

基于开花期氮素营养指标的冬小麦籽粒 蛋白质含量遥感预测

屈莎^{1,2,3,4}, 李振海^{2,3,4}, 邱春霞¹, 杨贵军^{2,3,4},
宋晓宇^{2,3,4*}, 陈召霞^{2,3,4}, 刘畅^{1,2,3,4}

(1. 西安科技大学测绘科学与技术学院, 西安 710054; 2. 国家农业信息化工程技术研究中心, 北京 100097;
3. 农业部农业信息技术重点实验室, 北京 100097; 4. 北京市农业物联网工程技术研究中心, 北京 100097)

摘要: 籽粒蛋白含量 (grain protein content, GPC) 是衡量小麦品质的重要指标, 及时准确的预测小麦 GPC 有利于小麦的分类收割和分级存储。为了能够选择一个合适的氮素营养指标作为中间变量来反演小麦 GPC, 该文研究分别以开花期植株氮素累积量 (plant nitrogen accumulation, PNA)、植株氮素含量 (plant nitrogen content, PNC)、叶片氮素累积量 (leaf nitrogen accumulation, LNA) 和叶片氮素含量 (leaf nitrogen content, LNC) 4 个氮素营养指标为中间变量, 并运用支持向量机 (support vector machines, SVM) 算法实现 4 个氮素营养指标的估测, 最后构建及评价基于开花期“植被指数 (vegetation index, VI)-氮素营养指标 (nitrogen nutrition index, NNI)-GPC”模式的冬小麦 GPC 预测模型。结果表明: 1) 通过分析植被指数与氮素营养指标的相关性, 选择植被指数 MSAVI、PSRI、DVI、RDVI 和 GNDVI 作为氮素营养指标模型的构建变量; 2) 运用 SVM 方法构建的 VI-NNI 模型中 LNC 的建模精度与验证精度相对最优, 其建模决定系数 (coefficient of determination, R^2) 和验证集标准均方根误差 (normalized root mean squared error, nRMSE) 及验证标准化平均误差 (normalized average error, NAE) 分别为 0.820、9.553%、-1.4%, 验证结果稳定性较好; 3) 构建 NNI-GPC 模型中 PNC 的建模精度与验证精度相对最好, 其建模 R^2 和验证 nRMSE 及 NAE 分别为 0.653、9.843%、-0.3%; 4) 最终构建的 VI-NNI-GPC 模型中, 以开花期 PNC 为中间变量的模型建模及反演精度最好, 其建模 R^2 和验证 nRMSE 及 NAE 分别为 0.631、8.564%、-0.9%。以氮素营养指标为中间变量的 GPC 遥感反演是可行的, 并且比较 4 个氮素营养指标为中间变量反演 GPC, PNC 具有较高精度的预测结果, 为精确反演 GPC 提供一个可靠的依据, 具有一定的应用前景。

关键词: 遥感; 模型; 支持向量机; 籽粒蛋白质含量; 氮素营养指标; 冬小麦

doi: 10.11975/j.issn.1002-6819.2017.12.024

中图分类号: S127

文献标志码: A

文章编号: 1002-6819(2017)-12-0186-08

屈莎, 李振海, 邱春霞, 杨贵军, 宋晓宇, 陈召霞, 刘畅. 基于开花期氮素营养指标的冬小麦籽粒蛋白质含量遥感预测[J]. 农业工程学报, 2017, 33(12): 186-193. doi: 10.11975/j.issn.1002-6819.2017.12.024

<http://www.tcsae.org>

Qu Sha, Li Zhenhai, Qiu Chunxia, Yang Guijun, Song Xiaoyu, Chen Zhaoxia, Liu Chang. Remote sensing prediction of winter wheat grain protein content based on nitrogen nutrition index at anthesis stage[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(12): 186-193. (in Chinese with English abstract)

doi: 10.11975/j.issn.1002-6819.2017.12.024 <http://www.tcsae.org>

0 引言

小麦是世界上主要的粮食作物之一^[1], 全世界大约有 35% 的人口以小麦为主要粮食。随着经济技术的发展, 人民生活水平已经从温饱上升到品质的要求, 所以农产品的品质就更受到了广泛的关注。因此, 如何快速, 准确, 无损地预测小麦品质成为现在的研究热点。小麦籽粒蛋白含量 (grain protein content, GPC) 是小麦品质好坏的

主要衡量标准, 国内外研究学者直接利用植被指数反演 GPC 或利用“植被指数 (vegetation index, VI)-氮素营养指标 (nitrogen nutrition index, NNI)-GPC”间接模式构建了大量的 GPC 预测模型。Liu 等^[2]利用开花期和灌浆期的遥感影像构建植被指数与冬小麦的 GPC 反演模型, 得到较好的反演精度; 金秀良等^[3]通过新构建的比率植被指数并用于 GPC 的预测, 结果表明比率植被指数比单一的植被指数预测 GPC 的精度要高; Song 等^[4]研究利用冬小麦挑旗期 QuickBird 影像, 通过分析植被指数与 GPC 的相关性, 并选择植被指数 GNDVI 用于 GPC 的预测。谭昌伟等^[5]研究利用 TM 影像和偏最小二乘法来构建多植被指数 GPC 模型, 模型的决定系数 (R^2) 为 0.642, 均方根误差 (RMSE) 为 0.307%, 反演效果较为理想; 田永超等^[6]研究表明利用小麦抽穗期后期的比值植被指数 (ratio of vegetation index, RVI) 进行 GPC 的预测有较高的反演精度。基于植被指数直接反演 GPC, 其模型经验性太强,

收稿日期: 2016-11-14 修订日期: 2016-12-13

基金项目: 国家自然科学基金项目 (41371349; 41471285); 国家重点研发计划 (2016YFD0300603)

