Original Article | Journal of Biosystems Engineering

J. Biosyst. Eng. 43(2):148-159. (2018. 6)
https://doi.org/10.5307/JBE.2018.43.2.148

eISSN : 2234-1862
pISSN : 1738-1266

Integrating UAV Remote Sensing with GIS for Predicting Rice Grain Protein

Tapash Kumar Sarkar[1,2], Chan-Seok Ryu[1,2]*, Ye-Seong Kang[1,2], Seong-Heon Kim[1,2], Sae-Rom Jeon[1,2], Si-Hyeong Jang[1,2], Jun-Woo Park[1,2], Suk-Gu Kim[3], Hyun-Jin Kim[3]

[1]*Department of Bio-Systems Engineering, College of Agricultural and Life Science, Gyeongsang National University, Jinju 52828, Republic of Korea*
[2]*Institute of Agriculture and Life Science, Gyeongsang National University, Jinju 52828, Republic of Korea*
[3]*Geomatics Total Service, Gwangju 61625, Republic of Korea*

Received: March 21th, 2018; Revised: May 25th, 2018; Accepted: May 31th, 2018

Abstract

Purpose: Unmanned air vehicle (UAV) remote sensing was applied to test various vegetation indices and make prediction models of protein content of rice for monitoring grain quality and proper management practice. **Methods:** Image acquisition was carried out by using NIR (Green, Red, NIR), RGB and RE (Blue, Green, Red-edge) camera mounted on UAV. Sampling was done synchronously at the geo-referenced points and GPS locations were recorded. Paddy samples were air-dried to 15% moisture content, and then dehulled and milled to 92% milling yield and measured the protein content by near-infrared spectroscopy. **Results:** Artificial neural network showed the better performance with R^2 (coefficient of determination) of 0.740, NSE (Nash-Sutcliffe model efficiency coefficient) of 0.733 and RMSE (root mean square error) of 0.187% considering all 54 samples than the models developed by PR (polynomial regression), SLR (simple linear regression), and PLSR (partial least square regression). PLSR calibration models showed almost similar result with PR as 0.663 (R^2) and 0.169% (RMSE) for cloud-free samples and 0.491 (R^2) and 0.217% (RMSE) for cloud-shadowed samples. However, the validation models performed poorly. This study revealed that there is a highly significant correlation between NDVI (normalized difference vegetation index) and protein content in rice. For the cloud-free samples, the SLR models showed R^2 = 0.553 and RMSE = 0.210%, and for cloud-shadowed samples showed 0.479 as R^2 and 0.225% as RMSE respectively. **Conclusion:** There is a significant correlation between spectral bands and grain protein content. Artificial neural networks have the strong advantages to fit the nonlinear problem when a sigmoid activation function is used in the hidden layer. Quantitatively, the neural network model obtained a higher precision result with a mean absolute relative error (MARE) of 2.18% and root mean square error (RMSE) of 0.187%.

Keywords: ANN, Grain protein, Spectral reflectance, UAV remote sensing

Introduction

Rice (*Oryza sativa* L.) is one of the most important cereal crops in the world which is mostly grown in the climatic zone. Rice is the staple food source for more than two-third of the world's population (Sasaki and Burr, 2000), especially in Southeast Asia (Nwugo and Huerta, 2011; Wang et al., 2011). The crop is planted about one-third of the world area mostly in Asia and provides 35-60% of the total calories intakes by the people. Rice has relatively lower protein content compared to other cereals but has higher protein quality due to its higher ratio of glutelin/prolamin (Kaul and Raghaviah, 1975). As a source of protein, rice is assumed to be less important than some other grains but, it is significant for the people whose daily food consumption is dominated by rice especially in South and Southeast Asia where people depend on rice as their main source of protein. Rice contributing approximately 29.1% of the dietary protein for human consumption in developing countries as an important source of protein for half the world

***Corresponding author:** Chan Seok Ryu
Tel: +82-55-772-1897; Fax: +82-55-772-1898
E-mail: ryucs@gnu.ac.kr

population (Sautter et al., 2006). In other point of view, grain quality is very important as it paid a premium price for high-quality grain. Therefore, it is very important to monitor the protein quality in the rice field.

Remote sensing is becoming increasingly important and gaining attention in agriculture that offers a more realistic alternative to the laboratory-based N analysis. As laboratory method is time-consuming and costly, remote sensing could provide spatial and temporal measurements of surface properties and was recognized as a reliable method for the estimation of various variables related to physiology and biochemistry (Hinzman et al., 1986; Diker and Bausch, 2003). Remote sensing offers to measure N status at site-specific, non-destructive, large-scale, and economical way and could be used to monitor N status as leaf chlorophyll concentration is mainly determined by N availability (Filella et al., 1995). Rice paddy field monitoring using visible and NIR camera mounted with radio-control helicopter has proposed to develop the method the rice quality evaluation through nitrogen content in rice leaves (Arai et al., 2014). Several studies have claimed to have found the effect of N availability on canopy spectral reflectance. Some particular bands or combined indices were established to detect N status of plants (Thomas and Oerther, 1972; Kleman and Fagerlund, 1987; Filella et al., 1995; Blackmer et al., 1996; Sembiring et al., 1998). Chlorophyll related pigments and leaf cell structure are predominantly influenced the plant canopy spectral reflectance in the visible (400-700 nm) and NIR (600-900 nm) wavelengths (Bonham-Carter, 1987; Vogelmann, 1993; Gitelson and Merzlyak, 1997). The most commonly employed vegetation index is normalized difference vegetation index (NDVI), which is sensitive to vegetation growth status, productivity, and other biophysical and biochemistry characteristics (Boken et al., 2002). Since, grain protein content is correlated with chlorophyll concentration and N concentration, it might be possible to use some vegetation indices and reflectance data to predict the protein content in rice. However, only a few studies have been conducted to predict rice grain protein prior to harvest (Ryu et al., 2011). Hence, this research aims to build the protein content (PC) prediction model using the unmanned air vehicle remote sensing reflectance data.

