Wheat grain protein content estimation based on multi-temporal remote sensing data and generalized regression neural network

Cunjun Li[1,1], Qian Wang[1], Jihua Wang[1], Yan Wang[1], Xiaodong Yang[1], Xiaoyu Song[1], Wenjiang Huang[1]

[1] Beijing Research Center for Information Technology in Agriculture, Beijing, 100097, China

Abstract. Monitoring grain protein content in large areas by remote sensing is very important for guiding graded harvest, and facilitates grain purchasing for processing enterprises. Wheat grain protein content (GPC) at maturity was measured and multi- temporal Landsat TM and Landsat ETM + images at key stages in 2003, 2004 growth stages were acquired in this study. GPC was estimated with multi-temporal remote sensing data and generalized regression neural network (GRNN) method. Results show that the GPC prediction accuracy of the GRNN model is higher, with the average relative deviation of self- modeling, average relative deviation of cross-validation as 0.003%,0.321%；4.300%, 7.349% for 2003 and 2004 respectively. GRNN method proves to be reliable and robust to monitoring GPC in large areas by multi-temporal and multi-spectral remote sensing data.

Keywords: wheat, grain protein content, multi-temporal, Landsat, generalized regression neural network

1   Introduction

Grain protein content (GPC) is a key indicator of wheat quality, which determines the usage of wheat for making bread, cookie or noodle. In world wheat trade, the higher the GPC is, the higher price grain production manufactures will pay for wheat farmers. Monitoring GPC aids to harvest wheat according to grain protein content, and facilitates processing enterprises to purchase graded grains. It becomes possible to monitor the quality of crops using quantitative remote sensing technology development in recent years.

Many scholars have monitored wheat protein content with remote sensing technology. Wright D.L et al. (2003) found that a good relationship existed between measured wheat organ nitrogen content and vegetation indices derived from airborne near infrared image and Quickbird image at heading stage, which can be applied to guide the fertilization at anthesis. The GPC was improved after twice fertilization for wheat under nitrogen stress. But poor relationships were found between vegetation

[1] Cunjun Li, Post address: P.O.Box 2449-26,Beijing, 100097, P.R.China, Phone: +86-10-51503692, Fax: +86-10-51503750, E-Mail: lcj963@163.com

indices derived from only the image derived from heading stage. Badri etc (2003) found that, vegetation indices derived from Landsat satellite images were greatly significantly correlated with wheat GPC (r > 0.75) two weeks before anthesis. Reflectance of near-infrared wavelengths of Aster image were significantly related to GPC three weeks after flowering with lower correlation coefficient (r < 0.5). Preliminary explanations were made for monitoring mechanism. Mieke Reyniers et al. (2006) used color infrared aerial image and Cropscan geophysics spectrometer to predict wheat GPC a month ahead of wheat harvest with prediction accuracy as 90%. Petterson (2007) got satisfied prediction GPC using TCARI index in planting. Huang et al.(2004) studied correlations between total wheat leaf nitrogen and grain quality components during late maturation period. The results showed that total nitrogen can predict GPC and gluten content. Wang et al. (2004) used PPR to monitor GPC based on field plots. Hansen et al. (2003) conducted plot experiment and measured ground-based multi-spectral, wide band (visible - near-infrared wavelengths) data. Wheat yield and protein partial least squares models were established. The high yield prediction accuracy is in contrast to low protein prediction accuracy. Song et al. (2006) monitored wheat protein content based on nitrogen translocation in wheat plant and Aster images at anthesis. Zhao et al. (2005) predicted winter wheat grain protein content with water vegetation indices derived from TM image acquired at anthesis. Analyzed relationships between GPC and nitrogen transport process and moisture stress.

Monitoring winter wheat grain protein content at field scale experiment with different level of nitrogen fertilization at the key growing stage is very common. However, at large scale wheat GPC will be affected by soil, weather conditions, management and other factors. These factors along with cultivars together contribute spatial variability of GPC. Only one remote sensing image at certain period can't untangle all the factors aforementioned. Multi-temporal images overcome the shortcomings of single image and can monitor the major factors that cause the spatial variety of wheat grain protein content. Therefore, in this paper we attempt to use multi-temporal remote sensing images and GRNN method to monitor GPC.

