVI	-0.66		0.76	0.75	-0.87	-0.37	0.43	0.10	-0.96	-0.96	-0.96	-0.95	-0.92	-0.92	0.81	0.87	0.81	0.99	1.00	
NSI	0.60	-0.63	-0.77	-0.77	0.95	0.34	0.39	-0.16	0.99	0.99	0.99	0.99	0.92	0.93	0.90	0.81	-0.82	0.99	0.97	1.00

注：样本数  $n=90$ ，其他同表 1。

2.3 PLS 模型

2.3.1 最佳主成分数目确定

PRESS 值越小，说明模型的预测能力越强。图 1 为用建模集样本得到的 PRESS 随主成分数的变化。起始时随着主成分数增加，PRESS 值不断减小，说明因主成分数偏少模型拟合不充分，即出现缺失拟合现象，直至主成分数为 5 时，PRESS 值最小，为 21.23；之后，随着主成分数继续增加，PRESS 值明显增加，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 PLS 模型的最佳主成分数为 5。

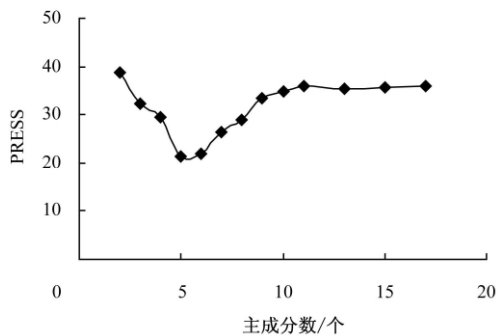


图 1 PRESS 随主成分数的变化

Fig.1 PRESS changes with the principal components

2.3.2 PLS 模型

基于 PLS 算法，以主成分数为 5 的 5 个植被指数即 NDVI、SIPI、RVI、NRI 和 PSRI 为自变量，以 GPC 为因变量，使用建模集样本和 2008-05-02 以及 2009-04-26 TM 影像构建的 GPC 预测模型为

$$GPC=4.596 \times NDVI+2.124 \times SIPI+2.719 \times RVI-2.109 \times NRI-1.642 \times PSRI+6.863 \quad (2)$$

利用该模型预测 GPC，将所得到的 GPC 预测值与实测值绘成 1:1 散点图，统计出最优直线回归方程及其  $R^2$  和 RMSE，图 2 为 GPC 模型预测能力的评价，建模集和验证集中的 GPC 预测值与实测值间的  $R^2$  均大于 0.6，RMSE 均小于 0.31，说明利用该 PLS 模型能较好地预测小麦 GPC，但 GPC 预测值多数高于实测值，尤其验证集样本，其可能原因是 GPC 测定时使用滴定的酸溶液体积偏小，造成实测值偏低。

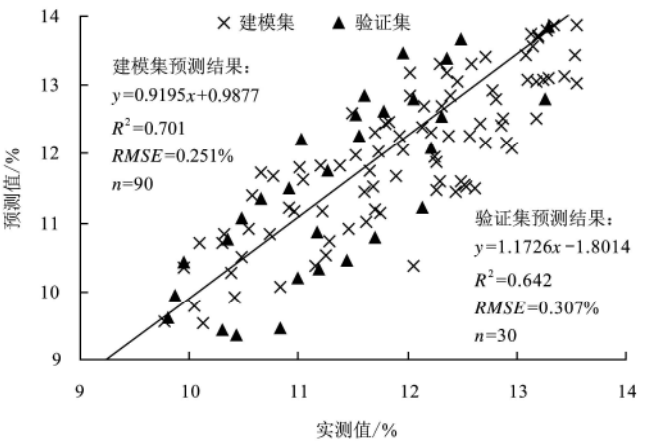


图 2 GPC 模型评价

Fig.2 Evaluation of grain protein content model

为了与 PLS 模型对比，又采用传统的 LR 和 PCR 建立 GPC 预测模型，并依据 GPC 预测值与实测值相关系数和 RMSE 对其评价。表 4 为 PLS、LR 和 PCR 模型预测验证集样本结果，PLS、LR 和 PCR 模型的 GPC 预测值与实测值相关系数分别为 0.802、0.669 和 0.727，RMSE 分别为 0.307、0.421 和 0.382，即 PLS 模型的相关系数大于 LR 和 PCR 模型，RMSE 小于 LR 和 PCR 模型，因此，PLS 模型对 GPC 的预测能力要好于 LR 和 PCR 模型。