In the field of agriculture, geographic information systems (GIS) finds its way to be used in a magnificent way and thus the use of GIS is being accelerated in pace over time. Global Positioning System (GPS) offers the possibility to attribute the spatial coordinates of the field data. Correspondingly, it is promising to determine and record the accurate position continuously. Considering this, with the availability of more details in the field of agriculture, it provides a larger database for users. GIS is essential in order to storage and handling of data (Lee, 1997). The information from remote sensing imagery into GIS database and maps of land cover or land use could be achieved simultaneously in addition to saving time and money (Mandal and Gghosh, 2000). The key management factors (timing and period of midseason drainage, the timing of fertilizer application and harvest) can be monitored by GIS intervention that may cause the variation of protein content over the large fields (Ryu et al., 2011).

Artificial neural networks (ANNs) have the ability to deal complicated spectral information with target attributes and complex nonlinear relationships which exist between spectral signatures and various crop conditions without any constraints for sample distribution (Gorr et al., 1994; Kimes et al., 1998). In addition, partial least squares regression (PLSR) is also an important statistical method that bears some relation to principal components analysis (PCA), instead of finding hyperplanes of maximum variance between the response and independent variables. It can use fewer new variables than the original ones to figure out the difficult analysis such as the superposition of a spectral band and find a linear regression model by projecting the predicted variables and the observable variables to a new space (Rännar et al., 1994; Tenenhaus et al., 2005).

Few studies implemented the regression model between reflectance and plant nitrogen by using MLR, ANNs, and PLSR (He et al., 2006; Yi et al., 2007). However, the predictive ability of the three modeling methods using fresh canopy-level spectral reflectance was not well compared. ANN model provided better accuracy in the retrieval of rice neck blasts compared with the results from the MLR model (Zhang et al., 2012).

The research was proposed to evaluate rice quality through protein content in rice crop with observation of VIs and reflectance data which were acquired with RGB, RE, and NIR camera mounted on low altitude unmanned air vehicle (UAV).

Materials and Methods

Test field

The study site was located at Pyeongteak of Gyeonggi province in South Korea (latitude: 37° 1' 34.4" N, longitude: 126° 49' 45.2" E, UTM Zone 52S, Elevation 27 m). A total area of around 240 ha comprised 54 fields ranged between 0.3 and 0.5 ha each was covered under the study.

Sampling and protein content measurement

Koshikari was the variety transplanted in all fields and usual farmland management was adopted. Sampling was done synchronously at the time of image acquisition at the geo-referenced points on September 11, 2016, and GPS locations were recorded one week before harvesting. In order to collect the rice sample, ten hills from each field were harvested manually from the center of the field and dried until the moisture content of the sample reached at 15%. Hereafter, the dried paddy sample was milled to produce brown rice. Each brown rice sample was milled to 92% milling yield by a polishing machine (VP-32T, Yamamoto Co. Ltd., Yamagata, Japan). The protein content (PC) of milled rice was then measured by near-infrared spectroscopy (Infratech TM 1241 Grain Analyzer, FOSS, Höganäs, Sweden).

Image acquisition

Three imaging sensors in RGB, NIR, and RE camera (PowerShot S110, CMOS, Canon, Tokyo, Japan) were mounted with fixed-wing unmanned air vehicle (UAV) (eBee, Sensefly, Switzerland) to acquire the images (Fig. 1).

The specification of the imaging sensors employed for detecting spectral reflectance of the electromagnetic spectrum from the canopy or plant tissue is shown in Table 1.

The UAV was operated by an Autopilot associated with autonomous flight (eMotion 2, Sensefly, Switzerland) based on differential GPS (DGPS), and flew at a height around 150 m. The ground resolution was around 4 cm during flight. The flight plan software (Emotion, Sensefly, Switzerland) on the ground that controls the flight path to acquire the images from the experimental field as shown in Figure 2. In addition, UAV was connected through a radio link where position, altitude, and status data were transmitted at 2.4 GHz frequency within 3 km range. The images were acquired at midday to reduce the influence

Figure 1. Image of drone and camera used for image acquisition

Table 1. Specifications of the cameras used for image acquisition

Camera	Band	Wavelength, nm	Camera resolution
NIR	Green	550	4896×3672
	Red	625	
	NIR	850	
RE	Blue	450	4896×3672
	Green	500	
	Red edge	715	
RGB	Blue	450	4048×3048
	Green	520	
	Red	660	

of incident light angle and dew on the leaf surface (Onoyama et.al., 2013).

