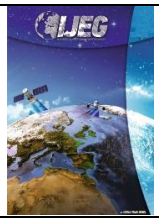
International Journal of Engineering and Geosciences
https://dergipark.org.tr/en/pub/ijeg
e-ISSN 2548-0960



# Estimating chlorophyll content of *Zizania latifolia* with hyperspectral data and random forest

**Adenan Yandra Nofrizal*[1], Rei Sonobe [1], Hiroto Yamashita [1], Akio Morita[1], Takashi Ikka [1]**

[1]*Shizuoka University, Faculty of Agriculture, Shizuoka, Japan*

**Keywords**
Dimensionality reduction
*Zizania latifolia*
Hyperspectral
Machine learning

**ABSTRACT**

The amount of chlorophyll in a plant useful to indicate its physiological activity and then changes in chlorophyll content have been used as a good indicator of disease as well as nutritional and environmental stresses on plants. Chlorophyll content estimation is one of the most applications of hyperspectral remote sensing data. The aim of this study is to evaluate dimensionality reduction for estimating chlorophyll contents from hyperspectral reflectance. Random Forest (RF) has been applied to assess biochemical properties such as chlorophyll content from remote sensing data; however, an approach integrating with dimensionality reduction techniques has not been fully evaluated. A total of 200 *Zizania latifolia* leaves with 5 treatments from Shizuoka University field were measured for reflectance and chlorophyll content. then, the regression models were generated based on RF with three dimensionality reduction methods including principal component analysis, kernel principal component analysis and independent component analysis. This research clarified that PCA is the best method for dimensionality reduction for estimating chlorophyll content in *Zizania Latifolia* with a RMSE value of 5.65 ± 0.58 μg $cm^{-2}$.

## 1. INTRODUCTION

Chlorophyll (chl) is a vital pigment which allows plants, to get energy from sunlight and convert it into organic compounds via photosynthesis (Ormeci, et. al. 2009). This content is critical for monitoring plant of health, particularly for the silicicolous plants. In this study, measurements of chlorophyll (chl *a* + chl *b*) content from Manchurian wild rice (*Zizania latifolia*) were used. *Zizania latifolia* is a tall emergent plant with well-developed underground parts and used as a food plant with both the stem and grain being edible and has been cultivated for more than 2000 years due to its rapid growth rate and high competitive ability (Wang et al. 2020). As a result, it is distributed widely in the eastern and southern areas of China with the rapid increase of the Southern Chinese population (Yan et al. 2018). Furthermore, this species is one of the silicicolous plants and then waste slag, which contain substantial proportions of heavy metals and pose a serious threat to the environment, could be used as siliceous fertilizer for cultivation (Zolotova et al. 2017, Krawiec, et. al. 2017). Then, the relationship between slag fertilization and growth rate can be expected to vary depending on the volume of slag fertilizer applied: lower rates of slag fertilization were found to increase antioxidant enzyme activity and chlorophyll content, while higher rates resulted in a reduction in these physiological and morphological properties (Chen, et al. 2019). Thus, monitoring chlorophyll content using field measurements would enable the determination of optimal slag fertilization rates and then some techniques have been required for monitoring chlorophyll content using field measurements and managing fertilizing schedules (Sonobe et al. 2021).

Generally, ultraviolet and visible (UV-VIS) spectroscopy or high-performance liquid chromatography (HPLC) measurements have been applied for quantifying leaf chlorophyll content (Prado-Cabrero et al. 2016), however, these techniques are not always applicable due to expensive, labor-intensive and require bulky equipment (Kalaji et al. 2017). Recently, the SPAD-502 Leaf Chlorophyll Meter (Konica Minolta

* **Corresponding Author**

*(adenan.y.nofrizal.20@shizuoka.ac.jp) ORCID ID 0000 – 0003 – 3176 – 0312
(sonobe.rei@shizuoka.ac.jp) ORCID ID 0000 – 0002 – 8330 – 3730
(yamashita.hiroto.16@shizuoka.ac.jp) ORCID ID 0000 – 0003 – 0571 – 8153
(morita.akio@shizuoka.ac.jp) ORCID ID 0000 – 0002 – 8418 – 2109
(ikka.takashi@shizuoka.ac.jp) ORCID ID 0000 – 0002 – 1845 – 5658

**Cite this article**

Nofrizal A Y, Sonobe R, Yamashita H, Morita A & Ikka T (2022). Estimating chlorophyll content of *Zizania latifolia* with hyperspectral data and random forest. International Journal of Engineering and Geosciences, 7(3), 221-228

**Research Article / DOI: 10.26833/ijeg.953188**

**Received: 24/06/2021; Accepted: 19/01/2022**

Inc.) was proposed for quick and *in-situ* quantifying chlorophyll. Nevertheless, some previous studies pointed out that variation in leaf thickness causes a variable relationship between SPAD readings and leaf structure, which is correlated with silica fertilizer levels (Kindomihou, et. al. 2006) and then an alternative method of quantifying chlorophyll is thus required.