作者简介: 屈莎, 女 (汉族), 陕西西安人, 西安科技大学测绘科学与技术学院, 研究方向为作物品质遥感监测。Email: 1508273100@qq.com

*通信作者: 宋晓宇, 女, 湖北人, 北京农业信息技术研究中心副研究员, 博士, 研究方向为农业定量遥感技术研究与應用。

Email: songxy@nercita.org.cn

扩展应用性较差。Wang 等^[7-8]通过构建以叶片氮含量 (leaf nitrogen content, LNC) 为中间变量的“VI-NNI-PC”的 GPC 预测模型, 都得到了较好的反演效果。王纪华等^[9]利用与 LNC 极显著相关的 820~1 100 nm 波段光谱反射率和显著相关的 1 150~1 300 nm 波段的光谱反射率建立了与各生育期 LNC 的回归模型, 且显著性检验水平表明模型的可靠性, 最终选择开花期来反演 GPC。李映雪等^[10]研究表明, 小麦开花后期 14 d 比值植被指数 RVI (1220, 710) 能较好地反映 LNC, 而且此时的 GPC 与 LNC 达显著性相关, 因此, 以花后 14 d 的 LNC 为中间变量能较好地间接反演 GPC。Huang 等^[11]研究利用植株总氮含量 (plant nitrogen content, PNC) 为中间变量构建 GPC 的预测模型, 得到了较好的模型反演精度。基于“VI-NNI-GPC”模式构建的 GPC 模型, 其氮素营养指标以叶片氮素含量选择较多, 机理性和解释性较单纯的“VI-GPC”模式有所提高, 但是国内外在构建大量“VI-NNI-GPC”模型的研究中不同的研究者选择的中间变量不同, 大都只是选择了单一的、经验的一种, 并没有对这些氮素营养指标中究竟哪一种作为中间变量会得到更好的结果作深入的研究, 并且试验数据也大都只是单一年限, 模型稳定性较差。所以本文就这两点展开研究。

本文拟采用支持向量机 (support vector machines, SVM) 方法, 充分利用光谱信息, 利用多植被指数反演 4 个氮素营养指标: 植株氮素累积量 (plant nitrogen accumulation, PNA)、植株氮素含量 (plant nitrogen content, PNC)、叶片氮素累积量 (leaf nitrogen accumulation, LNA) 和叶片氮素含量 (leaf nitrogen content, LNC), 分别构建 4 个氮素营养指标与 GPC 的关系模型, 耦合 2 个过程模型, 并采用其建模决定系数 (coefficient of determination, R^2) 和验证集标准均方根误差 (normalized root mean squared error, nRMSE) 及验证标准化平均误差 (normalized average error, NAE) 评价指标来评价 4 个“植被指数-氮素营养指标-GPC”模式的 GPC 模型, 以期为冬小麦籽粒蛋白质遥感精准预测提供相关理论依据和技术支持。

1 材料与方法

1.1 试验设计

试验于 2008—2011, 2012—2015 年度在国家精准农业示范基地进行, 基地位于北京市昌平区小汤山镇 (40°10'31"N~40°11'18"N, 116°26'10"E~116°27'05"E), 土壤类型为潮土, 且在 30 cm 以上的土壤中硝态氮、全氮、有效磷、有机质、速效钾质量分数各为 3.16~14.82 mg/kg, 1.0~1.2 g/kg, 3.14~21.18 mg/kg, 15.8~20.0 g/kg, 86.83~120.62 mg/kg。

试验一: 2008-2009 年度试验为品种的完全随机试验, 3 个冬小麦品种分别为农大 195、京 9428 和烟农 19, 没有重复, 总计 3 小区。播种期为 2008 年 9 月 28 日, 肥料水平为同一氮素水平的冬小麦试验, 底肥为 N 44.62 kg/hm², P₂O₅ 60 kg/hm², K₂O 76.5 kg/hm², 拔节期追 N 157.78 kg/hm², 灌溉量为 215 mm, 其他按照田间实际管理操作。

试验二: 2009—2010 年度试验为品种与施肥量的完全随机试验 A 和品种与播期完全随机的试验 B。试验 A 包括 3 个冬小麦品种: 农大 195、京 9428 和京东 13; 4 个氮肥施用量: (N₁) 44.62 kg/hm²、(N₂) 72.68 kg/hm²、(N₃) 109.02 kg/hm²、(N₄) 134.78 kg/hm², 播种期为 2009 年 9 月 25 日。试验 B 包括 3 个冬小麦品种 (同试验 A), 3 个播种期: 2009 年 9 月 25 日, 10 月 5 日和 10 月 15 日; 氮肥施用量为 N 109.02 kg/hm²。重复 2 个, 总计 36 小区。肥料水平为底肥为 N 56.12 kg/hm², P₂O₅ 60 kg/hm², K₂O 76.5 kg/hm², 灌溉量为 210 mm, 其他按照田间实际管理操作。数据利用过程中首先将重复求平均再进行数据分析。

试验三: 2010-2011 年度试验为品种完全随机试验。3 个冬小麦品种: 农大 195、京 9428 和烟农 19。没有重复, 总计 3 小区。播种期为 2010 年 10 月 3 日; 肥料水平为底肥施氮量为 N 114.08 kg/hm², P₂O₅ 60 kg/hm², K₂O 76.5 kg/hm², 拔节期追施 N 117.76 kg/hm², 灌溉量为 240 mm, 其他按照田间实际管理操作。

试验四: 2012—2013 年度试验为品种和氮肥水平的完全随机试验。4 个冬小麦品种: 农大 211、京 9843、中优 206 和中麦 175; 4 个氮肥施用量: (N₀) 0、(N₁) 104.88 kg/hm²、(N₂) 209.76 kg/hm²、(N₃) 419.25 kg/hm²。重复 2 个, 总计 32 小区。播种期为 2012 年 9 月 28 日; 肥料水平为除各处理施氮量外, 底肥为 P₂O₅ 60 kg/hm² 和 K₂O 76.5 kg/hm², 灌溉量为 187 mm; 其他按照田间实际管理操作。

试验五: 2013—2014 年度试验为品种、施氮量和灌溉量的正交试验, 2 个冬小麦品种: 京 9843 和中麦 175; 4 个氮肥施用量: (N₀) 0、(N₁) 90.16 kg/hm²、(N₂) 180.32 kg/hm²、(N₃) 269.56 kg/hm², 3 个灌溉量: 雨养、正常灌溉 (171 mm) 和过量灌溉 (317 mm)。重复 3 个, 总计 48 小区。播种期为 2013 年 10 月 4 日, 肥料水平为除各处理施氮量外, 底肥为 P₂O₅ 60 kg/hm² 和 K₂O 76.5 kg/hm²; 其他按照田间实际管理操作。数据利用过程中首先将重复求平均再进行数据分析。

试验六: 2014—2015 年度试验为品种、施氮量和灌溉量的正交试验, 各处理与 2013—2014 年试验完全相同。3 个重复, 48 个小区, 播种期为 2014 年 10 月 7 日, 除各处理施氮量外, 底肥为 P₂O₅ 60 kg/hm² 和 K₂O 76.5 kg/hm²; 其他按照田间实际管理操作。数据利用过程中首先将重复求平均再进行数据分析。