Image processing

All images acquired from each flight were combined with Gyro and GPS information using flight plan software, then these images were mosaicked by UAV mapping software (Pix4D mapper pro, Pix4D, Switzerland) considering each band and finally tagged image file format (TIFF) of 8 bit/pixel was produced. The sampling points in the georeferenced images were identified by UAV mapping software using GCPs (ground control points) through the quadratic equation shown in Figure 3. After that region of interest at sampling points is fixed and the reflectance data at each sampling point is calculated by the image processing software (ENVI 4.7, Exeils Visual Information Solution Inc., USA). Reflectance panel was set in the middle of the field to compensate the illumination and atmospheric effect and the reflectance spectrum of the canopy at sampling points was calculated from the mean spectral value of the canopy (Ref_{canopy}) dividing by the mean spectral value of the reference board (Ref_{board}) to minimize solar-induced disturbances (Ryu et al., 2014) as shown in equation (1)

$$Ref_{adj} = \frac{Ref_{canopy}}{Ref_{board}} \quad (1)$$

Data analysis

Simple linear regression (SLR) and polynomial regression (PR) model were employed to investigate the best vegetation indices to predict the protein content in rice grain by the statistical software Origin 9.0 (OriginLab Corporation, Northampton, Massachusetts). The main purpose of principal component analysis is to build the linear combinations of the original variables that represent the most original variations of the data set being investigated. The protein content prediction models were built by artificial neural networks (ANN) and partial least square (PLS) regression using the spectral reflectance data.

ANN analysis was done by MATLAB (R2015a, The Mathworks, Inc., Natick, MA, USA) to develop a model in order to predict protein content in rice before harvest. Usually, ANN states to the interconnected nodes or units known as neurons produce the basic building block to develop a network which works in the same manner as human central nervous system does. ANN is greatly applicable on very complex relations and situations particularly in the case of remaining strong non-linearity between different variables and the parameters. A two-layer feed forward neural network was established that trained with Levenberg-Marquardart (LM algorithm). In this study, "trainlim" train function was used as it is the fastest training function for back-propagation algorithm. Hence, ANN training was accomplished by adjusting the weight and bias values between the elements in accordance with Levenberg-Marquardt optimization. The performance of ANN models was assessed by root mean square error (RMSE) and coefficient of determination (R^2). RMSE and MARE for calibration, validation, and overall performance were measured to determine the accuracy and R^2 and NSE were calculated for ANN model to quantify the performance of the model.

Partial least square regression analysis was done by R project (version 3.2.3, R Foundation for Statistical Computing, Austria) to assess the predictive power of the relationship between protein content and the spectral reflectance of the Green, Red, and NIR bands. In principle,

Figure 2. Flight plan over the study area

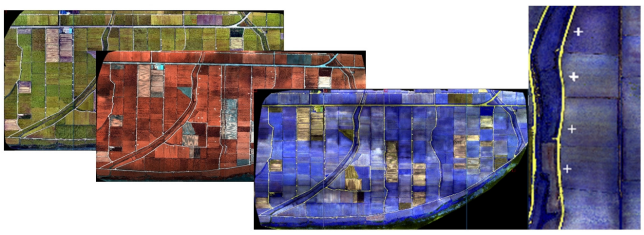

Figure 3. Mosaicked image and sampling point in the experimental field

PLS is a bilinear modeling approach for relating the variations in one or several response variables to the variations of several predictors (Esbensen, 2002).

The PLS regression was basically developed as an extension of the more familiar PCA. The most used fields of PLS is in chemometrics or spectroscopy being used to originate calibration equation to predict chemical composition of sample from NIR spectra (Reeves, 2001; Downey et al., 2002).

Model validation is very important if it is considered to be used predictively i.e. for prediction of estimation error and to avoid over-fitting and under-fitting that often results in a lower quality of prediction (Esbensen, 2002). In this research, full cross-validation (leave-one-out) technique was applied due to the lower number of samples. The validation process was carried out in the following order: a) one sample is omitted from the calibration, b) the model is built from the remaining samples, and c) lastly, the omitted measurement was used to predict using the model. This process is repeated until all samples have been omitted once.

The RMSE is the square root of the variance of the residual indicates the absolute fit of the model to the data-how close the actual data points are to the model's predicted values. The performance of the model was determined by calculating the RMSE which is a measure of the average difference between the measured and predicted values. It can be interpreted as the average error of prediction, expressed in the same unit as the original measured value (CAMO, 2004). The error can be calculated by the following equation (2) as follows:

$$RMSE = \sqrt{\frac{\sum (y_m - y_p)^2}{n}} \quad (2)$$

Where, y_m is the measured grain protein content, y_p is the grain protein predicted by the model, and n is the number of measurements in the training sample set during calibration or cross-validation.

The model performance was also assessed for calibration and validation data based on the model's correlation coefficient (R^2) for predicted versus measured compositions, and ratio of performance to deviation (RPD). The RPD is defined as the ratio of the standard deviation of the reference samples STDEV divided by the RMSE which is a measure of the robustness of the model.

Results and Discussion

The distribution of data excluding outliers is shown in boxplot in Figure 4. There is a small variation in data range since same variety and management practice were adopted and the range varies between 5.1 and 6.6. Fifty percent of the data lies between 5.4 and 5.9 in first and third quartile. However, 75% of the data for protein content are less than 5.9%, but the upper 25% data shows a higher protein content up to 6.6%. This data is skewed. One of the reasons of this variation of protein content even after the same management is topography of land. Few fields were high land resulting early drain out of water caused the early leaf senescence, on the other hand, some medium to low land fields retained water left the fields more green. Moreover, crop in a number of fields were lodged caused the variation and difference in not only protein content but also the reflectance value in the electromagnetic spectrum incurred a complex relation among variables.