Hyperspectral remote sensing, which provides spectral information on continuous wavelengths, is one of the most attractive alternative options for detection of leaf chlorophyll content and has played an important role in evaluating vegetation characteristic and management of agricultural fields (Huang et al. 2016). Especially, chlorophyll content estimation is one of the major topics on hyperspectral remote sensing (Lausch et al. 2013, Sonobe and Wang 2017b, Zhang et al. 2011). The numerical inversion of radiative transfer models (RTM) is one of the ways to estimate chlorophyll content using the hyperspectral remote sensing approaches and the PROSPECT model (Jacquemoud and Baret 1990) is the most famous RTM which simulates reflectance from given vegetation properties (Sonobe and Wang 2017b, Hernandez-Clemente, et al. 2014, Gu et al. 2016, Hunt, et al. 2016) and retrieving chlorophyll, carotenoid (Hernandez-Clemente et al. 2014, Féret et al. 2008), or dry matter content (Romero, et. al. 2012). However, most of RTMs were based on the datasets taken under relatively low-stress conditions, such as ANGERS, LOPEX and HAWAII (Féret et al. 2008). Besides, the inversion approach possesses some risks of "ill-posed" problem. Vegetation indices (VIs) have also been widely used to estimate various vegetation properties such as leaf area index (LAI), coverage, chlorophyll content, biomass, and photosynthetically active radiation (Zou et al. 2015, Huang et al. 2017). Although VIs are effective for removing variability caused by other features, such as soil background and atmospheric conditions (Blackburn and Steele 1999) and for reducing the data saturation problem (Mutanga and Skidmore 2004), the performances of the most indices changed to plant species (Sonobe and Wang 2017a).

Recently, machine learning has been applied to evaluate vegetation properties. Especially, random Forest (RF) machine learning methods performed well for assessing biochemical/physiological features vegetation (Sonobe et al. 2018, Cui et al. 2019, Fernandez-Delgado et al. 2019, Breunig et al. 2020, Sonobe, et. al. 2020). Which is the a regression technique that combine numerous decision trees to classify or predict the value of variable, has been used and reported its high performances for regression (Biau and Scornet 2016).

For applying hyperspectral remote sensing, a dimension reduction is one of the important tasks for improving the usability of hyperspectral data and generating robust regression models and it might be performed after data cleaning and data scaling and before training a predictive model (Duda, et al. 2001). Some statistical methodologies based either on Principal Component Analysis (PCA) or Independent Component Analysis (ICA), which reveals higher-order statistical information of the process data, have been widely applied to use multivariable techniques for data reduction (Fan et al. 2016, Rutledge 2018, Swiniarski and Skowron 2004). For detailing nonlinear dimension reduction, kernel principal component analysis has also been applied as the modified principal component analysis method (Li et al. 2020).

The potential main study is (1) to evaluate the potential hyperspectral data for estimation the chlorophyll content of Zizania latifolia and (2) to investigate the best dimensionality reduction method among PCA, KPCA and ICA for hyperspectral data.

## 2. METHOD

### 2.1. Study Area and Measurement

Manchurian wild rice (*Zizania latifolia*) plants were cultivated at within-row distances and inter-row spacing of 100 cm on a paddy field at Shizuoka University (Shizuoka, Japan, Figure 1) and grown in flooded conditions. As a basal fertilization, 18 kg of $NH_4Cl$, 12 kg of $P_2O_5$ and 12 kg of $K_2O$ were supplied per 1000 $m^2$. Two further supplementary fertilizations were administered, consisting of 12 kg of $NH_4Cl$, 12 kg of $P_2O_5$ and 12 kg of $K_2O$, and 6 kg of $NH_4Cl$, respectively (per 1000 $m^2$). The soluble silicic acid content of the provided molten slag was 32% and the standard amount of slag fertilizer was 120 kg per 1000 $m^2$. The experiment included a control without slag and four slag fertilizer treatments: a standard amount of slag (1×Slag), and double (2×Slag), 4 times (4×Slag) and 8 times (8×Slag) the standard concentration. A total of 200 leaves (40 leaves from each treatment) were measured for reflectance with 3 times reputation on each leaf samples and chlorophyll content on 2 and 5 October, 2020.

Hyperspectral reflectance was obtained using the FieldSpec4 (Malvern Panalytical, Almelo, Netherlands) and then a splice correction function was applied to minimize the inconsistency caused by the three detectors using ViewSpec Pro (Analytical Spectral Devices Inc., USA).

Dimethyl-formamide was used the prepare extracts and their chlorophyll contents were quantified using a dual beam scanning ultraviolet-visible spectrometer (UV-1900, Shimadzu, Japan) and Porra's method (Porra, et. al. 1989) and the below equations (1 to 3) were used to calculate chlorophyll–a (Chl–a) and b (Chl–b) content (in μg $ml^{-1}$). Finally, units were converted to μg $cm^{-2}$ using the area of leaf discs, since leaf optical properties are sensitive to chemistry in terms of quantity per surface area.

$$\text{Chl–a } (\mu g\ ml^{-1}) = 12.00 \times (A_{663.8} - A_{750}) - 3.11 \times (A_{646.8} - A_{750}) \quad (1)$$

$$\text{Chl–b } (\mu g\ ml^{-1}) = 20.78 \times (A_{646.8} - A_{750}) - 4.88 \times (A_{663.8} - A_{750}) \quad (2)$$

$$\text{Chla+b} = \text{Chl–a} + \text{Chl–b} \quad (3)$$

where $A$ is the absorbance, and the subscripts are the wavelengths (in nm).


