Quantitative analysis from unifying field and airborne hyperspectral in prediction biophysical parameters by using partial least square (PLSR) and Normalized Difference Spectral Index (NDSI)

Arief Darmawan[1], Nadirah[1], Agus Wibowo[1], Muhammad Evri[1], Sidik Mulyono[1], Anto Satrio Nugroho[1], Muhamad Sadly[1], Nani Hendiarti[1], Osamu Kashimura[2], Chiaki Kobayashi[2], Atsushi Uchida[3], Aya Uraguchi[3], Hozuma Sekine[3]

[1]*Agency for the Assessment and Application of Technology (BPPT), Jl. MH. Thamrin no. 8, Jakarta-10340, Indonesia, phone number: +62-21-3169706, e-mail: ariefd@webmail.bppt.go.id*

[2]*Earth Remote Sensing Data Analysis Center, FOREFRONT TOWER 3-12-1 Kachidoki, huo-ku, Tokyo 104-0054, Japan, phone number:+81-3-3533-9310, e-mail : kobayashi@ersdac.or.jp*

[3]*Mitsubishi Research Institute, Inc., 3-6, Otemachi 2-chome Chiyoda-ku, Tokyo 100-8141, Japan, phone number: +81-3-3277-0551, e-mail: a-uchida@mri.co.jp*

Abstract

Paddy rice canopy hyperspectral was measured by using ground-based spectroradiometer and HyMap sensor onboard an airplane from 350 nm up to 2500 nm in the same time that covered various growth stages. Coinciding with hyperspectral measurement, biophysical parameters such as leaf area index (LAI), SPAD value were measured on ground during the airplane passed over area of interest (AOI). Rice yield was measured at the time of final harvesting by random selected yield (ubinan method) for each sampling area. In finding the best correlation among canopy hyperspectral reflectance with crop development, optimal individual waveband explored with involving all possible waveband combinations to obtain the best fitted two-pair waveband related to crop biophysical parameters. Normalized Difference Spectral Index (NDSI) was applied from spectral transformations (obtained from optimal waveband selected by exploring all possible waveband) to improve sensitivity analysis on plant. Canopy hyperspectral provided information about plant, soil and water background when plant canopy don't completely cover soil surface yet. The present study was directed to examine spectral indices and establish the relationships between biophysical parameters of rice by using partial least square regression (PLSR) technique.

INTRODUCTION

Since the partial least squares regression (PLSR) was developed by Wold in the late 1960s for econometrics (Wold 1975), it has become the most reliable and popular technique which has been used in varies application of scientific researches. A lot of experimental methods introduced it as a tool to analyze chemical data in the late 1970s (Geladi and Kowalski 1986, Martens et al. 1986, Mevik and Wehrens 2007).

PLSR is a multivariate regression technique that easily handles many correlated and/or noisy variables. PLSR works by iteratively finding a multidimensional direction in the X (input or signal) space that explains the maximum multidimensional variance direction in the Y (output or response) space. Like PCA (principal component analysis), it reduces dimensionality of by identifying latent or hidden variables that correspond to eigenvectors in an eigenvalue problem. By choosing components in each step and after the first fewer components, this technique can reach the optimal model performance (Helland 1988, Martens and Naes 1989). It can be said that this technique is an extension of multiple regression analysis in which the effects of linear combinations of several predictors on a response variable (or multiple response variables) will be reanalyzed.

PLSR is especially useful when (1) the number of predictor variables is similar to or higher than the number of observations (i.e. over fitting) and/or (2) predictors are highly correlated (i.e. there is strong collinearity). The application fields of PLSR also cover situations in which there are more than one response variable. In these cases where multiple response variables are used, PLSR creates other latent factors from the linear negative regression coefficients showing a varying degree of association with the predicted response variable, from very intense to nearly null influence. After this first step, a random variable following a normal distribution was added to the predictions of the previous multiple regression model, thus establishing the final response variable.