Relation of NDVI and protein content in rice

In order to predict protein content, numerous vegetation indices (VIs) for NIR, RE, and RGB images were calculated by using different algorithms using different visible, near infra-red and red edge wavebands of the electromagnetic spectrum. The correlation between VIs and PC were assessed. The VIs calculated from RGB and RE images showed no correlation with PC while the NDVI derived from NIR image showed a correlation that

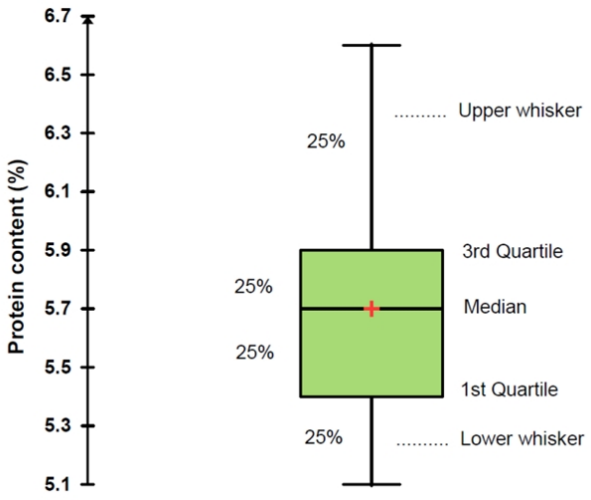

Figure 4. Protein content distribution in the experimental fields

was better than any other indices (R^2=0.210, RMSE= 0.327%) with final grain protein content followed by RVI (R^2=0.194, RMSE=0.329%) and GNDVI (R^2=0.133, RMSE=0.342%) considering all 54 samples.

As the shadow effect in the images was found therefore, the total images were divided into two groups as cloud-free and cloud-shadowed specified by (Ryu et al., 2011). All samples were grouped into cloud-free and cloud-shadowed according to the color value (screen) and pixel data value. The color value for cloud-free samples was set as NIR>180, Red>150, and Green>150. The samples with NIR<180, Red<150, and Green<150 were considered as cloud-shadowed and these two terms were used in the rest of the paper. Table 2 shows the results of the linear and polynomial regression models of rice protein content using NDVI. The NDVI is derived by the equation as NDVI = (NIR-Red)/(NIR+Red). All the F-value of the linear regression models were greater than F-critical value, revealed that all of the prediction models for PC passed the significance level test of 0.01. The results of SLR estimation model showed that the R^2 and the RMSE values were 0.553 and 0.210% respectively for the cloud-free samples. Also, for the cloud-shadowed samples, the established model yielded the R^2 and RMSE as 0.479 and 0.225% respectively after discarding five samples as outliers. Thus, the rice protein content can be predicted by the vegetation index NDVI as there is a significant correlation between NDVI and protein content.

According to the relationship between NDVI and PC, the statistical PR models were developed with a view to improve the model performance if there is any data fluctuation. From the PR models, it was observed that the R^2 and RMSE values were 0.627 and 0.199 for the cloud-free sample implied a better accuracy than SLR model while the model for cloud-shadowed samples showed R^2 value as 0.482 with similar RMSE which is identical to SLR model (Table 2). Moreover, the F-test value of the PC evaluation models was greater than the F-critical value (Table 2) indicated that the models passed a 0.01 significant level test. It revealed that there is a significant relationship between NDVI and PC and both the models are capable of predicting PC before harvest.

Principle component analysis and partial least square regression

Dimension reduction is one of the major task for multivariate regression analysis. PCA is applied without the consideration of the correlation between the dependent and independent variables termed as an unsupervised dimension reduction method while PLS is applied based on the correlation. PCA search for a few linear combinations of the variables that can be used to summarize the data without losing too much information in the whole process. The principle components are obtained by eigenvalue decomposition of the covariance or correlation matrix of the predictive variable under consideration. Eigenvalues are a special set of scalars associated with a linear system of equations (matrix equation) that are sometimes also termed as characteristic roots, characteristic values (Hoffman and Kunze, 1971), proper values, or latent roots (Marcus and Minc, 1988). The higher the eigenvalue refers to higher the variability present in the respective component. In PCA, the eigenvalues calculated from scree plot is shown in Table 3. The results show that

Table 2. Results of linear and polynomial regression models for cloud-free and cloud-shadowed samples

Model	Sample	R^2	F-value	F-critical	RMSE (%)	Equation (PC=slope×NDVI+intercept)
Linear	Cloud-free (n=14)	0.553	14.878	9.33	0.210	y=6.742x+2.472
	Cloud-shadowed (n=35)	0.480	30.450	7.47	0.225	y=3.938x+3.748
Polynomial	Cloud-free (n=14)	0.633	9.498	7.21	0.199	$y=63.921x^2-52.512x+16.131$
	Cloud-shadowed (n=35)	0.482	14.857	5.34	0.228	$y=-4.298x^2+8.016x+2.794$

Table 3. Variability explained in the principle components

Parameter	F1	F2	F3
Eigenvalue	2.69	0.30	0.01
Variability (%)	89.58	9.97	0.45
Cumulative (%)	89.58	99.55	100.0

the resultant eigenvalues for three components were 2.9, 0.30 and 0.01 respectively. Hence, the relative contribution of the first eigenvalue is 89.58% which indicate that about 90% of the variability of the data set is explained in the first principal component and more than 99% of the variability in the data is explained in the first two components (Table 4).