1.2 试验数据获取

1.2.1 冠层光谱测定

光谱测定分别在开花期 (2009 年 5 月 12 日、2010 年 5 月 19 日、2011 年 5 月 17 日、2013 年 5 月 22 日、2014 年 5 月 7 日和 2015 年 5 月 14 日) 对小麦的冠层光谱进行测定, 地面高光谱采用 Fieldspec FR2500 型野外光谱辐射谱仪 (ASD, USA), 光谱范围为 350~2 500 nm, 350~1 000 nm 光谱分辨率为 1.4 nm, 1 000~2 500 nm 为 2 nm, 光谱重采样间隔为 1 nm, 测量时探头垂直于冠层上方 1 m 左右进行测定, 测定前后均用白板进行标定, 测量时天气选择晴朗无云的时段, 时间为 10:00~14:00, 每个小区测量 20 次, 取平均作为该处理的冠层光谱。

1.2.2 氮素营养指标的获取

在每个小区获取光谱信息的同时在每个小区选取具有代表性的20个小麦茎,茎叶分离,分器官装入纸袋,烘箱105℃杀青30min,用烘箱75℃烘24~48h至恒量,称量测各器官的干物质量,最后用凯氏定氮仪(Buchi B-339, Switzerland)测量各个器官的全氮含量。以下分别是4个氮素营养指标的数据测量方法:

$$\text{PNA}(\text{kg}/\text{hm}^2) = N1 \cdot 10B1 + N2 \cdot 10B2 + N3 \cdot 10B3 \quad (1)$$

$$\text{PNC}(\%) = (N1 \cdot B1 + N1 \cdot B1 + N1 \cdot B1) / (B1 + B2 + B3) \quad (2)$$

$$\text{LNA}(\text{kg}/\text{hm}^2) = N1 \cdot 10B1 \quad (3)$$

$$\text{LNC}(\%) = N1 \quad (4)$$

其中, $N1$ (%)、 $N2$ (%)、 $N3$ (%) 分别为叶片氮含量、

茎氮含量、穗氮含量; $B1$ (g/m^2)、 $B2$ (g/m^2)、 $B3$ (g/m^2) 分别为叶片生物量、茎生物量、穗生物量。

1.2.3 籽粒蛋白质含量

小麦成熟后,在每个小区取1m²的小麦,晒干脱粒,利用 Infratec TM 1241 近红外谷物分析仪(FOSS, 丹麦)来测定冬小麦 GPC, 在本次试验中6a GPC 数据的范围为11.14%~19.30%, 建模集 GPC 的范围为11.14%~19.30%, 验证集 GPC 的范围为11.60%~18.90%。

1.3 方法

1.3.1 植被指数选择

根据冬小麦的光谱特征以及前人的研究成果^[12-13]选用14个常见的植被指数选择用于反演4个氮素营养指标的待选变量。

表1 植被指数的计算公式
Table 1 Vegetation index calculation formula

植被指数 Vegetation index	名称 Name	公式 Formula	参考文献 References
NDVI	归一化植被指数 Normalized difference vegetation index	$(R_{800} - R_{670}) / (R_{800} + R_{670})$	[14]
RDVI	重归一化植被指数 Re-normalized difference vegetation index	$\sqrt{\text{NDVI} \cdot \text{DVI}}$	[15]
SAVI	土壤调节植被指数 Soil-adjusted vegetation index	$1.5(R_{800} - R_{670}) / (R_{800} + R_{670} + 0.5)$	[16]
OSAVI	调整土壤植被指数 Optimized soil-adjusted vegetation index	$1.16(R_{800} - R_{670}) / (R_{800} + R_{670} + 0.16)$	[17]
MSAVI	修正土壤调整植被指数 Modified soil-adjusted vegetation index	$0.5 \left[\frac{2R_{800} + 1 - \sqrt{2(R_{800} + 1)^2 - 8(R_{800} - R_{670})}}{2} \right]$	[18]
NRI	氮反射指数 Nitrogen reflectance index	$(R_{570} - R_{670}) / (R_{570} + R_{670})$	[19]
GNDVI	绿色归一化植被指数 Normalized green difference vegetation index	$(R_{750} - R_{550}) / (R_{750} + R_{550})$	[20]
NPCI	归一化色素叶绿素植被指数 Normalized pigment chlorophyll vegetation index	$(R_{430} - R_{680}) / (R_{430} + R_{680})$	[21]
SIPI	结构不敏感指数 Structure intensive pigment index	$(R_{800} - R_{445}) / (R_{800} + R_{680})$	[22]
PSRI	植被衰减指数 Plant senescence reflectance index	$R_{705} / (R_{717} + R_{491})$	[23]
DVI	差值植被指数 Difference vegetation index	$R_{800} - R_{670}$	[24]
RVI	比值植被指数 Ratio vegetation index	R_{800} / R_{670}	[25]
NDCI	归一化差值叶绿素指数 Normalized difference chlorophyll index	$(R_{708} - R_{665}) / (R_{708} + R_{665})$	[26]
PPR	植被色素比率 Plant pigment ratio	$(R_{550} - R_{450}) / (R_{550} + R_{450})$	[27]

注: R 分别为各自波长处的反射率值。

Note: R are the reflectance values at the respective wavelengths respectively.

1.3.2 支持向量机回归

SVM 是机器学习反演氮素含量的一个研究热点,可以提高光谱数据的智能化与自动化^[28-29]。SVM 回归理论是在寻找一个最优的分类面使得所有的训练样本离该最优的分类面的误差最小。遵循统计学习中的结构风险最小原理,使错误的概率的上界最小化,使实际输出与理想输出之间的偏差最小。近几年, SVM 回归在各个方面的应用^[30-32]也是越来越广泛。本研究中以 MSAVI、PSRI、DVI、RDVI、GNDVI 作为自变量,4个氮素营养指标分别为因变量运用 SVM 回归进行训练和预测。

本研究的 SVM 回归是在 MATLAB 的 Libsvm^[33]软件包中实现回归和预测的,此软件包为台湾大学林智仁副教授等设计的用于模式识别和回归的,并且提供了源代码、很多参数以及交互检验功能,该软件可以解决很多的回归问题,本次研究核函数采用 RBF 核函数,利用交叉验证法寻找最佳参数 C (惩罚因子) 和参数 g (RBF

核函数中的方差), 利用最佳的参数训练模型。

1.3.3 统计分析

本次试验数据共计88个,随机抽取约2/3的样本($n=58$)用于建模,约1/3的样本($n=30$)用于模型验证。对于模型的精度评定采用的评价指标为建模集决定系数(coefficient of determination, R^2)及验证集标准均方根误差(Normalized root mean squared error, nRMSE)及标准化平均误差(Normalized average error, NAE)作为评价模型的优劣;nRMSE在描述验证模型的精度一般都是给定差异显著性界限,例如 nRMSE<10%为无差异,10%≤nRMSE<20%为较小差异,20%≤nRMSE<30%为中等,nRMSE≥30%为较大差异^[34-38]。其计算公式分别为

$$\text{nRMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} / \bar{y}_i \quad (5)$$

$$NAE = \frac{\bar{y}_j - \bar{y}_i}{\bar{y}_i} \quad (6)$$

式中 y_i (%) 为冬小麦籽粒蛋白质含量实测值, y_j (%) 为预测值, \bar{y}_i (%) 为实测值的平均值, \bar{y}_j (%) 为预测值的平均值, n 为样本数。