The objective of this study was to establish the relationships between crop parameters of rice and the examined spectral index ie. NDSI calculated from ground and airborne reflectance data by using PLSR technique.

DATA

Field Campaign was conducted in Subang and Indramayu district in Java island, Indonesia from June 25 – July 3, 2008. Two regions of interest have been determined and they include the sampling areas of 500 m x 500 m square. In each sampling area, some 'quadrats' (10 m x 10 m square) were set and field measurement were carried out at ten sampling points which randomly distributed. Crops parameters such as LAI, SPAD and yield were measured for the regions where airborne hyperspectral image was obtained, in order to develop regression models.

Reflectance measurements

Spectral measurements were carried out simultaneously with the airborne image acquisition. Spectral data was

obtained using FieldSpec Pro FR Spectroradiometer (Analytical Spectral Device Inc.) at three different growth stages for five times for each of measuring points which were apart from each other intervals of two meters. The height of sensor placed at a point of measurement was 10 cm and 50 cm above canopy. In this study the growth stages will be classified in 3 growth phases: Vegetative, Reproductive, and Ripening.

SPAD Measurement

A chlorophyll meter (SPAD-502 Minolta Camera Co. Osaka Japan) was used for chlorophyll measurement on five top fully expanded leaves. Three SPAD readings (dimensionless values, 650/940 nm wavelengths transmittance ratio) were taken around the midpoint of each leaf blade. The SPAD readings will be averaged to represent the mean SPAD readings of each sampling points.

Leaf Area Index (LAI) Measurement

LAI-2000 measurement were performed at sunset or on overcast days with a single sensor mode and a sequence of one above, four below and one above readings for each sampling point. In order to reduce the influence of the adjacent plots and of the operator, a 45^0 view-cap was applied on the optics. The standards LAI-2000 outputs (five rings, 5R) were reprocessed, using the LI-COR C2000 software, in order to discard the wide viewing angle reading and to estimate the four ring (4R) LAI.

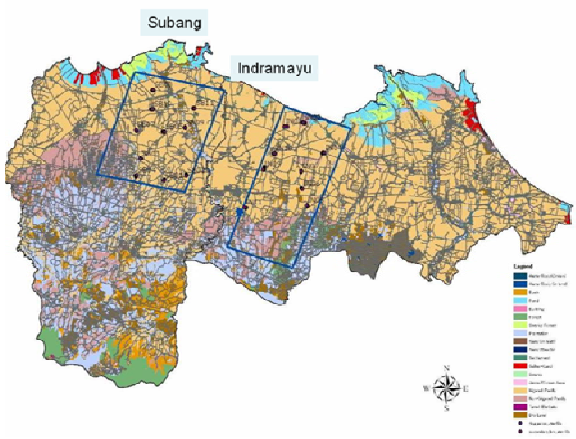


Fig. 1 ROI's and sampling areas

Yield Measurement

Rice yield was measured at the time of final harvesting for each sampling area. The measurement was implemented by Agriculture Officer and Statistical Officer using crop cutting (Ubinan).

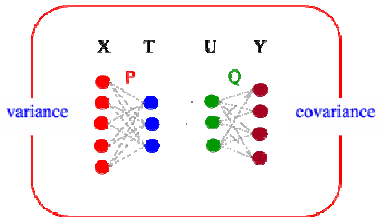
Hymap data

The airborne hyperspectral image (HyMap) was obtained over the areas with the Instantaneous Field of View (IFOV) 2.5 mr along track, 2.0 mr across track (Spatial resolution 3.5–10 m) and Field of View (FOV) 62 degrees (512 pixels). The airborne flew 11 lines started from south to east (altitude: 6500 feet or 1981 m), then one cross line and one low altitude (east to west, altitude: 4500 feet or 1372 m).