表 4 PLS、LR 和 PCR 模型预测结果比较

Table 4 Comparison of predicted results with PLS, LR and PCR				
算法	主成分数/个	样本数/个	相关系数	RMSE/%
PLS	5	30	0.802	0.307
PCR	7	30	0.727	0.382
LR	0	30	0.669	0.421

依据上述分析，利用 2008-05-02 TM 影像生成 NDVI、SIPI、RVI、NRI 和 PSRI 图，通过一一解算和二值化掩膜，再叠加小麦种植数据去掉非小麦区后，结合行政边界矢量数据，基于 PLS 模型，得到江苏中部地区小麦 GPC 空间分布遥感预测图(图 3)，沿海地区 GPC 11.3%~11.6%，尤其大丰地区较为突出；沿江地区 11.3%~11.6%；兴化及周边地区 11.3%~11.6%和 11.6%~12.6%；GPC 高于 12.6%和低于 10.7%的地区极少。该结果与收获期外业调

查及当地农技部门提供的小麦 GPC 实际分布情况是一致的。

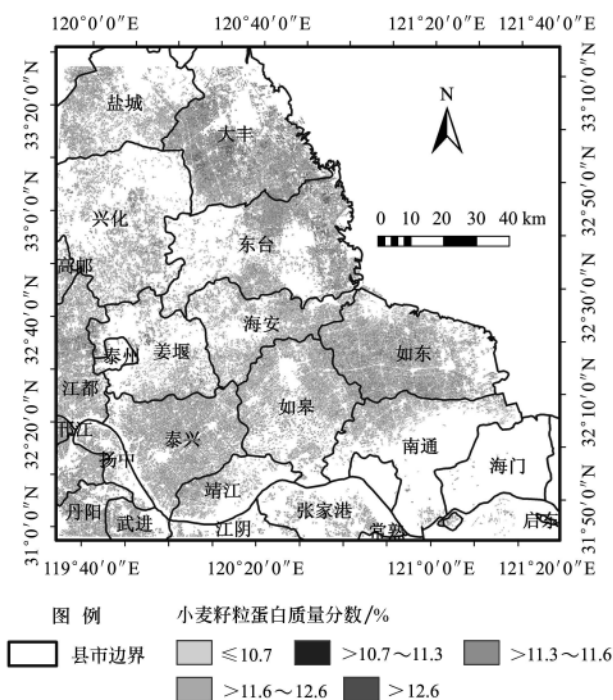


图3 江苏中部地区冬小麦 GPC 空间分布

Fig.3 Winter wheat grain protein content spatial distribution level in central Jiangsu Province

### 3 结论与讨论

1) GPC 与大多数遥感变量间存在密切关系，且多数遥感变量两两间存在严重的多重相关性，导致运用传统算法难以建立较高精度的 GPC 遥感预测模型。该研究利用 PLS 算法构建了以 *NDVI*、*SIPI*、*RVI*、*NRI* 和 *PSRI* 为自变量 GPC 遥感预测模型，经验证，*RMSE* 为 0.307%，小于传统算法中 LR 和 PCR 模型的 *RMSE*，对 2008 年江苏小麦 GPC 的预测结果与实际情况相符合，说明利用该 PLS 模型预测小麦 GPC 是可行的，且具有较高的精度。研究结果为利用航天航空遥感影像准确预测小麦品质提供了一种有效方法和技术支持。

2) 该研究缺少大于 13.6% 和小于 9.7% 的 GPC 样本，若增加 GPC 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 GPC 遥感预测模型，精度可靠，但是基于 TM 数据建立的，因此，该模型是否适用于其他遥感传感器数据，以及能否预测其他籽粒品质参数，有待于进一步研究。