2 结果与分析

2.1 植被指数与各氮素营养指标的相关性分析

如表 2 所示, 各个植被指数与 PNA、PNC、LNA、LNC 4 个氮素营养指标之间的相关性基本都达到了极显著相关水平 ($P < 0.01$)。在选取合适的植被指数进行 4 个氮素营养指标的反演时, 其选取原则为: 综合植被指数与 4 个氮素营养指标的相关性, 选取每个氮素营养指标相关性最好的植被指数, 如与 PNA 相关性最好的植被指数为 MSAVI, 再根据每个氮素营养指标与植被指数相关性排序, 选取在 4 个氮素营养指标中相关性排名都靠前的且共有的植被指数, 根据前人研究经验本试验最终选择前 5 个植被指数来反演 4 个氮素营养指标, 因此本文最终选取的植被指数为 MSAVI、PSRI、DVI、RDVI、GNDVI。

表 2 各植被指数与氮素营养指标间的相关性 ($n=58$)
Table 2 Correlation between vegetation index and nitrogen nutrition index ($n=58$)

植被指数 Vegetation index	PNA	PNC	LNA	LNC
NDVI	0.640**	0.522**	0.655**	0.636**
SAVI	0.733**	0.603**	0.757**	0.718**
OSAVI	0.706**	0.581**	0.725**	0.695**
MSAVI	0.750**	0.626**	0.771**	0.718**
GNDVI	0.714**	0.586**	0.723**	0.721**
NRI	0.509**	0.473**	0.533**	0.425**
SIPI	-0.588**	-0.462**	-0.609**	-0.612**
PSRI	-0.719**	-0.677**	-0.726**	-0.705**
DVI	0.739**	0.606**	0.773**	0.724**
RVI	0.721**	0.672**	0.719**	0.616**
NPCI	0.646**	0.632**	0.659**	0.584**
NDCI	0.429**	0.425**	0.473**	0.416**
PPR	-0.135 ^{Ns}	-0.155 ^{Ns}	-0.094 ^{Ns}	-0.192 ^{Ns}
RDVI	0.727**	0.597**	0.752**	0.714**

注: **为在 0.01 显性相关; Ns 表示不显著; n 为样本数量。下同。
Note: ** represents significant at the 0.01; Ns represents no significant; n is the number of samples. The same below.

2.2 氮素营养指标的遥感反演

本研究将 SVM 回归应用在氮素营养指标的反演中, 利用 5 个植被指数来分别反演 4 个氮素营养指标为非线性回归, 利用核函数, 将氮素营养指标的预测转换为高维空间中的线性问题^[39]。本研究主要是利用 MATLAB 的 Libsvm 软件包实现回归和预测的。其中用户需要给出两个重要的参数 C (惩罚因子) 和 g (核函数中的方差) 的范围用于模型的训练, 本试验中 C 和 g 的范围为 $[-8, 8]$, 并采用交叉验证法以 0.5 为步长来搜索 4 个氮素营养指标的最佳参数值, 从而实现对整个训练集进行训练获取最优预测结果的。根据所选的 5 个植被指数构建氮素营养

指标的反演模型, 其结果如图 1 所示: 利用植被指数反演的 4 个氮素营养指标, 建模集 R^2 最大的 PNA 为 0.872, 最小的 LNA 为 0.668, 整体都达到差异极显著水平, 所以在模型的构建上, 4 个氮素营养指标都可以用植被指数来反演, 且都能较好的预测植株各个氮素营养指标。根据验证集检验精度 nRMSE, 精度最高的 LNC 为 9.553%, 属无差异, PNC 为 15.796% 有较小差异, LNA 为 27.189% 属于中等差异, PNA 为 31.030% 属于较大差异。根据 NAE 显示, 误差最小的 PNC 为 -1.0%, 误差最大的 LNC 为 -1.4%, 各个模型在模拟时都会产生相应较小的偏差, 除冠层 PNA 的模型预测值高估于实测值, 其他 3 个氮素营养指标都相应的会低估于实测值。综合各个建模集及验证集的精度验证指标, 在 SVM 构建植被指数反演各个氮素营养指标的模型中, LNC 的建模精度 R^2 较高, nRMSE 较小, NAE 偏差较小, LNC 的模型较稳定。相比较于其他 3 个氮素营养指标, 应用遥感数据进行 LNC 模型的构建精度较高。

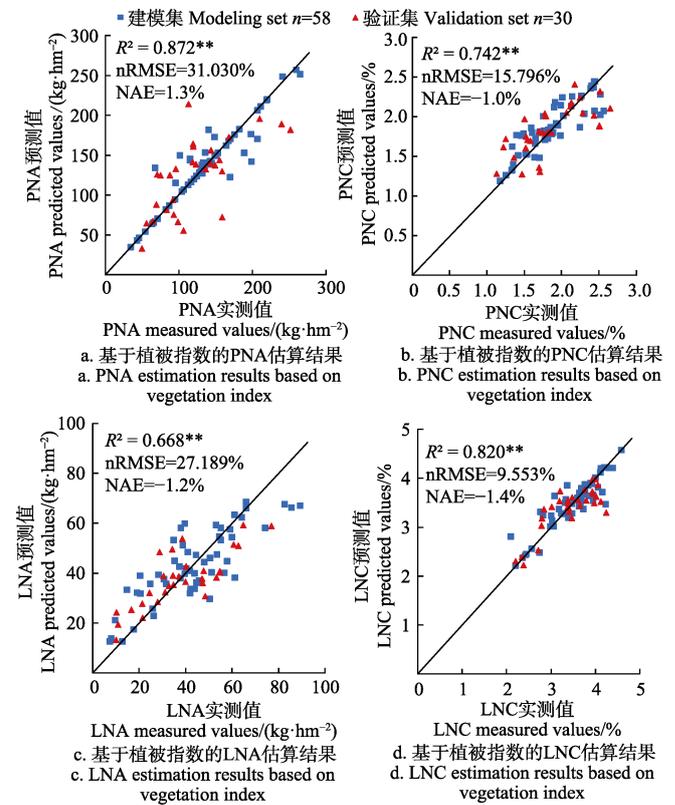


图 1 氮素营养指标的实测值与预测值的关系
Fig.1 Relationship between measured values and predicted values of nitrogen nutrition index