Method

The partial least squares regression

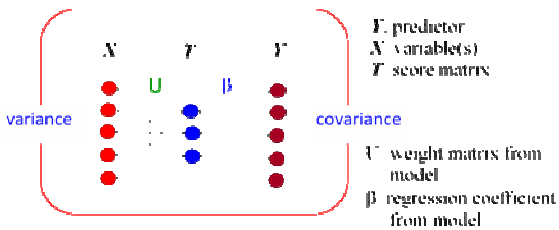
The PLSR can be illustrated as a following scheme:



Where:

Y: predictor
X: variable(s)
P : loading matrix of input latent variable
Q: loading matrix of output latent variable
U : PLS weight matrix
T : score matrix of input latent variable

It consists of outer relations (X and Y blocks individually) and an inner relation linking both blocks. The outer relation for the X block, X = TP' + E and the outer relation for the Y block, Y = UQ' + F. Simplified models would consist of a regression between the scores for the X and Y model as illustrated in the scheme below:



Finally, the best correlation among canopy hyperspectral reflectance with crop development can be estimated by using the following equation:

$$\hat{Y} = \bar{Y} + (Z - \bar{X})W_g\hat{C}$$

Where:

$\bar{X}$: Centered sample spectral matrix used in the model

W_g : Weight matrix derived from PLS1 algorithm

$\hat{C}$: Coefficient of regression derived from PLS1 algorithm

$\bar{Y}$: Centered sample biophysical parameter used in the model

$\hat{Y}$: Predictor

Z : Spectral matrix will be used for prediction

The Determinant Coefficient (R^2) and Root Mean Square Error (RMSE) can be calculated by using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - Y_p)^2}{\sum_{i=1}^{n}(Y_i - Y_m)^2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - Y_p)^2}{n}}$$

with Y_i represents the predicted crop variables from the PLSR model, Y_p denotes measured crop characteristics for the entire data set and Y_m represent the average value of Y_p.

The analysis was carried out by using fieldspec and Hymap data during field campaign (n=104). PLSR analysis was performed in the single reflectance to predict LAI, SPAD values and yield. To define the optimum number of latent variables (NLVs) on PLSR model, root mean square error was computed; those minimum values were used as the number of latent variables (Mevik and Cederkvist 2004).

Normalized Difference Spectral Index (NDSI)

Instead of Normalize Different Vegetation Index (NDVI), in this paper we used Normalized Difference Spectral Index (NDSI) to avoid confusion. The NDSI was derived from the following equation:

$$S_{NDSI} = \frac{\lambda_2 - \lambda_1}{\lambda_2 + \lambda_1}$$

where:
S_{NDSI} is Normalized Difference Spectral Index and λ_1, λ_2 ($\lambda_1 < \lambda_2$) is a reflectance at waveband 1 and 2 respectively.

Result (Indramayu site)

Table 1a, 1b, 1c, demonstrated the result of analysis when using the PLSR to predict LAI, SPAD value and yield. In PLSR, all wavelengths were thoroughly employed in the models, though their contributions were made indirectly through latent variable. The R^2 model performed the determination coefficient of crop variable and fieldspec data whereas the R^2 test performed the determination coefficient of R^2 model tested to Hymap data. R^2 test value especially for attributed SPAD and yield showed much lower value than that of R^2 model. One of the suspected reasons was because Hymap data was applied as spatial data (composite) whereas the fieldspec was applied as point data. For Hymap data does not only provide the vegetative reflectance but also its surroundings so it decreased the prediction capability. Therefore, it needs further treatments of data processing before applying it.

Leaf Area Index (LAI)

Table 1a.

Model	Height	LV	R^2 Model	R^2 Test
PLSR	FS10-HM	12	0.965	0.906
	FS50-HM	8	0.959	0.908

SPAD

Table 1b.