**Figure 1.** *Zizania latifolia* and location of each treatment

### 2.2. Data Analysis

Performance evaluation was conducted for RF regression and all processes were implemented using R version 3.5.3 (R Core Team 2020). RF regression creates multiple decisions tees called classification and regression trees (CART) based on randomly bootstrapped samples of training data (Breiman 2001) via generalization of the binomial variance (using a Gini index) and by nodes that are using by split variable from a group of randomly selected variable (Liaw and Wiener 2002). Since former research has described the effectiveness of RF (Hobley et al. 2018, Johansson et al. 2014), it was also used in this research. RF differs from CART in growing non-deterministically to decorrelate the trees and lessen variance using two-stage randomization scheme related to a bootstrap sample and random variable selection. The number of trees (ntree) and the number of variables used to split the nodes (mtry) are normally established by the user. For tuning these hyperparameters, Bayesian optimization was applied using the Gaussian process (Bergstra and Bengio 2012).

### 2.3 Dimension Reduction Techniques

RF-Based Regression models were generated after dimension reduction techniques including Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Independent Component Analysis (ICA). In this study, 10 components were applied for each technique.

#### 2.3.1 Principal Component Analysis (PCA)

PCA is the oldest and best-known technique of multivariate data analysis (Mishra et al. 2017). It was first coined by (Pearson 1901), and produced independently by (Hotelling 1933). PCA is the usual name for a technique which uses sophisticated underlying mathematical principles to transforms several probably correlated variables into smaller number of variables named principal components. The origin PCA lies in multivariate data analysis; however, it has a wide range of other applications. In general terms, PCA uses a vector space transform to reduce the dimensionality of large data sets. Using mathematical forecast, the original dataset, which may involve many variables (i.e., the principal component). The central idea of PCA is to reduce the dimensionality of the data set in which there are many interrelated variables. Recently, hyperspectral data set contain highly correlated spectral bands and often having redundant information. Some of study (Bioucas-Dias et al. 2013, Varshney and Arora 2004) have been using PCA to extract the band dependency or correlation through statistical properties (Gonzales and Woods 2008). Practical implementation this algorithm was according on (Rodarmel and Shan 2002).

$$X_i = [x_1, x_2, x_3, \ldots\ldots, x_N]^T_i, \quad (4)$$

$$m = \frac{1}{M}\sum_{i=1}^{M}[x_1, x_2, x_3, \ldots\ldots, x_N]^T_i, \quad (5)$$

In Eq. 4, $X_i$ represents a vector pixel for all the dimensions at a specific pixel location $i$. Similarly, $X_i$ represents each pixel in the respective $i^{th}$ dimension and the total number of dimensions are *N*. Total number of pixel vectors depend on the size of each band $M = m * n$ where $m$ is the number of rows and $n$ is the number of columns. The PCA depends on the eigen value decomposition of the covariance matrix and the equation 7.

$$C_x = ADA^T \quad (6)$$

$$Where\ D = diag(\lambda_1, \lambda_2, \lambda_3, \ldots\ldots, \lambda_N) \quad (7)$$

Is the diagonal matrix of eigenvalues $\lambda_1$, $i$ = 1, 2, 3,....$N$ of the covariance matrix $C_x$. A is the orthonormal matrix with eigenvectors $a_k$ (k = 1, 2,....,$N$) of $C_x$. Each PCA pixel vector after the transformation is defined by:

$$y_I = A^T x_I (I = 1, 2, 3, \ldots, M) \quad (8)$$

Each of this transformed pixel vectors contain compressed information of the entire data set and only a first few bands contribute to useful information. And then, the number of vectors to be chosen for each data set varies with size and dimensions of the data set (Mallapragada, et al. 2018).

#### 2.3.2 Kernel Principal Component Analysis (KPCA)

PCA only allows linear dimensionality reduction and then cannot be well represented in a linear subspace if the data has more complicated structures. Kernel PCA is the nonlinear form of PCA, which better exploits the complicated spatial structure of high-dimensional features (Benhart 1997). The Radial Basis kernel function kernel, which is the typical general-purpose kernel, was applied and the kernel bandwidth was set to 0.1. As tell before, PCA just preform linear transformations, and therefore cannot handle the complex nonlinear features that are widely present in hyperspectral data (Zhang et al. 2019). Therefore, KPCA is hypothesized to improve the performance of dimensionality reduction in this study.

