PLOS ONE

RESEARCH ARTICLE

Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm

Changwei Tan[1☯]*, Xinxing Zhou[1☯], Pengpeng Zhang[1], Zhixiang Wang[1], Dunliang Wang[1], Wenshan Guo[1]*, Fei Yun[2¤]*

1 Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops/Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education of China, Yangzhou University, Yangzhou, China, **2** National Tobacco Cultivation and Physiology and Biochemistry Research Centre/Key Laboratory for Tobacco Cultivation of Tobacco Industry, Henan Agricultural University, Zhengzhou, China

☯ These authors contributed equally to this work.
¤ Current address: College of Agricultural, Yangzhou University, Yangzhou, Jiangsu, China
* tanwei010@126.com (C.T.); yunfeifei55@henau.edu.cn (F.Y.); guows@yzu.edu.cn(W.G.)

Citation: Tan C, Zhou X, Zhang P, Wang Z, Wang D, Guo W, et al. (2020) Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm. PLoS ONE 15(3): e0228500. https://doi.org/10.1371/journal.pone.0228500

Editor: Claudionor Ribeiro da Silva, Universidade Federal de Uberlandia, BRAZIL

Received: July 18, 2019

Accepted: January 16, 2020

Published: March 11, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research was financially supported by the National Key Research and Development Program of China (2018YFD0300805), the National Natural Science Foundation of China (41271415, 31771711), the Project Funded by China Postdoctoral Science Foundation (2019M650125) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Abstract

Remote sensing has been used as an important means of modern crop production monitoring, especially for wheat quality prediction in the middle and late growth period. In order to further improve the accuracy of estimating grain protein content (GPC) through remote sensing, this study analyzed the quantitative relationship between 14 remote sensing variables obtained from images of environment and disaster monitoring and forecasting small satellite constellation system equipped with wide-band CCD sensors (abbreviated as HJ-CCD) and field-grown winter wheat GPC. The 14 remote sensing variables were normalized difference vegetation index (NDVI), soil-adjusted vegetation index (SAVI), optimized soil-adjusted vegetation index (OSAVI), nitrogen reflectance index (NRI), green normalized difference vegetation index (GNDVI), structure intensive pigment index (SIPI), plant senescence reflectance index (PSRI), enhanced vegetation index (EVI), difference vegetation index (DVI), ratio vegetation index (RVI), R_{blue} (reflectance at blue band), R_{green} (reflectance at green band), R_{red} (reflectance at red band) and R_{nir} (reflectance at near infrared band). The partial least square (PLS) algorithm was used to construct and validate the multivariate remote sensing model of predicting wheat GPC. The research showed a close relationship between wheat GPC and 12 remote sensing variables other than R_{blue} and R_{green} of the spectral reflectance bands. Among them, except PSRI and R_{blue}, R_{green} and R_{red}, other remote sensing vegetation indexes had significant multiple correlations. The optimal principal components of PLS model used to predict wheat GPC were: NDVI, SIPI, PSRI and EVI. All these were sensitive variables to predict wheat GPC. Through modeling set and verification set evaluation, GPC prediction models' coefficients of determination (R^2) were 0.84 and 0.8, respectively. The root mean square errors (RMSE) were 0.43% and 0.54%, respectively. It indicated that the PLS algorithm model predicted wheat GPC better than models for linear regression (LR) and principal components analysis (PCA) algorithms. The PLS algorithm model's prediction accuracies were above 90%. The improvement was by more than

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

20% than the model for LR algorithm and more than 15% higher than the model for PCA algorithm. The results could provide an effective way to improve the accuracy of remotely predicting winter wheat GPC through satellite images, and was conducive to large-area application and promotion.

Introduction

The grain quality index of winter wheat includes many parameters, of which grain protein content (GPC) has been the most important indicator for measuring wheat quality [1–3]. At present, GPC was mainly detected by chemical determination method based on manual measurement, which was costly and inefficient. In addition, the current sampling method for investigating the quality of winter wheat was point-like sampling. It meant that only a few sampling points were used to reflect the situation in a large area. Therefore, the samples lacked representativeness and made it difficult to grasp the overall quality information of winter wheat over requisite time. In the protein content monitoring method, a combined method was proposed for pretreatment of the NIR spectrum. This was based on both the empirical mode decomposition and the wavelet soft-threshold methods, presuming certain accuracy in the monitoring of GPC content [4]. A rapid and simplified decision support method to predict the wheat quality at a small range was established with an accuracy of more than 80% [5]. Compared with previous researches, remote sensing technology has the advantages of being fast, accurate and based on wide range in data collection. Therefore, the model based on remote sensing and corresponding algorithm could serve as an effective way to obtain wheat quality status in advance. With the urgent need for remote sensing in the agricultural field, more and more studies focused on crop quality prediction. To materialize this, the remote sensing has a wide range of application and development in large-scale regional crop management and monitoring [6, 7].

For many years, agricultural remote sensing focused mainly on crop growth monitoring and yield estimation, and formed a relatively complete technical system. Recently, by using different combinations of remote sensing variables, the reliability of remote sensing model of nitrogen concentration in wheat leaves has been improved [8]. Likewise, the remote sensing prediction of crop yield loss under soil salinization effect has also achieved some results [9]. Production forecasts based on advance very high-resolution radiometer (AVHRR) data in Kansas, USA were almost identical to production data from local government field surveys [10]. At present using multi-temporal radarsat-2 SAR image, wheat could be identified effectively with an accuracy rate of 0.929 [11]. The normalized difference vegetation index (NDVI), which extracted from the moderate resolution imaging spectroradiometer (MODIS) data, has been used in a wide range of applications for global agricultural monitoring, particularly in crop growth monitoring, quality prediction and yield estimation [12]. In another attempt from 2003 to 2015, using NDVI deduced MODIS data some researchers improved the estimation and prediction method of wheat yield in Hungary with good results [13]. However, there were few reports on crop quality remote sensing prediction using spectral reflectance. Later on, through the ground spectral data, some studies have assessed the metabolic energy, ash content, crude protein and other indicators of leguminous plants [14]. There were also corresponding breakthroughs in forestry, as well as the water index of olive forest was successfully detected by vegetation spectroscopy [15]. In recent years, many researchers studied the prediction of crop quality based on the space satellite remote sensing platform [16, 17]. With