The variables i.e. Green, Red and NIR contributed equally to the first factors, Green and NIR to the second while Green and Red contributed higher to the third factor. There was a very minimal contribution of the Red and NIR to the second and third factor respectively (Table 4).

It was observed from the PCA that the Green and the Red band were highly correlated, also the NIR and the Red band were significantly correlated while the Green and the NIR showed little lower correlation. Protein content is negatively and marginally correlated with the Green and the Red band and little but positively correlated with the NIR band (Table 5).

Three different PLS models were developed considering cloud-free, cloud-shadowed and all samples as shown in Table 6. The performance of the models was assessed by the model's R^2 for measured versus predicted compositions, RMSE, and RPD. This was done for both calibration and validation using the corresponding data sets. Goodness of prediction is evaluated following the classification by (Chang et al., 2001). According to the classification, R^2 greater than 0.80 and RPD values greater than 2.0 as indicators for excellent prediction models. R^2 between 0.50 and 0.80 and RPD values between 2.0 and 1.4, were considered as models of medium quality which are useful for quantitative predictions in most applications. Models with R^2 lower than 0.50 and RPD lower than 1.4 are to be ranked not useable. In this study, calibrated prediction models show R^2 between 0.66 and 0.336, RMSE between 0.169% and 0.287, and RPD between 1.789 and 1.268. Even the R^2 values of cloud-free and shadowed sample are reasonable, lower RPD values indicate reduced predictability. Validation results based on leave-one-out (LOO) approach confirm this with very low R^2 (0.232 to 0.395) higher RMSE (0.236 to 0.316), and poor RPD (1.115 to1.306). From the results of the PLS regression, it is evident that the protein content prediction models for calibration are of medium quality while validation reveals drawbacks in prediction accuracy. The lower number of sample might be the reason of lower value of R^2 in validation model for cloud-free sample. Hence, the models are not to be perfect to predict the protein content more precisely with the lower validation accuracy.

Table 4. Contribution of variables to the factors

Variable	F1	F2	F3
Green	34.25	24.72	41.03
Red	36.15	6.96	56.89
NIR	29.59	68.31	2.09

Table 5. Correlation matrix (Pearson (n))

Variable	Green	Red	NIR	PC
Green	1	0.98	0.73	-0.10
Red	0.98	1	0.81	-0.13
NIR	0.73	0.81	1	0.15
PC	-0.10	-0.13	0.15	1

Table 6. Results of PLSR estimation models for calibration and validation

Model		Calibration			Validation		
		R^2	RMSE (%)	RPD	R^2	RMSE (%)	RPD
PLSR	Cloud-free (n=14)	0.663	0.169	1.789	0.262	0.247	1.224
	Cloud-shadowed) (n=35)	0.491	0.217	1.421	0.395	0.236	1.306
	All samples (n=54)	0.366	0.287	1.268	0.232	0.316	1.151

Artificial neural network prediction model

A two-layer feed-forward neural network with sigmoid activation function in the hidden layers and linear activation function in the output layer was used to fit the multi-dimensional problem using MATLAB R2015a. The network consists of one input layer, one hidden layer and one output layer. In this study, the spectral reflectance Green, Red, and NIR was used as input variables while protein content values as target variable. The number of neurons in the hidden layer was optimized by using the training, validation, and testing data sets. The network was trained with Levenberg-Marquardt backpropagation algorithm. The dataset was divided randomly into 70% training, 15% validation, and 15% for testing. The numbers of neurons in the hidden layers were determined when the minimum values of bias were found. Hence, bias was used as an additional parameter to adjust the output along with the weighted sum of the inputs to the neuron. After long trials using different hidden neurons, the ANN model with a 3-14-1 architecture was developed. The mean square error (MSE) calculated owing to define the best performance of the model shown in equation 3.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(PC_{i(actual)} - PC_{i(pred)})^2 \quad (3)$$

The simulated performance of training, validation, and testing is shown in Table 7.

It was found that the maximum outcome produced in the case of 14 hidden neuron which was selected for the proposed model showed a significant performance as the final mean square error significantly lower than the acceptable range. The best validation performance obtained with the MSE of 0.060.

Regression value is very important to determine the strength of the actual and predicted value generated by ANN simulation during training the network. Correlation coefficient (R) measures the correlation between actual and predicted value that lies between 1 and 0. The R value 1 indicates an exact linear relation between the actual and predicted value while 0 indicates no linear or random relationship. From Table 7, it is seen that the R values for calibration, validation, testing, and all data sets are 0.918, 0.796, 0.712, and 0.860 respectively. This result revealed that the predicted protein content values are close to the actual values for all data sets and the overall accuracy of the model is significantly high.

The overall error and prediction accuracy of the ANN model for predicting protein content in rice is presented in Table 8. The results show that, MARE calculated from the simulated and actual value is very accurate as per prediction accuracy evaluation specified by (Lewis, 1982). It is also revealed that 74% of the variance can be explained by the proposed ANN model (R^2=0.740). Nash-Sutcliffe model efficiency coefficient (NSE=0.733) is found to be very close to the coefficient of determination R^2 which is calculated to quantify how well the model simulation can predict the target variable.