### 2.3.3 Independent Component Analysis (ICA)

Independent component analysis (ICA) is closely related to PCA, whereas ICA finds a set of source variable that are mutually independent, PCA finds a set variable that are mutually uncorrelated (Naik and Dinesh 2011, Shlens 2014). The independent component analysis technique is one of the most well-known algorithms which are used for solving this research. ICA is separating multivariate signal into additive subcomponent. This is done by assuming that the subcomponent is non-Gaussian signal and that they are statistically independent from each order. The goal of ICA is to create maximally independent components for an given data (Hyvrinen and Oja 2000, Richards 2013) and this methods is according in the blind source separation. ICA represent of the images based on (Bell and Sejnowski 1997).

$$x = As \tag{9}$$

where $x_i = \sum_{j=1}^{N} a_{ij}\, s_j$ is a pixel in an image, A is a basic function matrix and s image source? $x_i$ is represented as a linear combination of each image source $s_j$ and weight coefficient $a_{ij}$. $N$ is the number of image sources in the data set. All pixels each image source as linearly transformed with a matrix of filters $W$, so that the resulting vector:

$$u = Wx, \tag{10}$$

recovers the underlying causes, s, in a different order and re-scaled. It shows partial hindrance to process data with noise and becomes computationally expensive with hyperspectral data.

### 2.4 Statistical Criteria

To evaluate the performance of the regression model, the root-mean-square error (RMSE, equation (8)) was applied. RMSE is the square root of the mean of the square of all of the error. The use of RMSE is very common, and it is considered an excellent general-purpose error metric for numerical predictions (Neil and Hashemi 2018).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

Where n is number of samples, $y_i$ is measured chlorophyll content and $\hat{y}_i$ is estimated chlorophyll content. RMSE is a good measure of accuracy, but only to compare prediction errors of different models or model configurations for a particular variable and not between variable, as it is scale-dependent.

## 3. RESULTS

### 3.1 Chlorophyll Content

The measured chlorophyll content per leaf area ($cm^2$) ranged from 17.53 to 58.02 µg and the maximum value were obtained from the 2 × Slag treatment while the minimum values were from the control. Although there were significant differences in chlorophyll content between 2×Slag and other treatments such as control, 1× slag, 4 × slag and 8 × slag ($p < 0.05$, Tukey-Kramer test), the other combinations did not differ significantly.

### 3.2 Spectral Reflectance and Correlation

The mean reflectance spectra and pre-processed spectra for each slag fertilizer concentration are shown in Figure 2. Contrary to the relationships between treatments and Chla+b, the control samples showed the highest reflectance values while the lowest values were from the 2×Slag samples. The reflectance values near the green peak increased when plants were fertilized with more than the 2×Slag treatment, but decreased at lower treatments.

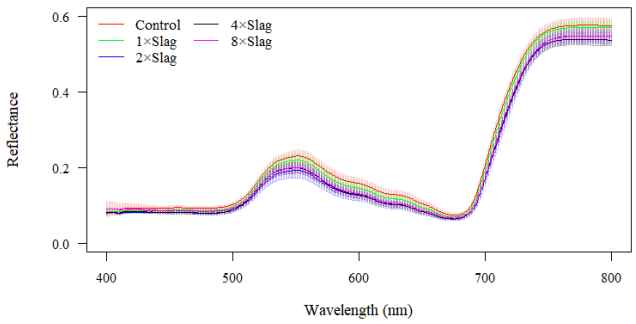



**Figure 2.** Spectral patterns for each slag fertilizer treatment

Significant negative correlations were confirmed between Chlorohyll content and reflectance values near 525 nm and 700 nm (Figure 3), the line of spectral reflectance at this number wavelength shown get around -0.6. Although the correlation coefficients were generally weakest for the control, the relationships between wavelength and the correlation coefficient were similar among the four treatments. Stepwise discriminant analysis ($p < 0.05$) showed that reflectance values at 20 wavelengths (547, 553, 557, 564, 566, 567, 570, 575, 583, 620, 631, 671, 681, 683, 696, 711, 714, 720, 766 and 771 nm) were useful for identifying samples at the five different slag fertilizer treatments. An overall accuracy of 95.5 % was achieved.

### 3.3 Accuracy Validation

Table 1 shows statistics for the RMSE values calculated using random forest machine learning regression models. Generally, PCA generally performed the best and PCA was selected as the best solution for estimating chlorophyll content 50 times, while KPCA was selected 12 times. Thus, it is not necessary to use kernel for expressing the relationships between chlorophyll

content of *Zizania latifolia* and reflectance data from FielSpec4.

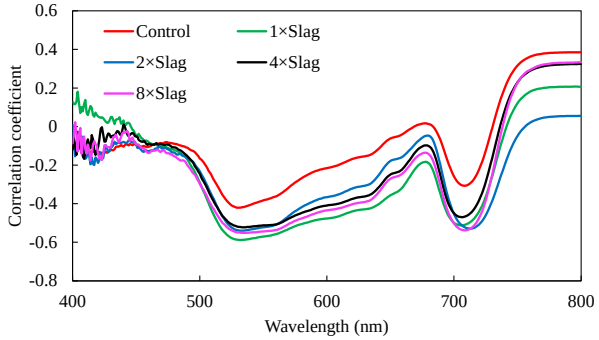



**Figure 3.** Correlation between spectral reflectance and chlorophyll content

**Table 1.** Root-mean-square error (RMSE, μg cm$^{-2}$) for each regression model after 100 repetitions.

| | PCA | KPCA | ICA |
|---|---|---|---|
| Minimum | 4.49 | 4.90 | 4.36 |
| Median | 5.65 | 5.86 | 5.77 |
| Mean | 5.65 | 5.94 | 5.76 |
| Maximum | 7.33 | 7.54 | 7.37 |
| Standard deviation | 0.58 | 0.56 | 0.59 |