advancement of geospatial technology in agriculture and the significant improvement of the resolution of remote sensing images, a large number of studies on the spatial pattern of farmland yield and quality have been reported successively [18]. Early repeated remotely sensed multispectral data reliably predicted the yield and quality of winter wheat and spring barley [19]. In the monitoring of quality fluctuation, some researchers combined remote sensing with geographic information system (GIS) to explain the changes of soybean oil and protein content [20]. Multi-temporal image monitoring might be the future trend. Recent study has also shown that three satellite images from each of landsat thematic map (TM) and advanced synthetic aperture radar (ASAR) successfully monitored the crop conditions and predicted yield and protein content [21]. According to a number of previous studies, remote sensing technology has been considered as a potential and effective method to predict the protein content and quality of wheat grains [22]. The results showed that the prediction of wheat GPC with TM and enhanced thematic mapper (ETM) data was effective [23]. Studies also suggest feasibility of using KODA-CIR (Eastman Kodak Co., USA) and Cropscan (NextInstruments Co., Australia) data to predict GPC of winter wheat one month before the harvest [24]. GPC prediction by using high-resolution satellite images to monitor the potential growth and development of wheat was also available [25]. Besides, the fusion of multi-sensor and multi-temporal remote sensing images as the data source provided a technical approach for predicting wheat GPC [26]. There had been many reports on remote sensing monitoring of agricultural conditions based on partial least squares method. Most of these primarily focused on crop pests and diseases as well as growth. Some researchers had successfully measured the canopy biomass and nitrogen status of wheat by using NDVI and partial least square (PLS) algorithm. In the growth of rice leaves, there were a number of breakthroughs in hyperspectral reflectance and PLS regression analysis [27–28]. Based on multi-temporal and multi-season satellite remote sensing data, PLS algorithm was used to monitor the host species distribution of spruce budworm in large forests [29]. However, there were very limited reports on quantitatively forecasting chemical components in grains such as GPC using satellite remote sensing data [8, 30]. On September 6, 2008, China has launched successfully satellites A and B (abbreviated as HJ-CCD) of the "Environment and Disaster Monitoring and Forecasting Small Satellite Constellation System" with independent intellectual property rights. The satellites were equipped with wide-band CCD sensors with spatial resolution of the sensor being 30 m. Time resolution was 2 d when satellites A and B were making observations simultaneously. This made them an ideal data source for agricultural remote sensing operation. Some studies on remote sensing prediction of wheat quality were still based on traditional algorithms and its accuracy was consequently affected [17, 20]. In this study, HJ-CCD images were used as remote sensing data sources and combined with PLS algorithm to construct GPC prediction model.

The objectives of the present study were to investigate the quantitative relationship between satellite remote sensing variables during flowering period and wheat GPC, and developed an effective way to improve the accuracy of predicting wheat GPC through remote sensing.

Materials and methods

Test design and data acquisition

Field sampling was used in this study for three years. The survey area was representative and the varieties are different. Samples were taken back to the laboratory for analysis, and corresponding satellite image data were collected.

For the present investigation, data collection was carried out in 5 counties namely, Taixing, Jiangyan, Yizheng, Xinghua and Dafeng in Jiangsu Province, the Peoples Republic of China. There were 15–20 sampling points in each county, totaling 92. The location of each sampling

site was determined by using a Juno ST hand-held GPS meter (Trimble Co., USA). The survey mainly included information collection on winter wheat varieties, growth period, population growth and disaster status (mainly by pests and diseases). Winter wheat varieties were of medium and weak gluten type, mainly *Yangmai 13*, *Yangmai 15 and Yangmai 16*. These varieties were available in the experimental counties. At harvest time, wheat grains were sampled by five-point sampling method in the field, and then brought back to the laboratory for wheat GPC determination [31].

A total of 3 tests were launched in the experimental counties from 2016–2018 to collect data. The satellite data was HJ-CCD images taken at flowering stage of the wheat crop. Data collection for Test 1, 2 and 3 were conducted on May 2, 2016; April 24, 2017 and April 26, 2018, respectively. The sampling points considered for the Test 1–3 were 92, 96 and 67, respectively.

Image preprocessing

Environment for Visualizing Images (ENVI 5.4) software (ESRI Co., USA) was used to preprocess satellite images. Firstly, the geometric rough correction of the satellite image was carried out by using the 1:100,000 topographic maps of Jiangsu area. Thereafter, the GPS control points for ground measuring were used to precisely correct the satellite image. This helped to ensure that the precision of geometric correction was better than one pixel. Atmospheric correction and reflectance conversion were carried out by empirical linear method [32]. According to the analysis of the results, the corresponding single-band value graph was obtained by band math. Data of wheat growing areas were obtained by supervised classification. The winter wheat planting data were superimposed and the non-winter wheat area was eliminated by one-to-one solution and binarization mask. By using the administrative boundary vector data and the PLS model, the spatial distribution map of winter wheat GPC prediction in Jiangsu province was produced.

Satellite remote sensing variables

In combination with the physical significance of spectral indices, selection of model parameters was based on the spectral characteristics of crops and the available literatures at home and abroad. Finally, in this study, four HJ-CCD bands and ten common spectral vegetation indices were selected (Table 1) as independent variables for PLS analysis in order to construct the model of predicting winter wheat GPC.

To extract spectral reflectance values of corresponding GPS positioning sampling points, ENVI 5.4 and geographic information system software (ArcGIS 10.2) (ESRI Co., USA) were used. These combined with the remote sensing vegetation index algorithm as provided in Table 1, satellite remote sensing variables were calculated using Excel 2016.

PLS regression

PLS regression was first applied to the field of chemometrics. Since then, it has been considered as a new multivariate analysis method with wide applicability. It was concentrated on the characteristics of principal component, linear regression and typical multiple regression analyses. It could effectively solve many problems. Such as, problems that cannot be solved by ordinary multiple regression, especially when there were many variables and multiple correlations. In these cases, PLS could effectively decompose and screen the comprehensive variables that were most explanatory to the dependent variables. Therefore, the established model is more reliable than the ordinary regression analysis. The PLS method first extracted a new variable called component as an independent variable, and established a linear combination

Table 1. Remote sensing vegetation indices used in this study.