2 Material and Method

2.1 Research region

More than twenty wheat fields, distributed randomly, flat, uniformly grown, with areas more than 50 acres, were selected for this study in Beijing rural area, China in 2003, 2004. Regular management were adopted on these fields. Local popular varieties were planted on the fields, such as Jingdong 8, Jing 9507 and so on. The fields are shown in figure 1.
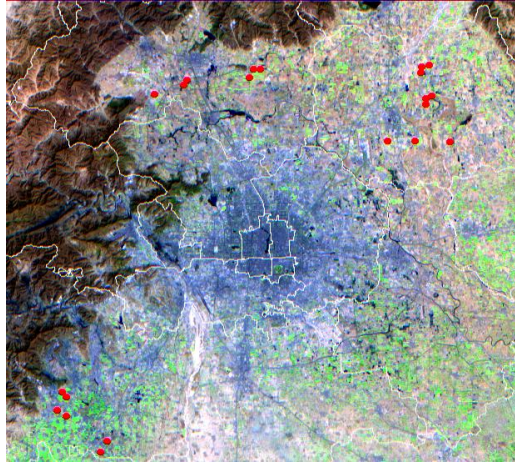
Fig. 1. Study area and the distribution map of wheat plot surveyed

2.2 Data acquisition and preprocessing

One month after wheat maturity, GPC was tested with NIR (Foss1241,Foss Tecator, Sweden).

The time and image type of the satellite data obtained include Landsat TM and Landsat ETM+ in wheat growing season of 2003 and 2004, which are shown in table 1.

Table 1. The satellite data Obtained.

YEAR	DATE	TYPE
2003	April 7,	ETM+
	May 9,	ETM+
	May 17,	TM
	May 25	ETM+
2004	April 1	TM
	April 17,	TM
	May 18	TM

The radiation correction adopts Empirical experience linear method (Liang, 2003). High radiation body and low radiation body with flat land were determined. Through experience linear spectral reflectance method and remote sensing image transformation, DN values of target true reflectance were calculated. Low radiation body and high radiation body were Ming tombs reservoir and drying yard of Xiaotangshan precision agriculture demonstration base respectively.

Geometric correction of satellite data includes the following steps. First of all, 1:100,000, topographic map in the study area was used as base map. Second, the April 7th, 2003 Landsat image was corrected with control points of the base map. Finally, the error should be controlled within a pixel. Other images were also

geometrically corrected in the same way. Beijing 54 coordinate system was adopted for projection with the center longitude as 117 °.

2.3   Data analysis

2.3.1   Extraction of remote sensing spectral index

According to Landsat band set and previous studies, seven multi-spectral vegetation indices were selected to monitor wheat nitrogen content, biomass, leaf chlorophyll density, and leaf equivalent water thickness. NDVI 21, NDVI 42 are sensitive to chlorophyll density, nitrogen content. NDVI47, RVI43 are sensitive to the biomass and leaf equivalent water thickness.

Table 2. Senven muti-spectral vegetation indices.

Abbreviation	Index	Formula
NDVI42	NDVI of 4 and 2	NDVI42=(B4-b2)/ (B4+b2)
NDVI43	NDVI of 4 and 3	NDVI43=(B4-b3)/ (B4+b3)
NDVI47	NDVI of 4 and 7	NDVI47=(B4-b7)/ (B4+b7)
NDVI21	NDVI of 2 and 1	NDVI21=(B2-b1)/ (B2+b1)
NDVI57	NDVI of 5 and 7	NDVI57=(B5-b7)/ (B5+b7)
RVI43	RVI of 4 and 3	RVI43=B4/B3
RVI54	RVI of 5 and 4	RVI54=B5/B4

*1，2，3…7 refer to band of LANDSAT TM/ETM+

2.3.2   Generalized regression neural network

Neural network is a modeling method in recent years. Generalized regression neural network (GRNN) is a memory-based feed forward network.

The GRNN consists four layers: input layer, pattern layer, summation layer and output layer (Specht, 1991; Chtioui, 1999; Haidar, 2011). Figure 2 shows a schematic diagram of generalized regression neural network architecture. GRNN has a special property in which no iterative training of the weight vectors is required. That is, any input–output mapping is possible by simply assigning the input vectors to the centroid vectors and fixing the weight vectors between the radial basis function units and outputs identical to the corresponding target vectors. This training algorithm is much better than the back propagation training, which involves long and iterative training as well as facing the problem of local minima. Moreover, a special property of GRNN is that it enables users to flexibly configure the network suitable for real hardware implementation, by adjusting only two parameters which are the center and width. Since the radial basis function acts as a detector for different input vectors, the weight

vectors are computed accordingly and there is no need to train the network. Thus, the GRNN is more straightforward and does not require a training process.
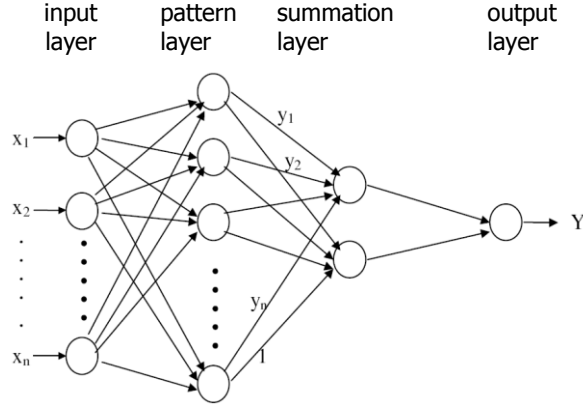

Fig. 2. Schematic diagram of GRNN architecture.