Residual is the difference between observed and predicted value. Figure 5 shows the cumulative distribution function calculated from residual which is termed as error of fit (actual-predicted protein content) and

Table 7. Performance of the ANN models with 14 neurons in the hidden layer

Performance of ANN model	14 hidden neuron	
	MSE	R
Training	0.018	0.918
Validation	0.060	0.796
Testing	0.090	0.712
Overall	0.035	0.860

Table 8. Results of ANN estimation models for calibration, validation, and overall performance

Model	Calibration		Validation		Overall performance			
	Sample	R	RMSE (%)	R	RMSE (%)	R^2	RMSE (%)	MARE (%)	NSE
ANN	All samples (n=54)	0.918	0.134	0.796	0.244	0.740	0.187	2.18	0.733

probability density function (measured by normal distribution function of error, mean, and standard deviation of error). The x value is the quantity of residual that is being measured. The y is the probability that the residual assumes the value x or in other word, the y value is the fraction of data set that have a value smaller than corresponding x value. Hence, for instance, 70% of the data has a residual value of 0.1 or less. Moreover, the S-curve indicates a nonlinear pattern of the data set with long tails.

Figure 6 represents the graphical representation of actual (target) and predicted (simulated) values of protein content which is a competent manner to compare the data.

The result shows that there is a high degree of likeness between the actual and predicted data that markedly indicates a high level of accuracy. Therefore, the reliability of the model is proven and hence, it is established that ANN has a good prediction ability.

Generally, healthy vegetation absorbs most of the visible light that falls on it, and reflects a large portion of the near-infrared light, unhealthy or sparse vegetation reflects more visible light and less near-infrared light, bare soils reflect moderately in both the red and infrared portion of the electromagnetic spectrum (Holme et al., 1987). Out of the four prediction models as SLR, PR, PLSR, and ANN, it was found that ANN model performed better than other three models. The results of SLR models showed that there is a highly significant correlation between NDVI and rice protein content with the R^2 and RMSE of 0.553, 0.479 and 0.210%, 0.225% for cloud-free and cloud-shadowed model respectively. This result is roughly better than the result reported by (Ryu et al., 2011) as R^2 = 0.401 for cloud-free, R^2 = 0.250 for cloud-shadow and a general purpose model with R^2 = 0.392 that might be because of different ground resolutions used for the two studies. The fitting accuracy of PLSR (Table 6) was much lower than other models developed for this study. The validation performance of cloud-free model showed very poor performance that might be the reason of very small sample size. Hence, compared to other models established for this study, ANN has the strong advantages to fit the nonlinear problem when an activation function is used in the hidden layer. ANN technique has been used for analyzing the spectral data in order to improving crop biochemical parameters (He et al., 2006, Yi et al., 2007, 2010, Zhang et al., 2011). Hence, the overall fitting precision of ANN model for predicting PC was obtained as 0.740 (R^2) which was little lower than the previous study (Zhang et Al., 2012) that might be because of ground-based platform and spectrometer was used by the investigator.

Conclusions

Grain protein content of rice can be predicted using the spectral reflectance of canopy at grain filling stage prior to harvest using UAV remote sensing. There is a significant correlation between spectral bands and grain protein contents of rice. In this study, Artificial Neural Network showed more promising potential for predicting protein content and found the coefficient of determination R^2 as 0.740 and NSE as 0.733 with relatively lower RMSE

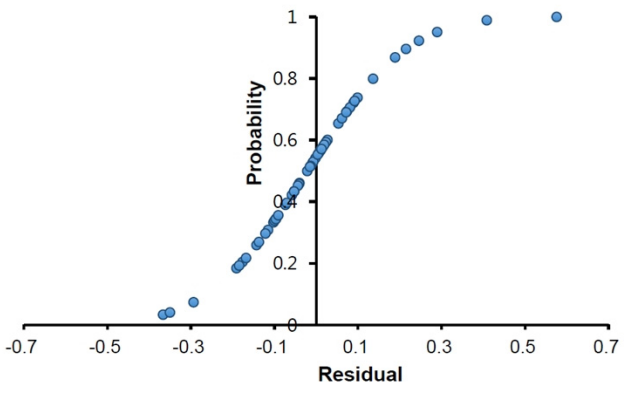

Figure 5. Cumulative distribution function calculated from measured and predicted PC

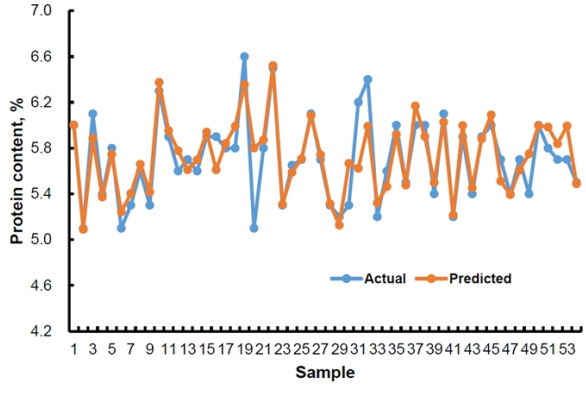

Figure 6. Graphical comparison of the actual and predicted protein content in the fields