Vegetation index	Abbreviation	Algorithm	Source
Normalized difference vegetation index	NDVI	$(R_{nir}-R_{red})/(R_{nir}+R_{red})$	[33]
Soil-adjusted vegetation index	SAVI	$(R_{nir}-R_{red})/(R_{nir}+R_{red}+0.5)^{*}1.5$	[34]
Optimized soil-adjusted vegetation index	OSAVI	$(R_{nir}-R_{red})/(R_{nir}+R_{red}+0.16)^{*}1.16$	[35]
Nitrogen reflectance index	NRI	$(R_{green}-R_{red})/(R_{green}+R_{red})$	[36]
Green normalized difference vegetation index	GNDVI	$(R_{nir}-R_{green})/(R_{nir}+R_{green})$	[37]
Structure intensive pigment index	SIPI	$(R_{nir}-R_{blue})/(R_{nir}+R_{blue})$	[38]
Plant senescence reflectance index	PSRI	$(R_{red}-R_{blue})/B_{nir}$	[39]
Enhanced vegetation index	EVI	$2.5^{*}(R_{nir}-R_{red})/(R_{nir}+6^{*}R_{red}-7.5^{*}R_{green}+1)$	[40]
Difference vegetation index	DVI	$R_{nir}-R_{red}$	[41]
Ratio vegetation index	RVI	R_{nir}/R_{red}	[42]

R_{blue}, R_{green}, R_{red} and R_{nir} denoted spectral reflectance at blue, green, red and near infrared bands, respectively.

https://doi.org/10.1371/journal.pone.0228500.t001

relationship between the dependent variable and the independent variable. The coefficient was determined by PLS calculation, and then the regression equation of the dependent variable was constructed. The regression model established by the PLS method could be expressed by Eq 1:

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

Where $x_1, \cdots x_p$ were linear combinations of remote sensing variables, $a_{0m}, a_{1m}, \cdots a_{pm}$ were parameters of the regression model and could be computed by PLS.

When the model was established by PLS algorithm, the increase of the number of principal components would improve the accuracy of the model. But too many principal components would cause over-fitting and the error would increase. Therefore, it was very important to determine the optimal principal components number of the PLS model. In this study, the sum of squared residuals was calculated by the cross-validation method. The smaller the predictive residual errors sum of square (PRESS) value, the stronger the prediction ability of the model is. Therefore, the optimal principal components number could be determined according to the minimum value of PRESS. PRESS can be expressed by Eq 2:

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

Where $y_i, y_{i,-i}$ were the measured value corresponding to the ith sample and the estimated value when the ith sample was excluded, and k was the number of validating iterations.

For the basic principles and specific practices of the PLS algorithm and PRESS, please refer to reference [43], which is not described here. Both the PLS and PRESS processes were performed by a self-written MATLAB program.

Evaluation of the model

Using the samples of the modeling set, and the verification set, the model was evaluated by plotting the 1:1 relationship graph between the predicted and measured values of winter wheat GPC. The evaluation indices were the determination coefficient (R^2) and the root mean square error (RMSE) [44]. On one hand, the larger the R^2, the better the model is. On the other hand, the smaller the RMSE, the stronger the estimation ability of the model is. RMSE and estimation

accuracy were calculated using Eqs 3 and 4, respectively:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (4)$$

Where y_i and $\hat{y}_i$ represented measured values and predicted values of winter wheat GPC, respectively, and n was the number of samples.

Results

GPC distribution

The GPC data measured in Tests 1–3 were arranged in the order of the GPC values in the winter wheat grain sample. To enhance the stability of the prediction model, under the premise that the maximum and minimum values of winter wheat GPC were guaranteed, needs to be included in the modeling sample set. To perform this, the numerical samples of 255 GPC were randomly divided into modeling set and verification set according to the ratio of 3:2. It could be seen from Table 2 that the amplitude of variation, mean, standard deviation and standard error of the modeling set and verification set samples were similar. At the same time, the modeling set and the verification set samples had desirable consistency.

Quantitative analysis between remote sensing variables and GPC

Table 3 shows the quantitative analysis of the GPC and remote sensing variables of 153 samples in the modeling set. It indicated that there was significant or extremely significant relationship between the GPC and 12 remote sensing variables except R_{blue} and R_{green}. The GPC was most closely related to NDVI, followed by enhanced vegetation index (EVI). The correlation coefficients (r) being 0.82 and 0.75, respectively for NDVI and EVI. The correlation between vegetation index and GPC was obviously better than single-band. All the other remote sensing variables had considerable multiple pairwise correlations. Except PSRI and R_{blue}, R_{green} and R_{red}, other remote sensing variables had pairwise correlation coefficients between 0.80 and 0.99. In particular, single-band B1-B4 pair wise correlation coefficients were between 0.93 and 0.98, and the pairwise correlation coefficient of most vegetation indices were above 0.90.

Determination of the number of optimal principal components

The smaller the PRESS values, the stronger the prediction ability of the model is. It means the number of optimal principal components could be determined based on the PRESS minimum value. Fig 1 shows the variation of PRESS with the number of principal components obtained from the GPC modeling set. At the beginning, as the number of principal components increased, the PRESS value decreased to a large extent. It has indicated that due to the small number of principal components, the model fitting was extremely inadequate. It means the missing fitting phenomenon occurred. When the principal components number of the GPC model was 4, the PRESS value was the smallest (PRESS = 21.39). After that, as the number of principal components increased, the PRESS value increased sharply, until they tend to be saturated. Via this, it was indicated that the over-fitting phenomenon occurred due to too many principal components. Therefore, it was reasonable to select the number of principal components corresponding to the minimum PRESS value. Since the optimal principal components

Table 2. Distribution of winter wheat GPC in the modeling and verification set (GPC unit: %).

Sample set	Number of samples	Amplitude of variation	Mean	Standard deviation	Standard error
Modeling set	153	9.36–14.58	11.99	1.33	0.11
Verification set	102	9.38–14.39	12.29	1.42	0.14

GPC refered to the grain protein content in dry matter.

https://doi.org/10.1371/journal.pone.0228500.t002

number of the PLS model, the optimal principal components number of the GPC model based on PLS algorithm were 4.