3) 该研究没有将 PLS 算法与人工神经网络 (artificial neural network, ANN)、支持向量机 (support vector machines, SVM)、地统计学等算法进行比较，也没有考虑到小麦栽培影响因素，事实上，这些算法和因素对小麦 GPC 遥感预测结果存在较大的影响，有待进一步研究。

#### 【参 考 文 献】

[1] Kyu-Sung L, Warren B C, Robert E K, et al. Hyperspectral versus multispectral data for estimating leaf area index in

four different biomes[J]. Remote Sensing of Environment, 2004, 91(1): 508—520.

- [2] Serrano L, Filella I, Penuelas J. Remote sensing of biomass and yield of winter wheat under different nitrogen supplies[J]. Crop Science, 2000, 40: 723—731.
- [3] Vaesen S, Gilliams K, Nackaerts, et al. Ground-measured spectral signatures as indicators of ground cover and leaf area index: the case of paddy rice[J]. Field Crops Research, 2001, 69(1): 13—25.
- [4] 黄文江, 赵春江, 王纪华, 等. 红边参数在作物营养诊断和品质预报上的应用[J]. 农业工程学报, 2004, 20(6): 1—5.  
Huang Wenjiang, Zhao Chunjiang, Wang Jihua, et al. Application of red edge variables to nutrition diagnosis and grain quality forecast of winter wheat[J]. Transactions of the CSAE, 2004, 20(6): 1—5. (in Chinese with English abstract)
- [5] 王纪华, 黄文江, 赵春江, 等. 利用光谱反射率估算叶片和籽粒品质指标研究[J]. 遥感学报, 2003, 7(4): 277—284.  
Wang Jihua, Huang Wenjiang, Zhao Chunjiang, et al. The inversion of leaf biochemical components and grain quality indicators of winter wheat with spectral reflectance[J]. Journal of Remote Sensing, 2003, 7(4): 277—284. (in Chinese with English abstract)
- [6] Delgado J A, Ristau R J, Dillon M A, et al. Use of innovative tools to increase nitrogen use efficiency and protect environmental quality in crop rotations[J]. Communications in Soil Science and Plant Analysis, 2001, 32: 1321—1354.
- [7] 宋晓宇, 王纪华, 黄文江, 等. 变量施肥条件下冬小麦长势及品质差异遥感监测[J]. 农业工程学报, 2009, 25(9): 155—162.  
Song Xiaoyu, Wang Jihua, Huang Wenjiang, et al. Monitoring spatial variance of winter wheat growth and grain quality under variable-rate fertilization conditions by remote sensing data[J]. Transactions of the CSAE, 2009, 25(9): 155—162. (in Chinese with English abstract)
- [8] Dennis L, Wright J V, Philip R J. Managing Protein in Hard Red Spring Wheat with Remote Sensing[C]//The 6th Annual National Wheat Industry Research Forum, 2003, Hyatt Regency Albuquerque, New Mexico.
- [9] Liu L Y, Wang J J, Bao Y S, et al. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images[J]. International Journal of Remote Sensing, 2006, 27(4): 737—753.
- [10] Badri B B, Armando A A, Rob M K. Relating satellite imagery with grain protein content[C]//Proceedings of the Spatial Sciences Conference, 2003, 22—27, Canberra.
- [11] Zhao C J, Liu L Y, Wang J H, et al. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress[J]. International Journal of Applied Earth Observation And Geoinformation, 2005, 7(1): 1—9.
- [12] Reyniers M, Vrindts E, De B J, et al. 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.