2.3 氮素营养指标与小麦 GPC 模型构建

由表 3 可以看出, 开花期 4 个氮素营养指标与冬小麦 GPC 的相关系数均大于 0.6, 达到差异极显著水平。比较 4 个氮素营养指标与 GPC 的相关性可知: PNC 与冬小麦 GPC 的相关性最高; GPC 与开花期 LNA 的相关性高于 LNC, 其原因可能是 LNA 考虑到了叶片干物质质量, 这一结果与薛丽红等的研究结果一致^[40]。从而利用 4 个氮素营养指标与冬小麦籽粒蛋白质含量构建线性模型。且建模精度及验证精度如图 2 所示。

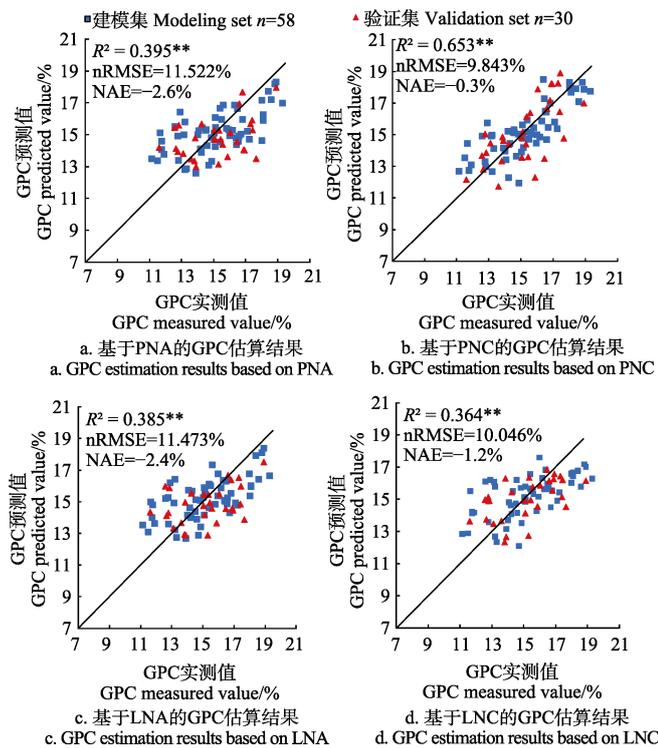
表3 氮素营养指标与冬小麦籽粒蛋白含量间的相关关系及回归模型 ($n=58$)

Table 3 Relationship between nitrogen nutrition index and winter wheat grain protein content and regression model ($n=58$)

氮素营养指标的 GPC 模型 GPC model of nitrogen nutrition index	相关系数 Correlation coefficient r	回归模型 Regression model
GPC _{PNA}	0.628**	$y = 0.025x + 11.735$
GPC _{PNC}	0.808**	$y = 4.727x + 6.386$
GPC _{LNA}	0.621**	$y = 0.069x + 12.194$
GPC _{LNC}	0.603**	$y = 2.209x + 7.472$

注: x 表示氮素营养指标, y 表示小麦 GPC。

Note: x represents nitrogen nutrition index, y represents wheat GPC.



注: PNA-GPC 表示 PNA 构建 GPC 模型, 同 PNC-GPC、LNA-GPC、LNC-GPC。
Note: PNA-GPC represents PNA construction GPC model, the same with PNC-GPC、LNA-GPC、LNC-GPC.

图2 氮素营养指标—冬小麦籽粒蛋白含量的实测值与预测值的关系

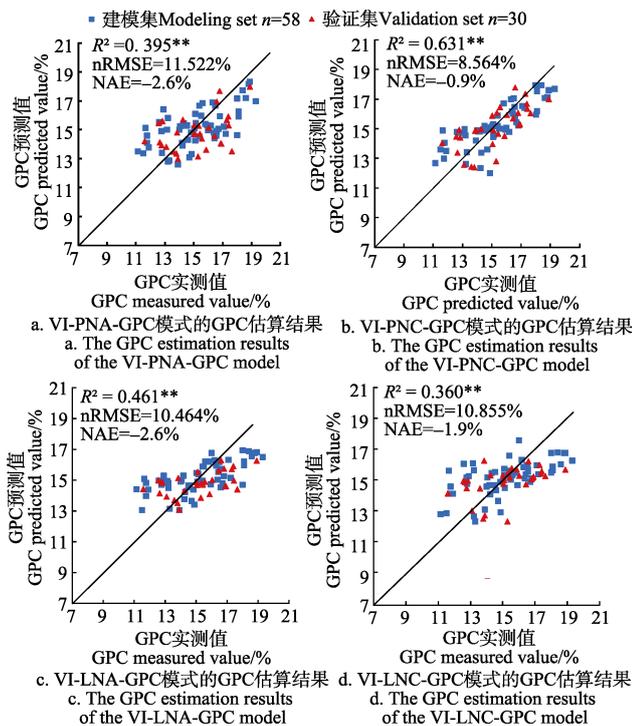
Fig.2 Relationship between measured values and predicted values of NNI-GPC model

本研究中冬小麦 GPC 的变幅范围为 11.14%~19.30%, 其变幅范围较大, 有利于各个氮素营养指标建立蛋白质的预测模型。比较冬小麦开花期 4 个氮素营养指标建模集和验证集的各个精度评价指标可知 (图 2), 建模集精度 R^2 , 最大 PNC 为 0.653, 最小的 LNC 为 0.364; 整体 4 个参数的建模精度都达到差异极显著水平, 所以, 以上 4 个氮素营养指标都可以用来反演冬小麦 GPC。验证集精度指标 nRMSE, PNC 精度最高为 9.843% 属无差异, LNC 为 10.046% 属于较小差异, LNA 为 11.473% 属于较小差异, 精度最小的 PNA 为 11.522% 属于较小差异; 验证精度 NAE, 4 个参数反演的验证都是小于 -2.6%, 偏差较小, 且相对于实测值有个较小的低估。

由此, 综合统计评价指标, PNC 较其他 3 个氮素营养指标具有较高 GPC 预测结果, 其原因与冬小麦 GPC 的形成因素有重要的关系: 第一: 冬小麦 GPC 形成主要受两方面因素的影响, 分别为小麦开花前存储于植株冠层部分 (茎、叶、穗) 的氮化合物的再运转 (约占籽粒总氮素累积量的 70%~80%), 和开花后植株对土壤中氮素的吸收与同化 (约占籽粒总氮素累积量的 20%~30%), 由此主要影响小麦 GPC 的因素在于冠层的氮素的存储运转^[41]。所以冠层氮素信息 (PNA 和 PNC) 与 GPC 的关系较叶片氮素信息 (LNA 和 LNC) 与 GPC 的关系, 一致性更好; 第二: 冬小麦 GPC 是冬小麦籽粒氮素的含量与籽粒的产量的比值计算得到的, 这与植株中的氮和碳的形成与累积有很大的关系; 而由 PNA、PNC、LNA、LNC 的计算公式 (1)、(2)、(3)、(4) 比较可知, LNA 仅仅表示了植株叶片的氮累积, LNC 只表示了叶片的氮含量, 不能代表整体的冠层信息, 而 PNA 只是表示了整个植株冠层的氮素累积量, 并没有涉及到植株的碳代谢, PNC 为植株氮素的累积量与植株生物量的比值, 这不仅考虑到了整个植株的氮累积量也涉及到了植株的碳代谢, 这与小麦籽粒蛋白质的含量相对应。所以 PNC 在反演冬小麦籽粒蛋白含量时建模精度和验证精度都明显优于 PNA、LNA 和 LNC。