PLS model

The structure of the PLS model was based on the PLS algorithm and the four vegetation indices with the principal components number of 4. All these were sensitive to the prediction of wheat GPC and could be easily extracted and calculated from the HJ-CCD image. These were NDVI, structure intensive pigment index (SIPI), plant senescence reflectance index (PSRI) and EVI. All these were considered as the independent variables and the GPC was the dependent variable for the PLS model of predicting the GPC. The GPC model constructed by the modeling set and the HJ-CCD images during the three days 2016-05-02, 2017-04-24, and 2018-04-26 were Eq 5:

$$GPC = 3.873 \times NDVI + 1.696 \times SIPI + 2.862 \times PSRI - 1.276 \times EVI + 5.821 \quad (5)$$

After the PLS model was built, it was used to predict winter wheat GPC. The predicted and measured GPC values were plotted as a 1:1 scatter plot. The optimal linear regression equation and its R^2 and RMSE were obtained. Fig 2 shows the evaluation of the PLS model's prediction ability. It could be seen from Fig 2 that the model set samples number was larger than the

Table 3. Correlation coefficients (r) between remote sensing variables and GPC.

	GPC	R_{blue}	R_{green}	R_{red}	R_{nir}	NDVI	OSAVI	SAVI	SIPI	PSRI	GNDVI	NRI	RVI	DVI	EVI
R_{blue}	-0.22	1.00													
R_{green}	-0.08	0.98	1.00												
R_{red}	-0.46	0.97	0.96	1.00											
R_{nir}	0.51	0.93	0.93	0.96	1.00										
NDVI	0.82	-0.67	-0.78	-0.88	0.93	1.00									
OSAVI	0.65	-0.67	-0.79	-0.85	0.94	0.95	1.00								
SAVI	0.59	-0.65	-0.81	-0.87	0.96	0.94	0.98	1.00							
SIPI	0.71	-0.64	-0.71	-0.69	0.95	0.98	0.97	0.98	1.00						
PSRI	0.63	-0.37	-0.26	-0.18	0.77	0.86	0.93	0.98	0.91	1.00					
GNDVI	0.67	-0.62	-0.79	-0.92	0.64	0.95	0.88	0.91	0.92	0.97	1.00				
NRI	-0.59	-0.68	0.68	0.87	-0.58	-0.87	-0.88	-0.86	-0.86	0.90	0.85	1.00			
RVI	0.61	-0.69	-0.82	-0.84	0.94	0.99	0.99	0.99	0.97	0.83	0.87	-0.84	1.00		
DVI	-0.63	0.66	0.72	0.77	-0.88	-0.97	-0.96	-0.96	-0.96	0.86	0.85	0.85	0.99	1.00	
EVI	0.75	-0.64	-0.78	-0.79	0.97	0.99	0.99	0.99	0.99	0.94	0.87	-0.83	0.98	0.98	1.00

All abbreviations were denoted by: normalized difference vegetation index (NDVI), soil-adjusted vegetation index (SAVI), optimized soil-adjusted vegetation index (OSAVI), nitrogen reflectance index (NRI), green normalized difference vegetation index (GNDVI), structure intensive pigment index (SIPI), plant senescence reflectance index (PSRI), enhanced vegetation index (EVI), difference vegetation index (DVI), ratio vegetation index (RVI), R_{blue} (reflectance at blue band), R_{green} (reflectance at green band), R_{red} (reflectance at red band) and R_{nir} (reflectance at near infrared band)

https://doi.org/10.1371/journal.pone.0228500.t003

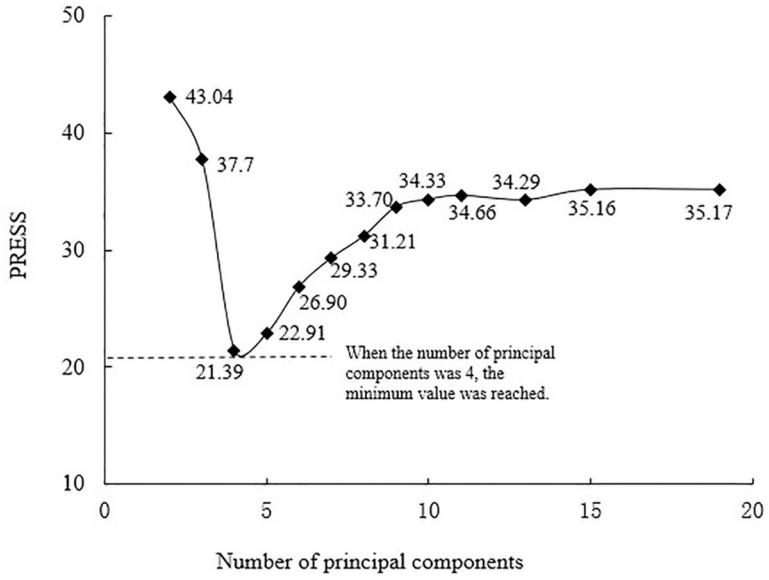

Fig 1. PRESS changes with the principal components.

https://doi.org/10.1371/journal.pone.0228500.g001

verification set samples number. The R^2 of the linear equation thus established by the modeling set was larger than R^2 of the verification set. The set RMSE was significantly smaller than the verification set RMSE. It indicates that the prediction model effect of the modeling set samples was better than the verification set. Thereby, it has theoretically conformed to the model's evaluation law [45]. In addition, the R^2 values between the predicted and measured GPC of the modeling and verification sets were greater than 0.8 and the RMSE were 0.43% and 0.54%, respectively. This result indicated that the PLS model could be used effectively to predict the winter wheat GPC.