Model	Height	LV	R^2 Model	R^2 Test
PLSR	FS10-HM	15	0.83	0.496
	FS50-HM	16	0.82	0.462

Yield

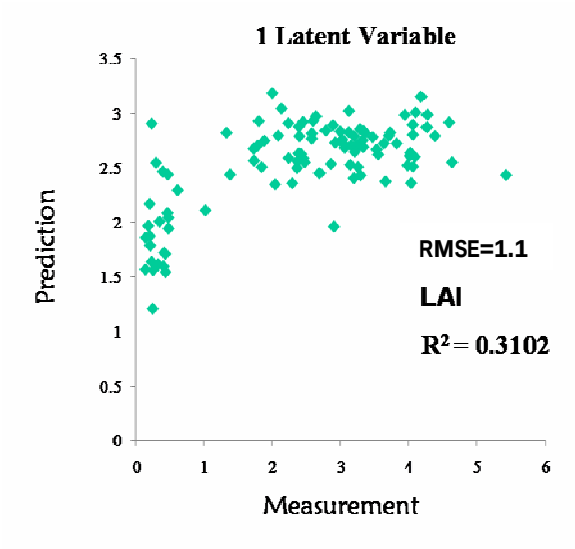
Table 1c.

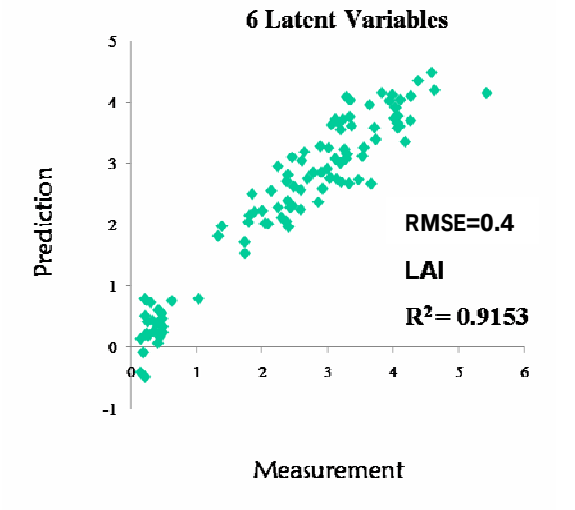
Model	Height	LV	R^2 Model	R^2 Test
PLSR	FS10-HM	11	0.864	0.147
	FS50-HM	14	0.885	0.147

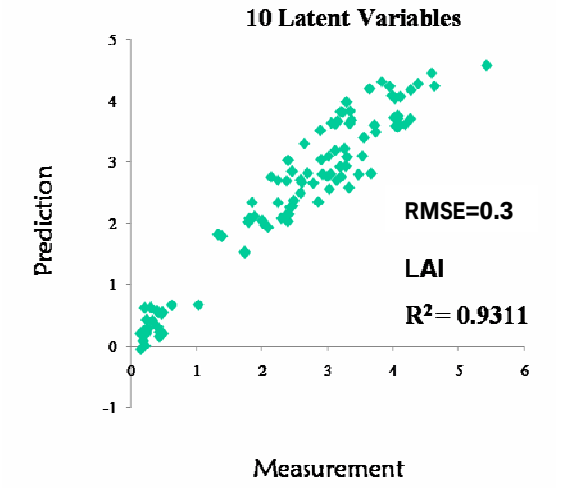
Table 1a, 1b, 1c. Coefficient of determination (R^2 model), number of latent variable (LV), cross validated coefficient of determination (R^2 test) attributed to LAI, SPAD and Yield.