Figure 4 and 5 show the relationships between measured and estimated chlorophyll contents when the results of 100 repetitions were combined. The coefficient of determinations ($R^2$) were 0.480 0.422 and 0.477 for PCA, KPCA and ICA, respectively and then the advantage of PCA was confirmed. However, the differences were too small to claim that any techniques should be applied.

## 4. Discussion

After 100 repetitions, the best and worst algorithms of each round were determined based on the RMSE value (Table 1). Although KPCA was selected 12 times as the best dimension reduction techniques, it possessed the worst results 54 times per 100 repetitions and then the advantages of using kernel was not confirmed in this study. (Lin et al. 2020) reported that the machine learning models combined with KPCA were effective for estimating soil composition contents, however, the advantages of KPCA were not confirmed for chlorophyll content estimation from *Zizania Latifolia*. The clear relationships between chlorophyll content and reflectance over green peak and rededge have been used for estimating chlorophyll content from reflectance (Sims and Gamon 2002) and then it was not effective to express their relationships using non-liner models.

PCA gave the best estimation results based on the RMSE value after 100 repetitions (Table 1) and it provides the highest accuracies 50 times while it was selected as worst 13 times. The high performances of ICA have been confirmed some studies using hyperspectral images and it was effective for separating two different combinations of illumination and reflectance components (Ahmad et al. 2017). In this study, the reflectance measurements were conducted the portable spectrometer (i.e., FieldSpec4) with leaf clip and then the necessity of separating illumination and reflectance components was quite small and then the advantage of ICA might be vague. Indeed, the high chlorophyll content estimation accuracies have been reported in some studies based on PCA and the reflectance from similar spectrometers (Liu et al. 2018, Saputro et al. 2018) and the results from our study also supported their results.

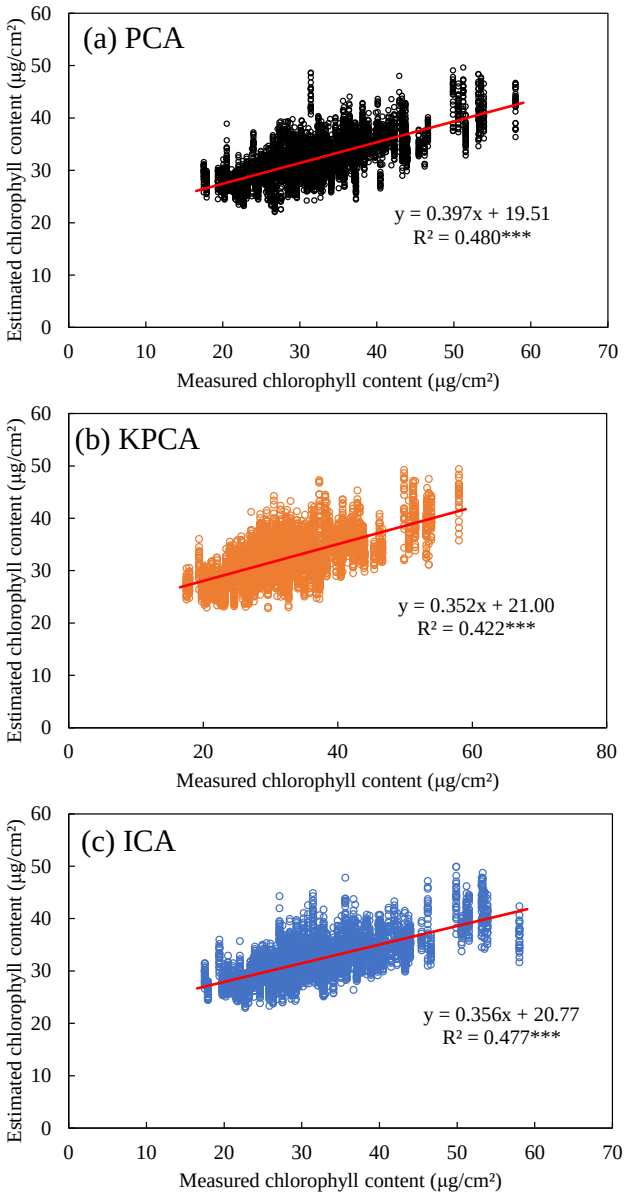



**Figure 4.** Relationship between estimation and measured chlorophyll contents

**Table 2.** Best- and worst-performing techniques after 100 repetitions. Results presented are number of times per 100 repetitions.