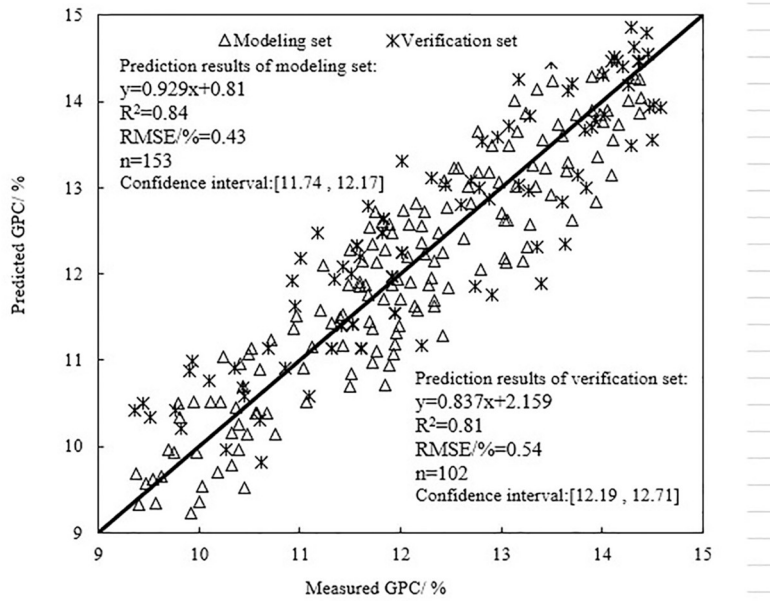

Fig 2. Evaluation of GPC model based on PLS algorithm.

https://doi.org/10.1371/journal.pone.0228500.g002

In order to compare with the traditional algorithm, the linear regression (LR) and principal components analysis (PCA) algorithm were used to establish the GPC estimation models using the model set and verification set samples, respectively. The GPC models were evaluated by R^2 and RMSE. The specific process was not described here. Table 4 shows the comparison of predicted results with PLS, LR and PCA based on the modeling set and verification set. It showed that the sample number was the same. The PLS algorithm models' R^2 were greater than those for LR and PCA algorithm models. But RMSE were smaller than those for LR and PCA algorithm models. This indicated that the PLS algorithm model was better than the LR and PCA algorithms in predicting winter wheat GPC. The modeling set and the verification set prediction accuracy were 20.6% and 22.4% higher than the LR algorithm, respectively, and were 15.4% and 16.3% higher than PCA algorithms, respectively.

According to the above analysis, NDVI, SIPI, PSRI and EVI maps were generated using 2018-04-26 HJ-CCD images. On those the winter wheat planting data was superimposed to remove the non-winter wheat area by one-to-one solution and binarization mask. Based on the administrative boundary vector data, as well as the above PLS model, the spatial distribution map for predicting winter wheat GPC in Jiangsu province was produced (Fig 3). The GPC distribution was mainly 11.3–11.8%. There was often more than 12.5% in Dafeng and Rudong wheat area and the north west of Jinhu wheat area. The GPC of some wheat regions in the north of the Yangtze River was 11.8–12.5%. The number in the southern wheat area was rarely higher than 11.8%. However, the number in the area along the Yangtze River was mainly 11.3–11.8%, particularly in the south.

Discussion

At present, the remote sensing images used in the crop estimation were mainly originated via MODIS, national oceanic and atmospheric administration (NOAA)/AVHRR, etc. [10, 33, 41]. These images had low spatial resolution and were difficult to apply to high-precision winter wheat remote sensing estimation in small areas. On the other hand, the high-resolution images such as Quickbird (Panchromatic image 0.61–0.72 m, multispectral image 2.44–2.88 m), SPOT (Panchromatic image 10 m, multispectral image 20 m), IKONOS (Panchromatic image 1 m, multispectral image 4 m) were costly [46, 47]. The medium-resolution TM images had revisiting periods of 16 days, making it difficult to obtain high-quality data in time. This limited continuous crop monitoring and made it inappropriate to predict crop quality [38]. The HJ-CCD satellites developed by China have been put into use one after another. The quality of the data obtained was continuously improved and was provided free of charge to users. This has created a convenient data platform for remote sensing and estimation of regional crop's quality and productivity [8, 48]. The experimental area of the present research has been located in the coastal area along the Yangtze River in Jiangsu Province, China. The whole wheat field has been fragmented and as a result the planting structure was complex. The time resolution of the selected HJ-CCD image was 2 d, and the scanning width of the single scene image was 750 km. These characteristics could meet the estimation demands for the actual regional winter wheat. Considering time resolution, spatial resolution and cost, the HJ-CCD image was more appropriate than the data of MODIS, TM, Quickbird, etc.

There was a close relationship between wheat GPC and 12 remote sensing variables except R_{blue} and R_{green}. In addition, there were considerable multiple correlations between all the other remote sensing variables except PSRI and R_{blue}, R_{green} and R_{red}. This made it difficult to establish a high precision remote sensing estimation model of wheat GPC using traditional algorithms [49]. In this study, the PLS algorithm was used to construct the remote sensing estimation model with NDVI, SIPI, PSRI and EVI as the independent variables. The correlation

Table 4. Comparison of predicted abilities with PLS, LR and PCA.

Algorithm	Number of principal components	Number of samples		R^2		RMSE/%		Accuracy/%	
		Modeling set	Verification set	Modeling set	Verification set	Modeling set	Verification set	Modeling set	Verification set
PLS	4	153	102	0.84	0.81	0.43	0.54	94.7	91.8
PCA	5	153	102	0.57	0.52	0.92	0.98	79.3	75.5
LR	0	153	102	0.49	0.45	1.05	1.23	74.1	69.4

PLS, LR, PCA, R^2 and RMSE denoted partial least square, linear regression, principal components analysis, determination coefficient and root mean square error, respectively.

https://doi.org/10.1371/journal.pone.0228500.t004

between the four remote sensing variables and GPC was extremely significant. They could be easily extracted and calculated from the HJ-CCD image. The RMSE values of the GPC prediction model based on NDVI, SIPI, PSRI and EVI as the independent variables were lower than the traditional LR and PCA models. The results showed that the PLS model, as a new multivariate analysis method, has a very high adaptability in wheat GPC prediction. In particular, there were many variables and multiple correlations in the analysis. The PLS algorithm could effectively optimize the dependent variables, and its model was significantly better than LR and PCA algorithms in wheat GPC prediction. The prediction accuracies were above 90%, and were improved by more than 20% compared to the LR algorithm and more than 15% higher than the PCA algorithm. The results were consistent with Hansen *et al.* [19] and Zhao *et al.* [23], and better than Liu *et al.* [21] and Xue *et al.* [22]. In order to reflect it in a better way, the