2.4 “植被指数—氮素营养指标—籽粒蛋白含量”的 GPC 定量反演

根据“植被指数—氮素营养指标—籽粒蛋白含量”之间的定量关系来构建综合模型。SVM 方法反演的氮素营养指标是根据确定最佳反演因子来确定训练模型的最佳预测结果, 所以在构建综合模型时, 是利用 SVM 方法预测出来的预测值作为输入值带入氮素营养指标构建的蛋白质预测模型中 (表 3), 从而获得最终的综合预测结果。其反演结果如图 3。综合以 4 个氮素营养指标为中间变量构建的模型, 其建模集精度 R^2 PNC (0.631) > LNA (0.461) > PNA (0.395) > LNC (0.360), 都达到差异极显著水平, 以 PNC 的建模精度最高; 4 个经植被指数反演的氮素营养指标都可以用于对冬小麦 GPC 的预测; 验证集检验精度 nRMSE PNC (8.564%) < LNA (10.464%) < LNC (10.855%) < PNA (11.522%) 都达较小差异及以上, 而 PNC 的模型稳定性较其他 3 个中间变量较好; 验证集验证精度 NAE PNC (-0.9%) < LNC (-1.9%) < PNA (-2.6%) = LNA (-2.6%)。经过中间变量 4 个氮素营养指标来反演蛋白含量都是有较小的低估, 且以 PNC 的误差最小。相比较于其他 3 个中间变量而言, PNC 都优于其他 3 个氮素营养指标, 所以, PNC 能更精准的预测冬小麦 GPC。其可能的原因主要是: 在植被指数反演 4 个氮素营养指标的建模精度差异不大, 而在氮素营养指标反演冬小麦 GPC 时, PNC 的建模精度明显高于其他 3 个氮素营养指标 (具体原因参见 2.3 节), 所以氮素营养指标反演冬小麦 GPC 的构建精度对于构建综合模型影响更大。因此, 以 PNC 作为中间变量是综合模型较优的选择。



注: VI-PNA-GPC 表示以 PNA 为中间变量的 GPC 反演模型, 同 VI-PNC-GPC、VI-LNA-GPC、VI-LNC-GPC。

Note: VI-PNA-GPC represents the model of GPC with PNA as the intermediate variable; the same with VI-PNC-GPC、VI-LNA-GPC、VI-LNC-GPC.

图 3 植被指数—氮素营养指标—籽粒蛋白含量的实测值与预测值的关系

Fig.3 Relationship of measured and predicted value of VI-NNI-GPC model

3 结论

本文利用冬小麦开花期构建“植被指数-氮素营养指标-GPC (VI-NNI-GPC) 的 GPC 反演模型; 首先利用支持向量机 (support vector machines, SVM) 的方法构建多植被指数分别反演 4 个氮素营养指数 (植株氮素累积量 (plant nitrogen accumulation, PNA)、植株氮素含量 (plant nitrogen content, PNC)、叶片氮素累积量 (leaf nitrogen accumulation, LNA) 和叶片氮素含量 (leaf nitrogen content, LNC)), 然后以 4 个氮素营养指标为中间变量, 分别构建冬小麦 GPC 反演模型, 比较这 4 个模型反演冬小麦 GPC 的最佳模型。本研究结果表明:

1) 通过所选的植被指数与 4 个氮素营养指标的相关性分析可知, 氮素营养指标与大多数植被指数都能达到极显著相关, 根据相关性分析结果选择植被指数: 修正土壤调整植被指数 (MSAVI)、植被衰减指数 (PSRI)、差值植被指数 (DVI)、重归一化植被指数 (RDVI) 和绿色归一化植被指数 (GNDVI) 作为氮素营养指标的模型构建变量。

2) 利用多植被指数和 SVM 方法反演氮素营养指标, 各个氮素营养指标的都有较高建模精度和验证精度, 所以, 利用多植被指数和 SVM 反演氮素营养指标是可行的; 且 LNC 的 R^2 为 0.820、 $nRMSE$ 为 9.553% 和 NAE 为 -1.4%, 综合考虑, 相比较于其他 3 个氮素营养指标, 为较优的一个反演模型。

3) 4 个氮素营养指标构建蛋白质的线性回归模型中, PNC 的 R^2 为 0.653、 $nRMSE$ 为 9.843% 和 NAE 为 -0.3%, 为 4 个氮素营养指标中建模及验证精度最高的一个, 所以在 GPC 反演的模型中以 PNC 反演 GPC 为最优的选择。

4) 比较 4 个氮素营养指标作为中间变量分别构建蛋白质反演模型的精度, 可知, 利用氮素营养指标作为中间变量反演蛋白质是可行的, 且 PNC 的 R^2 为 0.631、 $nRMSE$ 为 8.564% 和 NAE 为 -0.9%, 在建模及验证精度都要高于其他 3 个氮素营养指标作为中间变量的模型。所以最终在构建 GPC 的综合模型中, 以 PNC 为中间变量的综合模型是相对最优的选择。

本文主要利用 SVM 的方法构建的植被指数反演 4 个氮素营养指标, 最终反演 GPC, 没有将其与其他方法 (如偏最小二乘等) 做比较, 也没有考虑到基因、温度、气象、土壤等因素对冬小麦籽粒蛋白含量的影响, 实际上, 这些因素对小麦籽粒蛋白含量都有很大的影响, 所以, 今后需要加入生态因子, 运用不同的反演方法进行进一步验证, 从而得到最优蛋白含量预测模型。

[参考文献]