- [13] Steddom K, Bredehoeft M W, Khan M, et al. Comparison of visual and multispectral radiometric disease evaluations of cercospora leaf spot of sugar beet[J]. *Plant Disease*, 2005, 89(2): 153—158.
- [14] 王圆圆, 陈云浩, 李京, 等. 利用偏最小二乘回归反演冬小麦条锈病严重度[J]. *国土资源遥感*, 2007(1): 57—61. Wang Yuanyuan, Chen Yunhao, Li Jing, et al. The application of partial least square to the inversion of severity of winter wheat stripe rust disease[J]. *Remote Sensing for Land & Resources*, 2007(1): 57—61. (in Chinese with English abstract)
- [15] 竞霞, 黄文江, 琚存勇, 等. 基于 PLS 算法的棉花黄萎病高空间分辨率遥感监测[J]. *农业工程学报*, 2010, 26(8): 229—235. Jing Xia, Huang Wenjiang, Ju Cunyong, et al. Remote sensing monitoring severity level of cotton verticillium wilt based on partial least squares regressive analysis[J]. *Transactions of the CSAE*, 2010, 26(8): 229—235. (in Chinese with English abstract)
- [16] 王纪华, 黄文江, 劳彩莲, 等. 运用 PLS 算法由小麦冠层反射光谱反演氮素垂直分布[J]. *光谱学与光谱分析*, 2007, 27(7): 1319—1322. Wang Jihua, Huang Wenjiang, Lao Cailian, et al. Inversion of Winter Wheat Foliage Vertical Distribution Based on Canopy Reflected Spectrum by Partial Least Squares Regression Method[J]. *Spectroscopy and spectral analysis*, 2007, 27(7): 1319—1322. (in Chinese with English abstract)
- [17] 赵祥, 刘素红, 王培娟, 等. 基于高光谱数据的小麦叶绿素含量反演[J]. *地理与地理信息科学*, 2004, 20(3): 36—39. Zhao Xiang, Liu Suhong, Wang Peijuan, et al. A Method for inverting chlorophyll content of wheat using hyperspectral[J]. *Geography and Geo-Information Science*, 2004, 20(3): 36—39. (in Chinese with English abstract)
- [18] Hansen P M, Schjoerring J K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression[J]. *Remote Sensing of Environment*, 2003, 86(1): 542—553.
- [19] Hung T N, Byun-Woo L. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression[J]. *European Journal of Agronomy*, 2006(24): 349—356.
- [20] 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)

## Predicting grain protein content in winter wheat based on TM images and partial least squares regression

Tan Changwei<sup>1</sup>, Wang Jihua<sup>2</sup>, Huang Wenjiang<sup>2</sup>, Wang Junchan<sup>1</sup>, Zhu Xinkai<sup>1</sup>, Guo Wenshan<sup>1\*</sup>

(1. Key Laboratory of Crop Genetics and Physiology of Jiangsu Province/Key Laboratory of Crop Physiology, Ecology and Cultivation in Middle and Lower Reaches of Yangtse River of Ministry of Agriculture, Yangzhou University, Yangzhou 225009, China;

2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100089, China)

**Abstract:** In order to further improve the accuracy of predicting winter wheat grain protein content (GPC) by remote sensing, the study analyzed the quantitative relationship between satellite remote sensing variables and GPC. Depending on the partial least squares regression (PLS), the multivariable remote sensing prediction model and the space level distribution map of winter wheat grain protein content were constructed. For the PLS model, the number of the best principal components was 5, and normalized difference vegetation index (NDVI), structure insensitive pigment index (SIPI), ratio vegetation index (RVI), nitrogenous reflection index (NRI) and plant senescence reflectance index (PSRI) were identified as the sensitive remote sensing variables for predicting GPC. The determination coefficient ( $R^2$ ) and the root mean square error (RMSE) between estimated value and measured value of GPC were 0.642 and 0.307%, respectively. The results indicate that PLS method can provide an effective way to improve the accuracy of predicting wheat grain quality at large scale by remote sensing data.