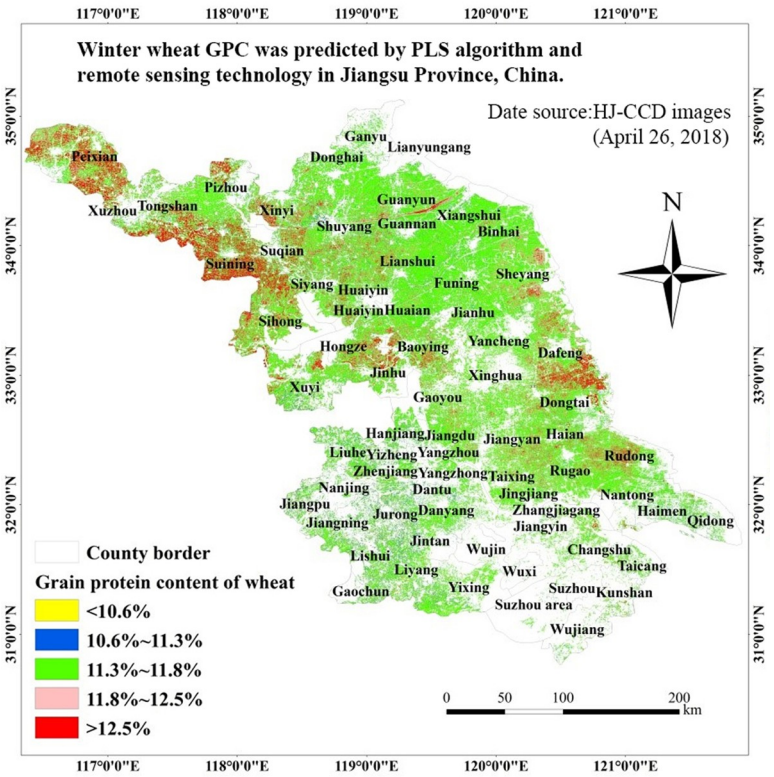

Fig 3. Spatial distribution of winter wheat GPC in Jiangsu province.

https://doi.org/10.1371/journal.pone.0228500.g003

actual situation of field planting and different varieties were selected in the experiment. With the data of different varieties as test samples, the results had more general significance. It was helpful to popularize and apply in practical production.

According to the spatial distribution map for predicting winter wheat GPC in Jiangsu province (Fig 3), the wheat GPC in northern of Jiangsu and eastern of Jiangsu was generally higher. The wheat GPC in the middle area of Jiangsu maintained a medium level of 11.3%-11.8%. The wheat GPC in southern of Jiangsu was relatively low. There was large scale wheat cultivation in northern and eastern of Jiangsu. Local agricultural facilities were well developed, and agricultural production was mainly in the form of farms for planting and management. Therefore, winter wheat planting could be managed uniformly, with good cultivation measures and maximum implementation. Overall agricultural production in the middle area of Jiangsu was slightly worse than that in north of Jiangsu. But the whole structure of agricultural facilities and agricultural production could meet the planting of winter wheat. Therefore, wheat GPC presented a general level range. Southern of Jiangsu was mostly metropolis and urban area with less farmland, and there were few areas for wheat cultivation. At the same time, the local farmland was chaotic and scattered, mainly operated by small farmers households. It might result in good cultivation measures and management could not be used effectively. Therefore, the wheat GPC in southern of Jiangsu was relatively low. The predicted results of the spatial distribution map for predicting winter wheat GPC in Jiangsu province were basically consistent with the actual situation of winter wheat production. It indicated that it was feasible to use the PLS model to predict winter wheat GPC with high precision. It has, therefore, provided an effective method and technical support for the high precision remote sensing prediction of wheat GPC.

The GPC values of the samples used in this study were basically ranging from 10–14%. Samples (GPC) with higher or lower content were relatively few, showing above 14% and less than 10%. There was a lack of samples more than 14.58% and less than 9.36%. If the variation of the GPC samples was increased, the PLS model may be further optimized and its application range would be further expanded. The remote sensing prediction model of winter wheat GPC would become more reliable. The results obtained were based only on the HJ-CCD data of the Jiangsu experimental area. Therefore, whether the model would be applicable to other remote sensing sensor data and/or be able to predict the winter wheat GPC in other areas needs further study.

The present study did not compare the PLS algorithm with artificial neural network (ANN) [50], support vector machines (SVM) [51], geostatistics [52], etc. Simultaneously, it also did not take into account the factors affecting winter wheat cultivation such as weather, soil and cultivation practices and so on. These algorithms and factors might have a wide range of influence on the results of predicting winter wheat GPC and needed further study.

Conclusion

NDVI, SIPI, PSRI and EVI were sensitive for predicting GPC based on PLS algorithm and HJ-CCD images. Through the modeling set and the verification set evaluation, the GPC model's R^2 were 0.84 and 0.81 and the RMSE were 0.43% and 0.54%, respectively. The prediction accuracies were above 90%. The improvements were by more than 20% than the LR algorithm and more than 15% higher than the PCA algorithm.

Supporting information

S1 File. Dataset.
(XLSX)

Acknowledgments

We thank National Science & Technology Infrastructure of China for providing the winter wheat distribution map and HJ-CCD data for 2016–2018. Thanks are also due to the anonymous reviewers and editor for their valuable comments in improving the quality of the manuscript.

Author Contributions

Conceptualization: Changwei Tan, Wenshan Guo.

Formal analysis: Xinxing Zhou, Pengpeng Zhang.

Funding acquisition: Changwei Tan, Wenshan Guo.

Investigation: Xinxing Zhou.

Methodology: Changwei Tan, Pengpeng Zhang.

Project administration: Changwei Tan, Wenshan Guo.

Software: Zhixiang Wang.

Supervision: Changwei Tan.

Validation: Dunliang Wang, Fei Yun.

Writing – original draft: Changwei Tan.

Writing – review & editing: Xinxing Zhou, Pengpeng Zhang, Fei Yun.