In GRNN, If $f(X,y)$ represents the known joint continuous probability density function of a vector random variable, X, and a scalar random variable, y, the conditional mean of y given X (also called the regression of y on X) is given by

$$E[y \mid X] = \frac{\int_{-\infty}^{\infty} y f(X,y) dy}{\int_{-\infty}^{\infty} f(X,y) dy} \qquad (1)$$

Where f (x,y) is the joint density and can be estimated by using Parzen's nonparametric estimator. Substituting Parzen's nonparametric estimator for f (x,y) and performing the integrations leads to the fundamental equation of GRNN:

$$\hat{Y}(X) = \frac{\sum_{i=1}^{n} Y^i \exp(-\frac{D_i^2}{2\sigma^2})}{\sum_{i=1}^{n} \exp(-\frac{D_i^2}{2\sigma^2})} \qquad (2)$$

Where σ is the smoothing parameter (sigma weight) and $D_i^2 = (x - x_i)^T (x - x_i)$. Training the GRNN involves finding the optimal values for the σ parameters in eq(2), whose optimal value is often determined experimentally (Kim,2003).
In this research, Matlab was used to train GRNN for protein estimation using muti-temporal Landsat TM/ETM+ data.

2.3.3 Model validation

The small number of selected sampling fields was about 20 in the study due to the large area, leading to some verification problems. Cross-validation method is used in this study. .Assume the number of all the samplings is n. Each time a sampling is left

out to validate the models established on the other n-1 sampling. In this way, all the n samplings are used to validate corresponding models. Deviation is calculated as: bias = | predicted value- actual value| / actual value× 100%. So according to the deviation the predictive power of modeling can be compared.

3   Results

Multi-temperal Landsat TM/ETM+ image from 2003 and 2004 were chosen as source data to estimate wheat grain protein content by GRNN method. Figure 3 shows the measured grain protein content versus predicted values derived from multi-temporal, multi-spectral parameters of all samples in 2003 with GRNN method.. Table 3 is the validation results.
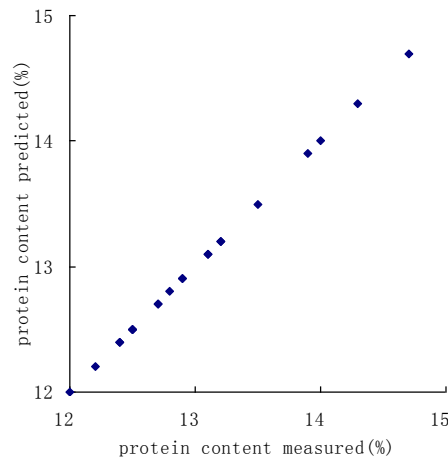

Fig. 3. Measured GPC versus predicted GPC in 2003

Table 3. Cross validation of predicted wheat GPC using muti-temporal airborne Landsat mutispectral data and GRNN method in 2003

Group	Deviation of cross-validation	Average deviation of self-modeling	Group	Deviation of cross-validation	Average deviation of self-modeling
1	0.995	0.002	12	6.459	0.001
2	2.336	0.001	13	6.733	0.001
3	7.919	0.001	14	7.541	0.001
4	2.291	0.001	15	1.676	0.001
5	6.429	0.002	16	2.613	0.001
6	6.041	0.001	17	3.914	0.007
7	1.643	0.001	18	0.688	0.001

8	11.567	0.001	19	0.355	0.003
9	9.794	0.001	20	1.078	0.013
10	0.692	0.001	21	3.788	0.015
11	5.752	0.001	Average	4.300	0.003

As shown in figure 3 and table 3, the GRNN method has potential to simulate GPC for multi-temporal and multi-spectral parameters of all samples in 2003, with high consistency and little deviation between measured and simulated ones. The average relative deviation of self-modeling and cross-validation were 0.003% and 4.300% respectively. The low deviation shows that the GRNN method has powerful ability to construct models and is a robust method to predict grain protein content.