**Key words:** models, predictive control systems, proteins, landsat TM, winter wheat

谭昌伟, 等. 基于 TM 和 PLS 的冬小麦籽粒蛋白质含率预测 (图 3)

Tan Changwei, et al. Predicting grain protein content in winter wheat based on TM images and partial least squares regression (Figure 3)

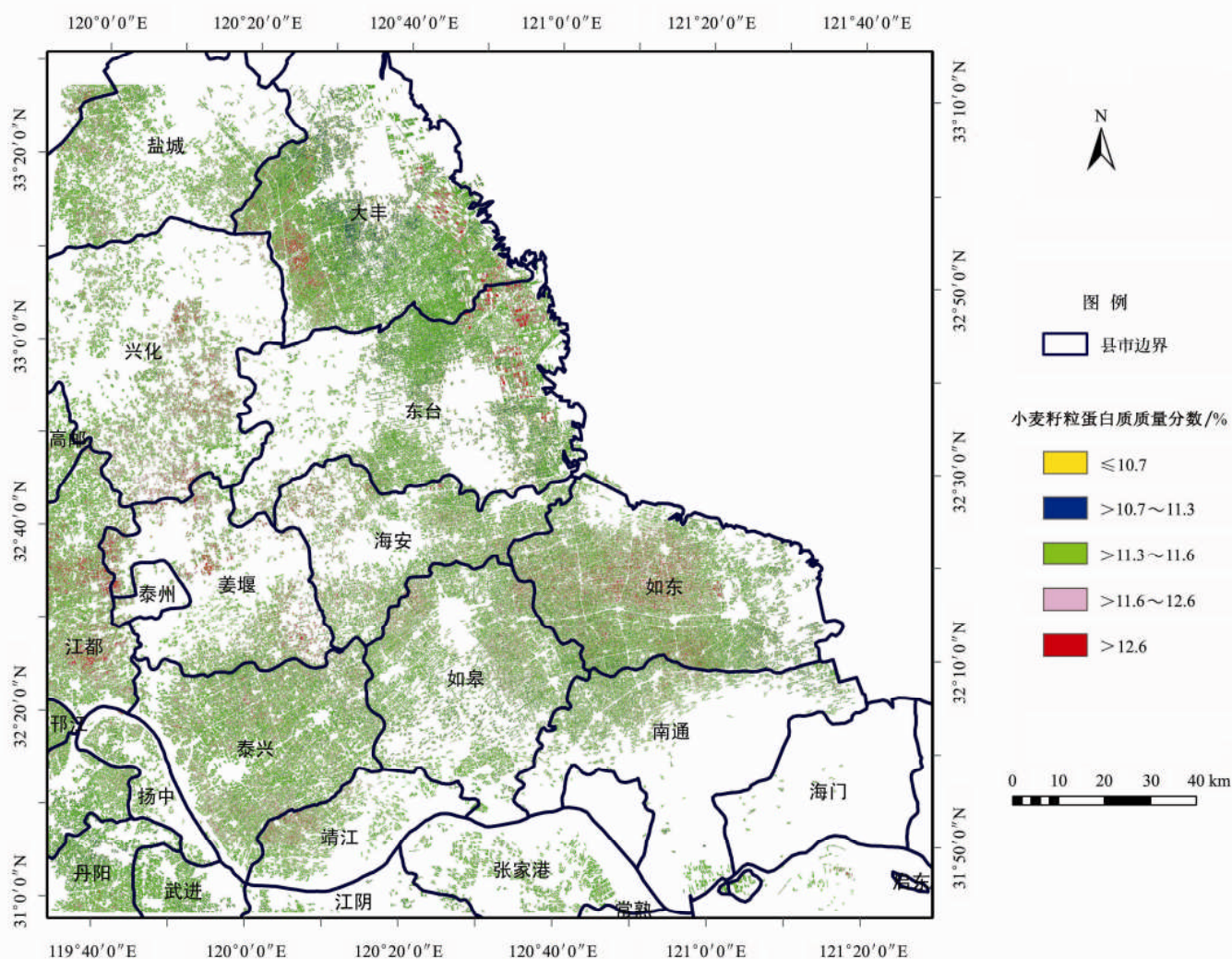


图 3 江苏中部地区冬小麦 GPC 空间分布

Fig.3 Winter wheat grain protein concent spatial distribution level in central Jiangsu Province