References

1. Soo C, Hee B, Baik BK. Protein quality of wheat desirable for making fresh white salted noodles and its influences on processing and texture of noodles. Cereal. Chem. 2003; 80(3):297–303. https://doi.org/10.1094/CCHEM2003.80.3.297
2. Voon JP, Edwards GW. Research payoff from quality improvement: The case of protein in australian wheat. Am. J.Agr. Econ. 1992; 74(3):564–572. https://doi.org/10.2307/1242569
3. Katyal M, Virdi AS, Kaur A, Singh N. Diversity in quality traits amongst Indian wheat varieties I: Flour and protein characteristics. Food Chem. 2016; 194:337–344. https://doi.org/10.1016/j.foodchem.2015.07.125 PMID: 26471563
4. Cai JH. Near-infrared spectrum detection of wheat gluten protein content based on a combined filtering method. J AOAC Int. 2017; 100(5):1565–1568. https://doi.org/10.5740/jaoacint.17-0008 PMID: 28425394
5. Bonfil DJ, Karnieli A, Raz M, Mufradi I, Asido S, Egozi H, et al. Decision support system or improving wheat grain quality in the Mediterranean area of Israel. Field Crop Res. 2004; 89(1):153–163. https://doi.org/10.1016/j.fcr.2004.01.017
6. Wang Z, Wang J, Zhao C, Zhao M. Vertical distribution of nitrogen in different layers of leaf and stem and their relationship with grain quality of winter wheat. J. Plant Nutr. 2005; 28(1):73–91. https://doi.org/10.1081/PLN-200042175
7. Gnyp ML, Bareth G, Li F, Lenz-Wiedemann V. Development and implementation of a multiscale biomass model using hyperspectral vegetation indices for winter wheat in the North China Plain. Int. J. Appl. Earth Obs.Geoinf. 2014; 33:232–242. https://doi.org/10.1016/j.jag.2014.05.006
8. Tan C, Wang D, Zhou J, Du Y, Luo M, Guo W. Estimation of leaf nitrogen concentration in wheat by the combinations of two vegetation indexes using HJ-CCD images. Int. J. Agric. Biol. 2018; 20(8):1908–1914. https://doi.org/10.17957/IJAB/15.0754
9. Satir O, Berberoglu S. Crop yield prediction under soil salinity using satellite derived vegetation indices. Field Crop Res. 2016; 192:134–143. https://doi.org/10.1016/j.fcr.2016.04.028
10. Salazar L, Kogan F, Roytman L. Use of remote sensing data for estimation of winter wheat yield in the United States. Int. J. Remote Sens. 2007; 28(17):3795–3811. https://doi.org/10.1080/01431160601050395

11. Xu J, Li Z, Tian B, Huang L. Polarimetric analysis of multi-temporal RADARSAT-2 SAR images for wheat monitoring and mapping. Int. J. Remote Sens. 2014; 35(10):3840–3858. https://doi.org/10.1080/01431161.2014.919679

12. Becker-Reshef I, Chris J, Mark S, Vermote E. Monitoring global croplands with coarse resolution earth observations: The global agriculture monitoring (GLAM) project. Remote Sens. 2010; 2(6):1589–1609. https://doi.org/10.3390/rs2061589

13. Bognár P, Kern A, Pásztor S, Lichtenberger J, Koronczay D, Ferencz C. Yield estimation and forecasting for winter wheat in Hungary using time series of MODIS data. Int. J. Remote Sens. 2017; 38(11): 3394–3414. https://doi.org/10.1080/01431161.2017.1295482

14. Biewer S, Fricke T, Wachendorf M. Development of canopy reflectance models to predict forage quality of legume–grass mixtures.Crop Sci. 2009; 49(5):1917–1926. https://doi.org/10.2135/cropsci2008.11.0653

15. Rallo G, Mario M, Ciraolo G, Provenzano G. Detecting crop water status in mature olive groves using vegetation spectral measurements. Biosyst. Eng. 2014; 128(SI):52–68. https://doi.org/10.1016/j.biosystemseng.2014.08.012

16. Maestrini B, Basso B. Predicting spatial patterns of within-field crop yield variability. Field Crop Res. 2018; 219:106–112. https://doi.org/10.1016/j.fcr.2018.01.028

17. Orlando F, Marta AD, Mancini M, Motha R. Integration of remote sensing and crop modeling for the early assessment of durum wheat harvest at the field scale. Crop Sci. 2015; 55(3):1280–1289. https://doi.org/10.2135/cropsci2014.07.0479

18. Tan C, Du Y, Zhou J, Wang D, Luo M, Zhang Y. et al. Analysis of different hyperspectral variables for diagnosing leaf nitrogen accumulation in wheat. Front. Plant Sci. 2018; 9:1–11. https://doi.org/10.3389/fpls.2018.00001

19. Hansen PM, Jørgensen JR, Thomsen A. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J. Agr. Sci. 2002; 139(3):307–318. https://doi.org/10.1017/S0021859602002320

20. Nutter FW, Tylka GL, Guan J, Moreira AJD. Use of remote sensing to detect soybean cyst nematode-induced plant stress. J. Nematol. 2002; 34(3):59–61. https://doi.org/10.1016/S0022-2011(02)00111-8

21. Liu L, Wang J, Bao Y, Huang W, Ma Z, Zhao C. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images. Int. J. Remote Sens. 2006; 27(4):737–753. https://doi.org/10.1080/01431160500296867

22. Xue L, Cao W, Yang L. Predicting grain yield and protein content in winter wheat at different N supply levels using canopy reflectance spectra. Pedosphere. 2007; 17(5):646–653. https://doi.org/10.1016/s1002-0160(07)60077-0

23. Zhao C, Liu L, Wang J, Huang W, Song X, Li C. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. Int. J. Appl. Earth Obs. 2005; 7(1):1–9. https://doi.org/10.1016/j.jag.2004.10.002

24. Reyniers M, Vrindts E, Baerdemaeker JD. Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat. Eur. J.Agron. 2006; 24(2):87–94. https://doi.org/10.1016/j.eja.2005.05.002

25. Marta AD, Grifoni D, Mancini M, Orlando F, Guasconi F, Orlandini S. Durum wheat in-field monitoring and early-yield prediction: assessment of potential use of high resolution satellite imagery in a hilly area of Tuscany, Central Italy. J.Agr. Sci. 2015; 153(1):68–77. https://doi.org/10.1017/S0021859613000877

26. Wang L, Tian Y, Yao X, Zhu Y, Cao W. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crop Res. 2014; 164:178–188. https://doi.org/10.1016/j.fcr.2014.05.001

27. Hansen PM, Schjoerring JK. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens. Environ. 2003; 86(4):542–553. https://doi.org/10.1016/s0034-4257(03)00131-7

28. Nguyen HT, Lee BW. Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. Eur. J.Agron. 2006; 24(4):349–356. https://doi.org/10.1016/j.eja.2006.01.001