| Dimension Reduction Techniques | Selected times | | Net score (Best-Worst) |
|---|---|---|---|
| | Best | Worst | |
| PCA | 50 | 13 | 37 |
| KPCA | 12 | 54 | -42 |
| ICA | 38 | 33 | 5 |

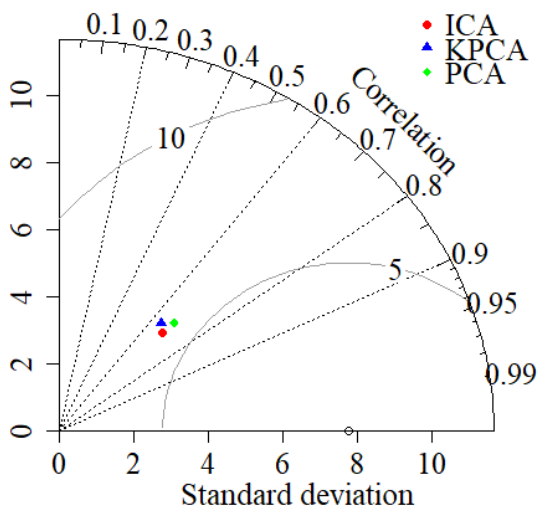



**Figure 5.** Taylor diagram showing the performance of each dimensionality reduction methods

## 5. CONCLUSION

This study has evaluated hyperspectral data and three-dimension reduction techniques for estimating chlorophyll contents from reflectance. According of result, hyperspectral data has useful to estimation chlorophyll contents and spectral reflectance had shown control sample are highest reflectance values more than another slag fertilizer treatment. However, based on the machine learning algorithm applied for regression, resulted a combination of Random Forest (RF) and Principal Component Analysis (PCA) is the best method for dimensionality reduction for the estimation chlorophyll compared another advanced method such as Kernel Principal Component Analysis (KPCA) and Independent Component Analysis (ICA). For future study, it will be better to add the other machine learning methods to estimate chlorophyll contents to increase the quantitively of model.

## ACKNOWLEDMENT

We thank the members of the Laboratory of Plant Functional Physiology and the Laboratory of Macroecology, Shizuoka University, for their support during both fieldwork and laboratory analyses.

## Author contributions

**Adenan Yandra Nofrizal:** Writing-Original draft preparation, Methodology, Software. **Rei Sonobe**: Investigation, Data Curation, Writing-Reviewing and Editing. **Hiroto Yamashita, Akio Morita, Takashi Ikka:** Conceptualization

## Conflicts of interest

The authors declare no conflicts of interest.

## REFERENCES

Ahmad F A, Khan I U, Islam M, Uzair & Ullah H (2017). EIllumination normalization using independent component analysis and filtering. *Imaging Science Journal,* 65**,** 308-313.

Bell A J & Sejnowski T J (1997). The ''independent components'' of natural scenes are edge filters. *Vision Research,* 37**,** 3327-3338.

Benhart S (1997). Kernel principal component analysis. In *Artificial Neural Networks-ICAN'97*, 583-588.

Bergstra J & Bengio Y (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research,* 13**,** 281-305.

Biau G & Scornet E (2016). A random forest guided tour. *Test,* 25**,** 197-227.

Bioucas-Dias J M, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N M & Chanussot J (2013) Hyperspectral Remote Sensing Data Analysis and Future Challenges. *Ieee Geoscience and Remote Sensing Magazine,* 1**,** 6-36.

Blackburn G A & Steele C M (1999). Towards the remote sensing of matorral vegetation physiology: Relationships between spectral reflectance, pigment, and biophysical characteristics of semiarid bushland canopies. *Remote Sensing of Environment,* 70**,** 278-292.

Breiman L (2001). Random forests. *Machine Learning,* 45**,** 5-32.

Breunig F M, Galvao L S, Dalagnol R, Dauve C E, Parraga A, Santi A L, Della Flora D P & Chen S S (2020). Delineation of management zones in agricultural fields using cover crop biomass estimates from PlanetScope data. *International Journal of Applied Earth Observation and Geoinformation,* 85.

Chen D, Meng Z W & Chen Y P (2019). Toxicity assessment of molybdenum slag as a mineral fertilizer: A case study with pakchoi (Brassica chinensis L.). *Chemosphere,* 217**,** 816-824.

Cui L, Jiao Z T, Dong Y D, Sun M, Zhang X N, Yin S Y, Ding A X, Chang Y X, Guo J & Xie R (2019). Estimating Forest Canopy Height Using MODIS BRDF Data Emphasizing Typical-Angle Reflectances. *Remote Sensing,* 11**,** 21.

Duda R, Hart O P E & Stork D G. (2001). *Pattern Classification*. Wiley-Interscience.

Fan S X, Zhang B H, Li J B, Huang W Q & Wang C P (2016). Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. *Biosystems Engineering,* 143**,** 9-19.

Fernandez-Delgado M, Sirsat M S, Cernadas E, Alawadi S, Barro S & Febrero-Bande M (2019). An extensive experimental survey of regression methods. *Neural Networks,* 111**,** 11-34.