0.187% indicates adequately good prediction ability of the model with satisfactory accuracy. The MARE calculated from the actual and predicted protein content is 2.18% indicates adequately accurate prediction with the ANN model. The results in PLS regression show that the protein content prediction model for calibrations are of medium goodness of fit while validation reveals drawbacks in prediction accuracy. Thus the PLS models are not to be perfect to predict the protein content precisely with lower validation accuracy. The NDVI calculated from NIR and Red wavebands performed better than other indices calculated from visible, red edge, and NIR bands in both linear and polynomial regression. The absolute measure of fit RMSE calculated as 0.210% for cloud-free and 0.225% for cloud-shadowed sample in linear regression whereas, the RMSE as 0.199% for cloud-free and 0.228% for cloud-shadowed sample in polynomial regression analysis. The results established that there is a highly significant correlation between NDVI and grain protein content. The RMSE in both models are less than 0.30% indicate acceptable predictive models specified by (Veerasamy et al., 2011). In summary, it is evident that ANN model is more applicable with a higher accuracy for predicting grain protein content before harvest followed by PR, LR, and PLS models.

Conflict of Interest

The authors have no conflicting financial or other interests.

References

Arai, K., M. Sakashita, O. Shigetomi and Y. Miura. 2014. Estimation of protein content in rice crop and nitrogen content in rice leaves through regression analysis with NDVI derived from camera mounted radio-control helicopter. International Journal of Advanced Research in Artificial Intelligence 3(3): 12-19. https://dx.doi.org/10.14569/IJARAI.2014.030303

Blackmer, T. M., J. S. Schepers, G. E. Varvel and E. A. Walter-Shea. 1996. Nitrogen deficiency detection using shortwave radiation from irrigated corn canopies. Agronomy Journal 88(1): 1–5. https://doi.org/10.2134/agronj1996.00021962008800010001x

Boken, V. K. and C. F. Shaykewich. 2002. Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index. International Journal of Remote Sensing 23(20): 4155–4168. https://doi.org/10.1080/014311602320567955

Bonham-Carter, G. F. 1988. Numerical procedures and computer program for fitting an inverted gaussian model to vegetation reflectance data. Computers & Geosciences 14(3): 339–356. https://doi.org/10.1016/0098-3004(88)90065-9

Chang, C. W., D. A. Laird, M. J. Mausbach and C. R. Hurburgh, Jr. 2001. Near-infrared reflectance spectroscopy-Principal components regression analyses of soil properties. Soil Science Society of America Journal 65(2): 480–490. https://doi.org/10.2136/sssaj2001.652480x

Diker, K. and W. C. Bausch. 2003. Potential use of nitrogen reflectance index to estimate plant parameters and yield of maize. Biosystems Engineering 85(4): 437–447. https://doi.org/10.1016/S1537-5110(03)00097-7

Downey, G., P. McIntyre and A. Davies. 2002. Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern mediterranean by visible and near-infrared spectroscopy. Journal of Agricultural and Food Chemistry 50(20): 5520-5525. https://doi.org/10.1021/jf0257188

Esbensen, K. H. 2002. *Multivariate Data Analysis – in practice*, Oslo, Norway: CAMO Software AS.

Filella, I., L. Serrano, J. Serra and J. Peñuelas. 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminate analysis. Crop Science 35: 1400–1405. https://doi.org/10.2135/cropsci1995.0011183X003500050023x

Gitelson, A. A. and M. N. Merzlyak. 1997. Remote estimation of chlorophyll content in higher plant leaves. International Journal of Remote Sensing 18(12): 2691-2697. https://doi.org/10.1080/014311697217558

Gorr, W. L., D. Nagin and J. Szczypula. 1994. Comparative study of artificial neural network and statistical models for predicting student grade point averages. International Journal of Forecasting 10(1): 17–34. https://doi.org/10.1016/0169-2070(94)90046-9

He, Y., X. L. Li and Y. N. Shao. 2006. Discrimination of varieties of apple using near infrared spectra based on principal component analysis and artificial neural network model. Spectroscopy and Spectral Analysis

26(5): 850–853. (In Chinese, with English abstract).

Hinzman, L. D., M. E. Bauer and C. S. T. Daughtry. 1986. Effects of nitrogen fertilization on growth and reflectance characteristics of winter wheat. Remote Sensing of Environment 19(1): 47–61. https://doi.org/10.1016/0034-4257(86)90040-4

Hoffman, K. and R. Kunze. 1971. Characteristic values. In: Linear Algebra, 2nd ed. 182-190. Englewood Cliffs, NJ, USA: Prentice-Hall Inc.

Holm, A. M., D. G. Burnside and A. A. Mitchell. 1987. The development of a system for monitoring trend in range condition in the arid shrublands of Western Australia. The Australian Rangeland Journal 9(1): 14-20. https://doi.org/10.1071/RJ9870014

Kaul, A.K. and P. Raghaviah. 1975. Influence of nitrogen fertilization on some nutritional quality characters in rice. Plant Foods for Human Nutrition 24(3-4): 391-403. https://doi.org/10.1007/BF01092224

Kimes, D. S., R. F. Nelson, M. T. Manry and A. K. Fung. 1998. Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. International Journal of Remote Sensing 19(14): 2639–2663. https://doi.org/10.1080/014311698214433

Kleman, J. and E. Fagerlund. 1987. Influence of different nitrogen and irrigation treatments on spectral reflectance of barley. Remote Sensing of Environment 21(1): 1–14. https://doi.org/10.1016/0034-4257(87)90002-2

Lewis, C. D. 1982. *International and Business Forecasting Methods. A practical guide to exponential smoothing and curve fitting.* London, UK: Butterworths Scientific Ltd.