- [1] 丛振涛, 王舒展, 倪广恒, 等. 气候变化对冬小麦潜在产量影响的模型模拟分析[J]. 清华大学学报: 自然科学版, 2008, 48(9): 1426—1430.
Cong Zhentao, Wang Shuzhan, Ni Guangheng, et al. Simulations of the impact of climate change on winter wheat production[J]. J Tsinghua Univ (Sci & Tech), 2008, 48(9): 1426—1430. (in Chinese with English abstract)
- [2] Liu Liangyun, Wang Jihua, Bao Yansong, 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.
- [3] 金秀良, 徐新刚, 李振海, 等. 基于新型植被指数对冬小麦蛋白质含量的估算研究[J]. 光谱学与光谱分析, 2013, 33(9): 2541—2545.
Jin Xiuliang, Xu Xingang, Li Zhenhai, et al. Estimation of winter wheat protein content based on new indexes[J]. Spectroscopy and Spectral Analysis, 2013, 33(9): 2541—2545. (in Chinese with English abstract)
- [4] Song Xiaoyu, Wang Jihua, Huang Wenjiang. Winter wheat growth and grain protein uniformity monitoring through remotely sensed data[J]. Proceedings of SPIE-The International Society for Optical Engineering, 2010, 7824: 7824G-1—7824G-8.
- [5] 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)
- [6] 田永超, 朱艳, 曹卫星, 等. 利用冠层反射光谱和叶片 SPAD 值预测小麦籽粒蛋白质和淀粉的积累[J]. 中国农业科学, 2004, 37(6): 808—813.
Tian Yongchao, Zhu Yan, Cao Weixing, et al. Monitoring

- protein and starch accumulation in wheat grains with leaf SPAD and canopy spectral reflectance[J]. *Scientia Agricultura Sinica*, 2004, 37(6): 808—813. (in Chinese with English abstract)
- [7] Wang Zhijie, Wang Jihua, Liu Liangyun, et al. Prediction of grain protein content in winter wheat (*Triticum aestivum*, L.) using plant pigment ratio (PPR)[J]. *Field Crops Research*, 2004, 90(2): 311—321.
- [8] Li Hong, Wei Xing, Yang, et al. Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra[J]. *Pedosphere*, 2007, 17(5): 646—653.
- [9] 王纪华, 黄文江, 赵春江, 等. 利用光谱反射率估算叶片生化组分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)
- [10] 李映雪, 朱艳, 田永超, 等. 小麦冠层反射光谱与籽粒蛋白质含量及相关品质指标的定量关系[J]. *中国农业科学*, 2005, 38(7): 1332—1338.
Li Yingxue, Zhu Yan, Tian Yongchao, et al. Relationship of grain protein content and relevant quality traits to canopy reflectance spectra in wheat[J]. *Scientia Agricultura Sinica*, 2005, 38(7): 1332—1338. (in Chinese with English abstract)
- [11] Huang Wenjiang, Wang Jihua, Liu Liangyun, et al. Study on grain quality forecasting method and indicators by using hyperspectral data in wheat[J]. *Proceedings of SPIE-The International Society for Optical Engineering*, 2005, 5655: 291—300.
- [12] Lemaire G, Jeuffroy M H, Gastal F. Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management[J]. *European Journal of Agronomy*, 2008, 28(4): 614—624.
- [13] Teruo Matsunaka, Yuji Watanabe, Tadashi Miyawaki, et al. Prediction of grain protein content in winter wheat through leaf color measurements using a chlorophyll meter[J]. *Soil Science and Plant Nutrition*, 1997, 43(1): 127—134.
- [14] Rouse J W. Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation[R]. Nasa, 1974.
- [15] Huete A, Justice C, Liu H. Development of vegetation and soil indices for MODIS-EOS[J]. *Remote Sens. Environ*, 1994, 48(2): 224—234.
- [16] Huete A R. A soil-adjusted vegetation index (SAVI)[J]. *Remote Sensing of Environment*, 1988, 25(3): 295—309.
- [17] Rondeaux G, Steven M, Baret F. Optimization of soil-adjusted vegetation indices[J]. *Remote Sensing of Environment*, 1996, 55(2): 95—107.
- [18] Qi J, Chehbouni A, Huete A R, et al. A modified soil adjusted vegetation index[J]. *Remote Sensing of Environment*, 1994, 48(2): 119—126.
- [19] Schleicher T D, Bausch W C, Delgado J A, et al. Evaluation and refinement of the nitrogen reflectance index (NRI) for site-specific fertilizer management[C]// 2001 Sacramento, CA July 29-August 1, 2001. 1998.
- [20] Baret F, Guyot G. Potentials and limits of vegetation indices for LAI and APAR assessment[J]. *Remote Sensing of Environment*, 1991, 35(2): 161—173.
- [21] Gitelson A A, Merzlyak M N. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll[J]. *Journal of Plant Physiology*, 1996, 148(3/4): 494—500.
- [22] Penuelas J, Baret F, Filella I. Semiempirical indexes to assess carotenoids chlorophyll-a ratio from leaf spectral reflectance[J]. *Photosynthetica*, 1995, 31(2): 221—230.
- [23] Tian Y C, Yao X, Yang J, et al. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance[J]. *Field Crops Research*, 2011, 121(3): 464—464.
- [24] Jordan C F. derivation of leaf - area index from quality of light on the forest floor[J]. *Ecology*, 1969, 50(4): 663—666.
- [25] Pearson R L, Miller L D. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie[C]// *Remote Sensing of Environment*, VIII. *Remote Sensing of Environment*, VIII, 1972: 7—12.
- [26] Mishra S, Mishra D R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll- a, concentration in turbid productive waters[J]. *Remote Sensing of Environment*, 2012, 117(2): 394—406.
- [27] Metternicht G. Vegetation indices derived from high-resolution airborne videography for precision crop management[J]. *International Journal of Remote Sensing*, 2003, 24(14): 2855—2877.
- [28] 张学工. 关于统计学习理论与支持向量机[J]. *自动化学报*, 2000, 26(1): 36—46.
Zhang Xuegong. Introduction to statistical learning theory and support vector machines[J]. *Acta Autom Atica Sinica*, 2000, 26(1): 36—46. (in Chinese with English abstract)
- [29] 李国政, 王猛, 曾华军. 支持向量机导论[M]. 北京: 电子工业出版社, 2004
- [30] 梁栋, 管青松, 黄文江, 等. 基于支持向量机回归的冬小麦叶面积指数遥感反演[J]. *农业工程学报*, 2013, 29(7): 117—123.
Liang Dong, Guan Qingsong, Huang Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(7): 117—123. (in Chinese with English abstract)
- [31] 戴栋, 黄筱婷, 代洲, 等. 基于支持向量机的输电线路覆冰回归模型[J]. *高电压技术*, 2013, 39(11): 2822—2828.
Dai dong, Huang Xiaoting, Dai Zhou, et al. Regression model for transmission lines icing based on support vector machine[J]. *High Voltage Engineering*, 2013, 39(11): 2822—2828. (in Chinese with English abstract)
- [32] 傅贵, 韩国强, 遼峰, 等. 基于支持向量机回归的短时交通流预测模型[J]. *华南理工大学学报(自然科学版)*, 2013, 41(9): 71—76.
Fu Gui, Han Guoqiang, Lu Feng, et al. Short-term traffic flow forecasting model based on support vector machine regression[J]. *Journal of South China University of Technology (Natural Science Edition)*, 2013, 41(9): 71—76. (in Chinese with English abstract)
- [33] Chang C C, Lin C J. LIBSVM: A library for support vector machines[M]. ACM, 2011.
- [34] 姜志伟, 武雪萍, 华路, 等. 洛阳旱地夏玉米生产潜力长期定量模拟与评价[J]. *生态学报*, 2009, 29(1): 315—324.