29. Wolter PT, Townsend PA, Sturtevant BR, Kingdon CC. Remote sensing of the distribution and abundance of host species for spruce budworm in Northern Minnesota and Ontario. Remote Sens. Environ. 2008; 112(10):3971–3982. https://doi.org/10.1016/j.rse.2008.07.005

30. Hagstrum DW. Using five sampling methods to measure insect distribution and abundance in bins storing wheat. J. Stored Prod. Res. 2000; 36:253–262. https://doi.org/10.1016/s0022-474x(99)00047-8 PMID: 10758264

31. Lugassi R, Zaady E, Goldshleger N, Shoshany M, Chudnovsky A. Spatial and temporal monitoring of pasture ecological quality: Sentinel-2-based estimation of crude protein and neutral detergent fiber contents. Remote Sens. 2019; 11:1–28 https://doi.org/10.3390/rs11070799

32. Hamm N, Atkinson PM, Milton EJ. A per-pixel, non-stationary mixed model for empirical line atmospheric correction in remote sensing. Remote Sens. Environ. 2012; 124:666–678. https://doi.org/10.1016/j.rse.2012.05.033

33. Sarmah S, Jia G, Zhang A, Singha M. Assessing seasonal trends and variability of vegetation growth from NDVI3g, MODIS, NDVI and EVI over South Asia. Remote Sens. Lett. 2018; 9:1195–1204. https://doi.org/10.1080/2150704X.2018.1519270

34. Ren HR, Zhou GS. Determination of green aboveground biomass in desert steppe using litter-soil-adjusted vegetation index. Eur. J. Remote Sens. 2014; 47:611–625. https://doi.org/10.5721/EuJRS20144734

35. Clevers JGPW, Kooistra L, Brande MVD. Using sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. Remote Sens. 2017; 9(5):405. https://doi.org/10.3390/rs9050405

36. Knox NM, Skidmore AK, Schlerf M, Boer WF, Wieren S, Waal CV. et al. Nitrogen prediction in grasses: effect of bandwidth and plant material state on absorption feature selection. Int. J. Remote Sens. 2010; 31:691–704. https://doi.org/10.1080/01431160902895480

37. Cicek H, Sunohara M, Wilkes G, McNairn H, Pick FR, Topp E. et al. Using vegetation indices from satellite remote sensing to assess corn and soybean response to controlled tile drainage. Agric. Water Manage. 2010; 98:261–270. https://doi.org/10.1016/j.agwat.2010.08.019

38. Liu L, Wang J, Bao Y, Huang W, Ma Z, Zhao C. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images. Int. J. Remote Sens. 2006; 27:737–753. https://doi.org/10.1080/01431160500296867

39. Ren S, Chen X, An S. Assessing plant senescence reflectance index-retrieved vegetation phenology and its spatiotemporal response to climate change in the Inner Mongolian Grassland. Int. J.Biometeorol., 2017; 61(4):601–612. https://doi.org/10.1007/s00484-016-1236-6 PMID: 27562030

40. Jiang Z, Huete A, Didan K, Miura T. Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 2008; 112(10):3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

41. Trombetta A, Iacobellis V, Tarantino E, Gentile F. Calibration of the aquacrop model for winter wheat using MODIS LAI images. Agric. Water Manage. 2015; 164:304–316. https://doi.org/10.1016/j.agwat.2015.10.013

42. Xie Q, Huang W, Liang D, Chen P, Wu C. Leaf area index estimation using vegetation indices derived from airborne hyperspectral images in winter wheat. IEEE J-STARS. 2017; 7:3586–3594. https://doi.org/10.1109/JSTARS.2014.2342291

43. Delaigle A, Hall P. Methodology and theory for partial least squares applied to functional data. Ann. Stat. 2012; 40:322–352. https://doi.org/10.1214/11-AOS958

44. Tan C, Wang D, Zhou J, Du Y, Luo M, Zhang Y. et al. Remotely assessing fraction of photosynthetically active radiation (FPAR) for wheat canopies based on hyperspectral vegetation indexes. Front. Plant Sci. 2018; 9:776. https://doi.org/10.3389/fpls.2018.00776 PMID: 29930568

45. Tan C, Wang D, Zhou J, Du Y, Luo M, Zhang Y. et al. Assessment of Fv/Fm absorbed by wheat canopies employing in-situ hyperspectral vegetation indexes. Sci. Rep. 2018; 1:8. https://doi.org/10.1038/s41598-018-27902-3 PMID: 29934625

46. Wu J, Wang D, Bauer ME. Image-based atmospheric correction of QuickBird imagery of Minnesota cropland. Remote Sens. Environ. 2005; 99:315–325. https://doi.org/10.1016/j.rse.2005.09.006

47. Turker M, Ozdarici A. Field-based crop classification using SPOT4, SPOT5, IKONOS and QuickBird imagery for agricultural areas: a comparison study. Int. J. Remote Sens. 2011; 32:9735–9768. https://doi.org/10.1080/01431161.2011.576710

48. Cheng Z, Meng J, Wang Y. Improving spring maize yield estimation at field scale by assimilating time-series HJ-1 CCD data into the WOFOST model using a new method with fast algorithms. Remote Sens. 2016; 8:303. https://doi.org/10.3390/rs8040303

49. Castaldi F, Casa R, Castrignanò A, Pascucci S, Palombo A, Pignatti S. Estimation of soil properties at the field scale from satellite data: a comparison between spatial and non-spatial techniques. Eur. J. Soil Sci. 2015; 65:842–851. https://doi.org/10.1111/ejss.12202

50. Shabani A, Ghaffary KA, Sepaskhah AR, Kamgar-Haghighi AK. Using the artificial neural network to estimate leaf area. Sci.Hortic. 2017; 216:103–110. https://doi.org/10.1016/j.Scienta.2016.12.032

51. Li D, Yang F, Wang X. Study on ensemble crop information extraction of remote sensing images based on SVM and BPNN. J Indian Soc. Remote Sens. 2016; 45:1–9. https://doi.org/10.1007/s12524-016-0597-yfigure

52. Pringle MJ, Marchant BP, Lark RM. Analysis of two variants of a spatially distributed crop model, using wavelet transforms and geostatistics. Agric. Syst. 2008; 98:135–146. https://doi.org/10.1016/j.agsy.2008.06.002