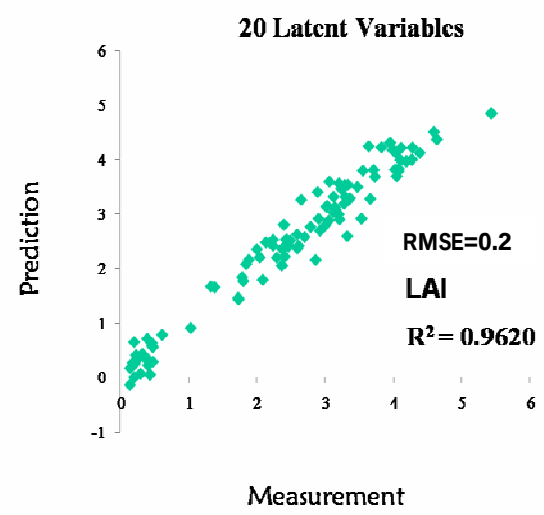
Analysis of SPAD was resulted lower R^2 value than LAI. For SPAD measurement, 1 hill was represented by 1 leaf blade measurement so the number of data (n) was very limited compared to the spectral data. For Yield data in addition to the limited of data, there is also a time-lag between the spectral measurement and rice harvested. During the time-lag, some of the vegetative disturbances could be happened probability from weather disruptions and/or from other factors

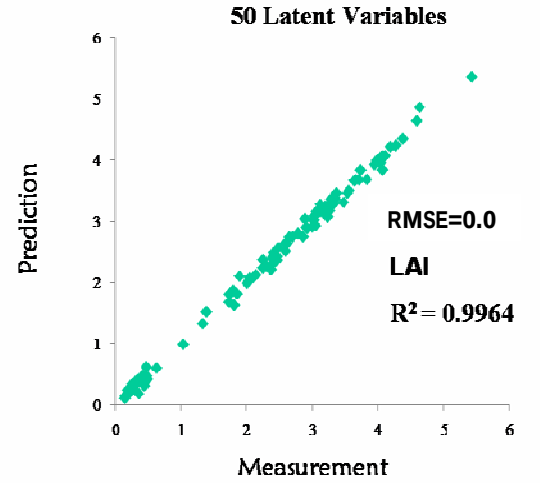
such as fertilizer and water deficiency and pest attacked.

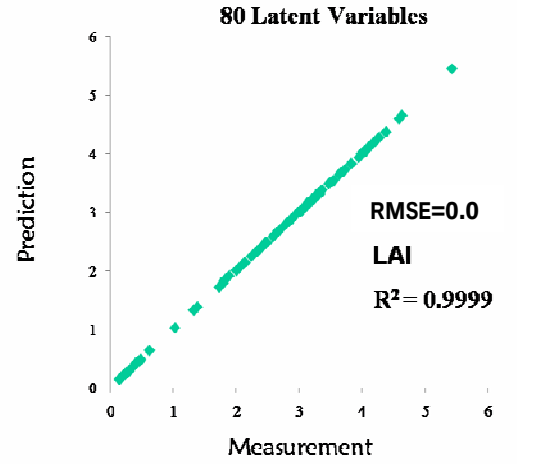












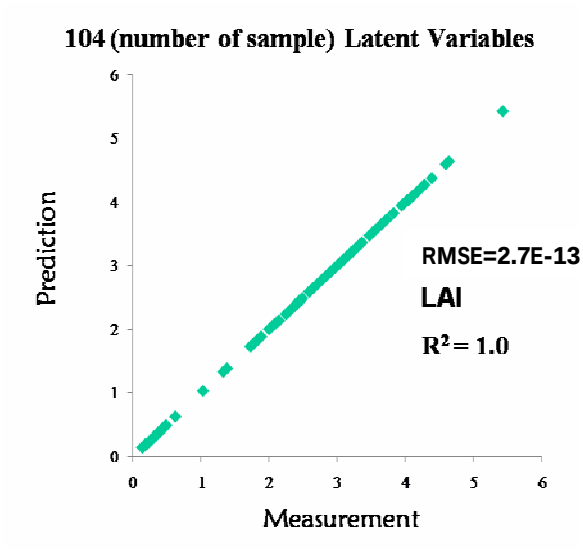
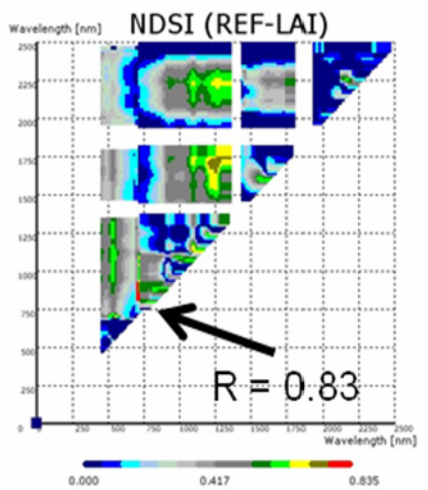
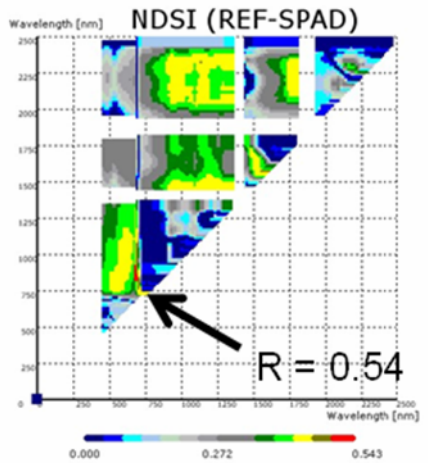