Féret J-B, Francois C, Asner G P, Gitelson A A, Martin R E, Bidel L P R, Ustin S L, G. le Maire & Jacquemoud S (2008). PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment,* 112**,** 3030-3043.

Gonzales C G & Woods R E (2008). *Digital image processing: third edition*. Pearson Prentice Hall.

Gu C Y, Du H Q, Mao F J, Han N, Zhou G M, Xu X J, Sun S B & Gao G L (2016). Global sensitivity analysis of PROSAIL model parameters when simulating Moso bamboo forest canopy reflectance. *International Journal of Remote Sensing,* 37**,** 5270-5286.

Hernandez-Clemente R, Navarro-Cerrillo R M & Zarco-Tejada P J (2014). Deriving Predictive Relationships of Carotenoid Content at the Canopy Level in a Conifer Forest Using Hyperspectral Imagery and Model Simulation. *IEEE Transactions on Geoscience and Remote Sensing,* 52**,** 5206-5217.

Hobley E, Steffens M, Bauke S L & Kogel-Knabner I (2018). Hotspots of soil organic carbon storage revealed by laboratory hyperspectral imaging. *Scientific Reports,* 8.

Hotelling H (1933). Analysis of a complex of statistical variables into principal components. *Journal of Education Psychology,* 24**,** 417-441.

Huang J, Sun Y H, Wang M Y, Zhang D D, Sada R & Li M C (2017). Juvenile tree classification based on hyperspectral image acquired from an unmanned aerial vehicle. *International Journal of Remote Sensing,* 38**,** 2273-2295.

Huang Y B, Lee M A, Thomson S J & Reddy K N (2016). Ground-based hyperspectral remote sensing for weed management in crop production. *International Journal of Agricultural and Biological Engineering,* 9**,** 98-109.

Hunt E R, Daughtry C S T & Li L (2016). Feasibility of estimating leaf water content using spectral indices from WorldView-3 ' s near-infrared and shortwave infrared bands. *International Journal of Remote Sensing,* 37**,** 388-402.

Hyvrinen A & Oja E (2000). Independent component analysis: algorithms and applications. *Neural nertworks,* 13**,** 411-430.

Jacquemoud S & Baret F (1990). PROSPECT: A Model of leaf optical properties spectra. *Remote Sensing of Environment,* 34**,** 75-91.

Johansson U, Bostrom H, Lofstrom T & Linusson H (2014). Regression conformal prediction with random forests. *Machine Learning,* 97**,** 155-176.

Kalaji H M, Dabrowski P, Cetner M D, Samborska I A, Lukasik I, Brestic M, Zivcak M, Tomasz H, Mojski J, Kociel H & Panchal B M (2017). A comparison between different chlorophyll content meters under nutrient deficiency conditions. *Journal of Plant Nutrition,* 40**,** 1024-1034.

Kindomihou V, Sinsin B & Meerts P (2006). Effect of defoliation on silica accumulation in five tropical fodder grass species in Benin. *Belgian Journal of Botany,* 139**,** 87-102.

Krawiec G, Czaplicka M & Czernecki J (2017). Physicochemical properties of slags produced at various amounts of iron addition in lead smelting. *Journal of Material Cycles and Waste Management,* 19**,** 959-967.

Lausch A, Pause M, Schmidt A, Salbach C, Gwillym-Margianto S & Merbach I (2013). Temporal hyperspectral monitoring of chlorophyll, LAI, and water content of barley during a growing season. *Canadian Journal of Remote Sensing,* 39**,** 191-207.

Li L M, Zhao J, Wang C R & Yan C J (2020). Comprehensive evaluation of robotic global performance based on modified principal component analysis. *International Journal of Advanced Robotic Systems,* 17**,** 11.

Liaw A & Wiener M (2002). Classification and Regression by random forest. *R news,* 2**,** 18-22.

Lin N, Liu H Q, Yang J J & Liu H L (2020). Hyperspectral estimation of soil composition contents based on kernel principal component analysis and machine learning model. *Journal of Applied Remote Sensing,* 14.

Liu H J, Li M Z, Zhang J Y, Sun H, Long Y W, Wu L & Zhang Q (2018). PCA based model on chlorophyll content diagnosis of winter wheat. *Ifac Papersonline,* 51**,** 643-647.

Mallapragada S, Wong M & Hung C C (2018). Dimensionality reduction of Hyperspectral Images for Classification. In *Ninth International Conference on Information*, 1-7. Tokyo, Japan.

Mishra S P, Sarkar U, Taraphder S, Datta S, Swain D P, Saikhom R, Panda S & Laishram M (2017). Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research,* 7**,** 60-78.

Mutanga O & Skidmore A K (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing,* 25**,** 3999-4014.

Naik N R & Dinesh K K (2011). An Overview of Independent Component Analysis and Its Applications. *Informatica,* 53**,** 63-81.

Neil S P & Hashemi M R (2018). *Fundamentals of Ocean Renewable Energy; Generating electricity from the sea*. Elsevier.

Ormeci C, Sertel E & Sarikaya O (2009). Determination of chlorophyll-a amount in Golden Horn, Istanbul, Turkey using IKONOS and in situ data. *Environmental Monitoring and Assessment,* 155**,** 83-90.