Figure 4 shows the measured GPC versus the simulated values in 2004. 22 samples were collected in 2004, table 4 shows 22 cross-validation results for corresponding models.
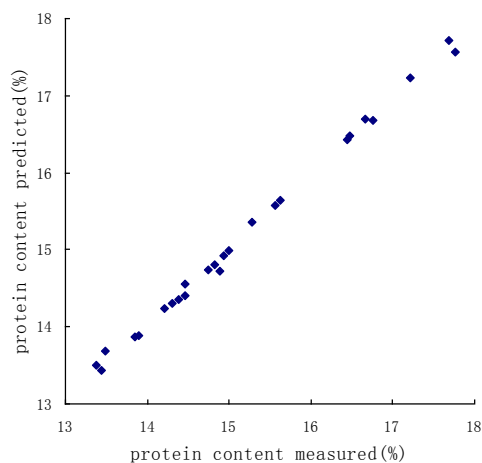

Fig. 4. Measured GPC versus predicted GPC in 2004

Table 4. Cross validation of predicted wheat GPC using muti-temporal Landsat mutispectral data and GRNN method in 2004

Group	Deviation of cross-validation	Average deviation of self-modeling	Group	Deviation of cross-validation	Average deviation of self-modeling
1	6.660	0.085	13	10.188	1.044
2	0.659	0.152	14	13.359	0.491
3	22.515	1.502	15	3.697	0.098
4	2.628	0.126	16	10.975	0.043
5	6.703	0.005	17	7.446	0.13

6	7.241	0.036	18	15.625	0.613
7	2.433	0.034	19	0.987	0.134
8	13.256	0.026	20	6.014	0.51
9	10.467	0.008	21	1.879	0.228
10	0.135	0.047	22	1.992	0.375
11	5.589	0.467			
12	11.233	0.904	Average	7.349	0.321

Figure 4 and table 4 shows that GRNN method has the potential to predict GPC with multi-temporal and multi-spectral parameters in 2004. The average relative deviation of self-modeling and cross-validation are 0.321% and 7.349% respectively. The high prediction accuracy shows that GRNN is suitable to simulate GPC.

By comparing the prediction results in 2003 and 2004, the former was better than the latter. This is perhaps because that only one image (May 8th during late period of wheat growth) was acquired in 2004, while three images including May 1st, May 9th and May 25th, were acquired during wheat growth in 2003. This is consistent with the previous conclusions that images for wheat anthesis and grain filling periods are of great importance to predict GPC (Wang, 2004).

4   Conclusion and Discussion

In this study, GRNN method proves to accurately predict GPC in large areas. Although wheat GPC were affected by soil, weather conditions, management and other factors, multi-temporal images acquired during critical growth development can overcome the low accuracy problems caused by single image.

However the distinguished weather conditions between 2003 and 2004 make it difficult to apply the models established on one year's data to predict the other year's grain protein content. Besides, it is necessary to combine meteorological data to establish prediction model in the future studies.

References

1. Badri B B, Apan A A, Kelly R M, Jensen T A, Strong W M, Butler D G. Relating satellite imagery with grain protein content. Proceedings of the Spatial Sciences Conference, 22-27, Canberra, 2003.

2. Chtioui Y, Panigrahi S, Francl L. A generealized regression neural network and its application for leaf wetness prediction to forecast plant disease. Chemometr Intell Lab Syst, 1999, 48: 47－58

3. Haidar, A.M.A., Mustafa, M.W., Ibrahim, F.A.F., Ahmed, I.A. Transient stability evaluation of electrical power system using generalized regression neural networks. Applied Soft Computing Journal 2011,11 (4):3558-3570

4. Hansen P M, Jorgensen J R, Thomsen A. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. Journal Of Agricultural Science,2002,139(3):307-318.

5. Huang W J,Zhao C J,Wang J H,Wang J D,Ma Z H.Application of red edge variables to nutrition diagnosis and grain quality forecast of winter wheat.Transactions of The Chinese Society of Agricultural Engineering,2004, 20(6):1-6.(in Chinese)

6. Kim, B., Kim, S., Kim, K., "Modelling of Plasma Etching using a Generalized Regression Neural Network", Vacuum 2003. 71:497–503.

7. Liang S.L.Quantitative Remote Sensing of Land Surfaces. Wiley & Sons, Hoboken, NJ, USA 2004.

8. Reyniers M, Vrindts E, Baerdemaeker J D. Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat. European Journal Of Agronomy,2006,24(2):87-94.

9. Song X Y, Wen J H, Wang J H, Liu L Y and Li C J. Preliminary application of ASTER images in winter wheat quality monitoring. Transactions of the Chinese Society of Agricultural Engineering, 2006, 9: 148-154

10. Specht D. F., A General Regression Neural Network , IEEE Transactions on Neural Networks,1991,2(6): 568–576.

11. Wang Z J, Wang J H, Liu L Y, Huang W J, Zhao C J, Wang C Z. Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR). Field Crops Research,2004,90(2-3):311-321.

12. Wright D L, Ritchie G, Rasmussen V P, Ramsey R D, Baker D. Managing Protein in hard red spring wheat with remote sensing. The 6th Annual National Wheat Industry Research Forum, Hyatt Regency Albuquerque, New Mexico , 2003.

13. Zhao C J, Liu L Y, Wang J H, Huang W J, Song X Y, Li C J. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. International Journal Of Applied Earth Observation And Geoinformation, 2005,7(1):1-9.