- Jiang Zhiwei, Wu Xueping, Hua Luo, et al. Long-period quantitative simulation and evaluation of summer maize potential productivity in Luoyang dryland[J]. *Acta ecologica Sinica*, 2009, 29(1): 315–324. (in Chinese with English abstract)
- [35] Timsina J, Humphreys E. Performance of CERES-Rice and CERES-Wheat models in rice–wheat systems: A review[J]. *Agricultural Systems*, 2006, 90(1/3): 5–31.
- [36] Biernath C, Gayler S, Bittner S, et al. Evaluating the ability of four crop models to predict different environmental impacts on spring wheat grown in open-top chambers[J]. *European Journal of Agronomy*, 2011, 35(2): 71–82.
- [37] Liu H L, Yang J Y, Drury C F, et al. Using the DSSAT-CERES-Maize model to simulate crop yield and nitrogen cycling in fields under long-term continuous maize production[J]. *Nutrient Cycling in Agroecosystems*, 2011, 89(3): 313–328.
- [38] Rinaldi M, Ventrella D, Gagliano C, et al. Comparison of nitrogen and irrigation strategies in tomato using CROPGRO model: A case study from Southern Italy[J]. *Agricultural Water Management*, 2007, 87(1): 91–105.
- [39] Courant R, Hilbert D, Hoyt F C. Methods of mathematical physics[J]. *Physics Today*, 1962, 15(11): 62–63.
- [40] 薛丽红, 朱艳, 张宪, 等. 利用冠层反射光谱预测冬小麦籽粒品质指标的研究[J]. *作物学报*, 2004, 30(10): 1036–1041.
- Xue Lihong, Zhu Yan, Zhang Xian, et al. Predicting wheat grain quality with canopy reflectance spectra[J]. *Acta Agronomica Sinica*, 2004, 30(10): 1036–1041. (in Chinese with English abstract)
- [41] 于振文. 作物栽培学总论[M]. 北京: 中国农业出版社, 2005.

Remote sensing prediction of winter wheat grain protein content based on nitrogen nutrition index at anthesis stage

Qu Sha^{1,2,3,4}, Li Zhenhai^{2,3,4}, Qiu Chunxia¹, Yang Guijun^{2,3,4}, Song Xiaoyu^{2,3,4*},
Chen Zhaoxia^{2,3,4}, Liu Chang^{1,2,3,4}

(1. College of Geomatics, Xi'an University of Science and Technology, Xi'an, 710054, China; 2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; 3. Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100097, China; 4. Beijing Engineering Research Center of Agricultural Internet of Things, Beijing 100097, China)

Abstract: Grain protein content (GPC) is an important quality index for wheat to meet a variety of needs of the commodity. Advanced site-specific knowledge of GPC would provide opportunities to the classification of wheat harvest and graded storage. Areas with higher GPC can be distinguished from the rest to maximize the price premium. Advanced knowledge of grain protein of the wheat may also provide opportunities to manipulate inputs to optimize outputs. In order to select an appropriate nitrogen nutrition index as an intermediate variable to improve the inversion accuracy of wheat GPC, in this study, the GPC predicting models at anthesis with vegetation index (VI)–nitrogen nutrition index (NNI)–GPC pattern were constructed and evaluated. The NNI included 4 nitrogen nutrition index, i.e. plant nitrogen accumulation (PNA), plant nitrogen content (PNC), leaf nitrogen accumulation (LNA) and leaf nitrogen content (LNC). In previous studies, only a single nitrogen nutrition index was used as the intermediate variable to construct GPC model, and it did not indicate which of the parameters could be utilized as the intermediate variable to obtain the best result. To improve the prediction model accuracy of GPC, we chose the optimal intermediate variable to retrieve the GPC of winter wheat in this study. Field experiments of 6 winter wheat cultivars in Beijing during the growing seasons of 2008–2011 and 2012–2015 were carried out for model building. Firstly, suitable vegetation indices were selected through analyzing the correlation between vegetation indices and nitrogen nutrition index to construct the model of VI–NNI by the support vector machines (SVM) algorithm and the optimal one was selected from the 4 nitrogen nutrition index. Secondly, the measured nitrogen nutrition index and winter wheat GPC were used to construct the NNI–GPC model, and another optimal one was obtained from the 4 nitrogen nutrition index. Thirdly, the 4 nitrogen nutrition index would be used as the intermediate variables to construct and evaluate VI–NNI–GPC prediction model of winter wheat, and we could get the optimal comprehensive model to retrieve GPC of winter wheat. We used modeling determination coefficient (R^2), normalized root mean squared error (nRMSE) and normalized average error (NAE) to evaluate the accuracy of models. The results showed that: 1) The selected 5 vegetation indices, MSAVI (modified soil-adjusted vegetation index), PSRI (plant senescence reflectance index), DVI (difference vegetation index), RDVI (re-normalized difference vegetation index) and GNDVI (green normalized difference vegetation index), which were used to retrieve the 4 nutrition index, produced higher correlation than the other vegetation indices. 2) In the model of VI–NNI, the LNC estimated by SVM was relatively better with the R^2 of 0.820 in the modeling, and the nRMSE of 9.553% and the NAE of -1.4% in the validation. 3) The modeling precision and validation precision of PNC in constructed NNI–GPC model were relatively high, with the R^2 , nRMSE and NAE values of 0.653, 9.843% and -0.3%, respectively. 4) The 'VI–NNI–GPC' model with the PNC as the intermediate variable performed better than the other intermediate variables, with the R^2 of 0.631 in the modeling, and the nRMSE and NAE values of 8.564% and -0.9% in the validation, respectively. It demonstrates that it is feasible to use nitrogen nutrition index as intermediate variables to retrieve GPC by remote sensing, and using PNC as the intermediate parameter achieves more accurate prediction results. This study provides a reliable reference for the accurate prediction of GPC and has a broad range of potential applications.

Keywords: remote sensing; models; support vector machine; grain protein content; nitrogen nutrition index; winter wheat