Pearson K (1901) On lines and planes of closet fit to system of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journals of Science,* 2**,** 559-572.

Porra R J, Thompson W A & Kriedemann P E (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 975**,** 384-394.

Prado-Cabrero A, Beatty S, Howard A, Stack J, Bettin P & Nolan J M (2016). Assessment of lutein, zeaxanthin and meso-zeaxanthin concentrations in dietary supplements by chiral high-performance liquid chromatography. *European Food Research and Technology,* 242**,** 599-608.

Richards J A (2013). *Remote Sensing Digital Image Analysis fitfh edition*. Springer.

Rodarmel C & Shan J (2002). Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science,* 62**,** 115-122.

Romero A, Aguado I & Yebra M (2012). Estimation of dry matter content in leaves using normalized indexes

and PROSPECT model inversion. *International Journal of Remote Sensing,* 33**,** 396-414.

Rutledge D N (2018). Comparison of Principal Components Analysis, Independent Components Analysis and Common Components Analysis. *Journal of Analysis and Testing,* 2**,** 235-248.

Saputro A H, Juansyah S D, Handayani W & Ieee (2018). Banana (Musa sp.) Maturity Prediction System based on Chlorophyll Content using Visible-NIR Imaging. In *International Conference on Signals and Systems (ICSigSys)*, 64-68. Bali, INDONESIA: Ieee.

Shlens J (2014). A tutorial on principal component analysis. *arXiv***,** arXiv:1404.1100.

Sims D A & Gamon J A (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment,* 81**,** 337-354.

Sonobe R, Hirono Y & Oi A (2020). Non-Destructive Detection of Tea Leaf Chlorophyll Content Using Hyperspectral Reflectance and Machine Learning Algorithms. *Plants,* 9.

Sonobe R, Miura Y, Sano T & Horie H (2018). Monitoring Photosynthetic Pigments of Shade-Grown Tea from Hyperspectral Reflectance. *Canadian Journal of Remote Sensing,* 44**,** 104-112.

Sonobe R & Wang Q (2017a). Hyperspectral indices for quantifying leaf chlorophyll concentrations performed differently with different leaf types in deciduous forests. *Ecological Informatics,* 37**,** 1-9.

Sonobe R & Wang Q (2017b). Towards a universal hyperspectral index to assess chlorophyll content in deciduous forests. *Remote Sensing,* 9**,** 191.

Sonobe R, Yamashit H, Nofrizal A Y, Seki H, Morita A & Takashi I (2021). Use of spectral reflectance from a compact spectrometer to assess chlorophyll content in Zizania latifolia. *Geocarto International***,** 1914747.

Swiniarski R W & Skowron A (2004). Independent component analysis, principal component analysis and rough sets in face recognition. In *Transactions on Rough Sets I,* eds. J. F. Peters, A. Skowron, J. W. GrzymalaBusse, B. Kostek, R. W. Swiniarski & M. S. Szczuka, 392-404.

Varshney P K & Arora M K (2004). *Advanced image processing techniques for remotely sensed hyperspectral data*. Springer.

Wang H L, Zhang X K, Jin B S, Wang M D, Chen W X & Liu D (2020). WATER LEVEL FLUCTUATIONS DETERMINE THE SPATIAL AND TEMPORAL DISTRIBUTION OF MANCHURIAN WILD RICE (ZIZANIA LATIFOLIA) IN SIX YANGTZE RIVER FLOODPLAIN LAKES, CHINA. *Applied Ecology and Environmental Research,* 18**,** 5491-5503.

Yan N, Du Y M, Liu X M, Chu C, Shi J, Zhang H B, Liu Y H & Zhang Z F (2018). Morphological Characteristics, Nutrients, and Bioactive Compounds of Zizania latifolia, and Health Benefits of Its Seeds. *Molecules,* 23.

Zhang H, Duan Z, Li Y Y, Zhao G Y, Zhu S M, Fu W, Peng T, Zhao Q Z, Svanberg S & Hu J D (2019). Vis/NIR reflectance spectroscopy for hybrid rice variety identification and chlorophyll content evaluation for different nitrogen fertilizer levels. *Royal Society Open Science,* 6.

Zhang K W, Hu B X, Wang J G, Pattey E & Smith A M (2011). Improving the retrieval of the biophysical parameters of vegetation canopies using the contribution index. *Canadian Journal of Remote Sensing,* 37**,** 643-652.

Zolotova E S, Ivanova N S, Ryabinin V F, Ayan S & Kotelnikova A L (2017). Kotelnikova Element mobility from the copper smelting slag recycling waste into forest soils of the taiga in Middle Urals. *Environmental Science and Pollution Research*.

Zou X C, Hernandez-Clemente R, Tammeorg P, Torres C L, Stoddard F L, Makela P, Pellikka P & Mottus M (2015). Retrieval of leaf chlorophyll content in field crops using narrow-band indices: effects of leaf area index and leaf mean tilt angle. *International Journal of Remote Sensing,* 36**,** 6031-6055.