Fig. 2. Changes of the determining factor (determination coefficient, R^2) of LAI in response to the number of latent variables (NLV).

a



b



c

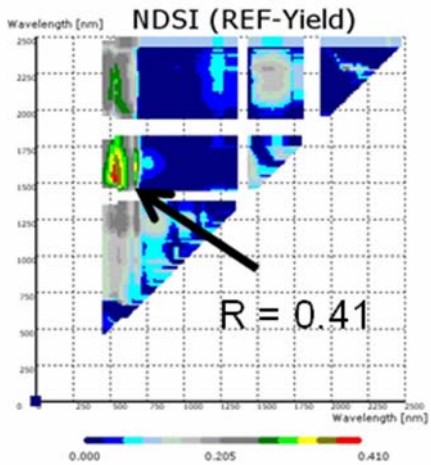


Fig. 3a, 3b, 3c. Distribution plot of coefficients of Determination (R^2) for wavebands selection. The R^2 values were obtained from the linear fitted regressions between measured rice LAI, SPAD and yield and all two band combinations in NDSI spectral index.

Conclusion

PLSR can improve the predictive capability, as indicated by increasing of R^2 and decreasing of RMSE and NLV.

For internal data, R^2 will increase when a selected number of latent is increased. It reachs value 1 when the number of latent approach or equal to the number of samples

In validation, the number of latent must be selected to get maximum of R^2.

REFERENCES

Carrascal, L.M, Ismael Galvan and Oscar Goldo. 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Oikos 118, 681-690.

Chwen-Ming Yang and Rong-Kuen Chen (2004). Modelling Rice Growth with Hyperspectral Reflectance Data. Crop Science 44:1283-1290.

Evri, M. Tsuyoshi Akiyama and Kensuke Kawamura (2008). Optimal visible and near-infrared waveband used in hyperspectral indices to predict crop variables of rice. J. JASS, 24(1).

Evri, M. Tsuyoshi Akiyama and Kensuke Kawamura (2008). Coupling hyperspectral data with principle component regression (PCR) and partial least square regression (PLSR) to improve prediction accuracy of rice crop variables. J. JASS, 24(1).

Gaston Sanchez (2009). Understanding Partial Least Squares Path Modeling (*An introduction with R*). Academic Paper Universitat Politècnica de Catalunya (UPC), Spain.

N. Yatawara (1996). Multivariate Process Monitoring with Partial Least Squares Regression: A case study. IIQP Research report RR-96-09.

S. Engelen, M. Hubert, K. Vanden Branden, and S. Verboven (September 20, 2003). Robust PCR and Robust PLSR: a comparative study. Mathematics Subject Classification. 62F35; 62H20; 92E99.