Mandal, D. and S. K. Ghosh. 2000. Precision farming – The emerging concept of agriculture for today and tomorrow. Current Science 79(12): 1644-1647.

Marcus, M. and H. Minc. 1988. *Introduction to Linear Algebra.* Mineola, NY, USA: Dover Publications.

Nwugo, C. C. and A. J. Huerta. 2011. The effect of silicon on the leaf proteome of rice (*Oryza sativa L.*) plants under cadmium-stress. Journal of Proteome Research 10(2): 518–528. https://doi.org/10.1021/pr100716h

Onoyama, H., C. Ryu, M. Suguri and M. Iida. 2013. Potential of hyperspectral imaging for constructing a year-invariant model to estimate the nitrogen content of rice plants at the panicle initiation stage. IFAC Proceedings Volumes 46(18): 219-224. https://doi.org/10.3182/20130828-2-SF-3019.00054

Rännar, S., F. Lindgren, P. Geladi and S. Wold. 1994. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. Journal of Chemometrics 8(2): 111–125. https://doi.org/10.1002/cem.1180080204

Reeves, J. B. 2001. Near-infrared diffuse reflectance spectroscopy for the analysis of poultry manures. Journal of Agricultural and Food Chemistry, 49(5): 2193-2197. https://doi.org/10.1021/jf0013961

Ryu, C., M. Suguri, M. Iida, M. Umeda and C. Lee. 2011. Integrating remote sensing and GIS for prediction of rice protein contents. Precision Agriculture 12(3): 378–394. https://doi.org/10.1007/s11119-010-9179-0

Ryu, C., H. Onoyama, M. Suguri and Y. Kim. 2014. Estimation of the main properties in potherb mustard (Mizuna) using hyperspectral imagery. Journal of Agriculture & Life Science 48(6): 375-386 (In Korean, with English abstract). https://doi.org/10.14397/jals.2014.48.6.375

Sasaki, T. and B. Burr. 2000. International rice genome sequencing project: the effort to completely sequence the rice genome. Current Opinion in Plant Biology 3(2): 138–142. https://doi.org/10.1016/S1369-5266(99)00047-3

Sautter, C., S. Poletti, P. Zhang and W. Gruissem. 2006. Biofortification of essential nutritional compounds and trace elements in rice and cassava. Proceedings of the Nutrition Society 65(2): 153–159. https://doi.org/10.1079/PNS2006488

Sembiring, H., W. R. Raun, G. V. Johnson, M. L. Stone, J. B. Solie and S. B. Phillips. 1998. Detection of nitrogen and phosphorus nutrient status in winter wheat using spectral radiance. Journal of Plant Nutrition 21(6): 1207–1233. https://doi.org/10.1080/01904169809365478

Tenenhaus, M., V. E. Vinzi, Y. M. Chatelin and C. Lauro. 2005. PLS path modeling. Computational Statistics and Data Analysis 48(1): 159–205. https://doi.org/10.1016/j.csda.2004.03.005

The Unscrambler. 2004. Ver. 9.1, Oslo, Norway: CAMO Software AS.

Thomas, J. R. and G. F. Oerther. 1972. Estimating nitrogen content of sweet pepper leaves by reflectance measure-

ments. Agronomy Journal 64(1): 11–13. https://doi.org/10.2134/agronj1972.00021962006400010004x

Veerasamy, R., H. Rajak, A. Jain, S. Sivadasan1, C. P. Varghese and R. K. Agrawal. 2011. Validation of QSAR models - strategies and importance. International Journal of Drug Design and Discovery 2(3): 511-519.

Vogelmann, T. C. 1993. Plant tissue optics. Annual Review of Plant Physiology and Plant Molecular Biology 44: 231–251. https://doi.org/10.1146/annurev.pp.44.060193.001311

Wang, Y., S. G. Kim., S. T. Kim, G. K. Agrawal, R. Rakwal and K. Y. Kang. 2011. Biotic stress-responsive rice proteome: An overview. Journal of Plant Biology 54: 219–226. https://doi.org/10.1007/s12374-011-9165-8

Yi, Q. X., J. F. Huang, F. M. Wang, X. Z. Wang and Z. Y. Liu. 2007. Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. Environmental Science and Technology 41(19): 6770–6775. https://doi.org/10.1021/es070144e

Yi, S. L., L. Deng, S. L. He, Y. Q. Zheng and S. S. Mao. 2010. A spectrum based models for monitoring leaf potassium content of citrus sinensis (L.) cv. Jincheng orange. Scientia Agricultura Sinica 43(4): 780–786. (In Chinese, with English abstract). https://doi.org/10.3864/j.issn.0578-1752.2010.04.015

Zhang, H., H. Hu, X. B. Zhang, L. F. Zhu, K. F. Zheng, Q. Y. Jin and F. P. Zeng. 2011. Estimation of rice neck blasts severity using spectral reflectance based on BP-neural network. Acta Physiologiae Plantarum 33(6): 2461–2466. https://doi.org/10.1007/s11738-011-0790-0

Zhang, H., T. Q. Song, K. L. Wang, G. X. Wang, H. Hu and F. P. Zeng. 2012. Prediction of crude protein content in rice grain with canopy spectral reflectance. Plant, Soil and Environment 58(11): 514–520. https://doi.org/10.17221/526/2012-PSE