

Article



Estimations of Nitrogen Concentration in Sugarcane Using Hyperspectral Imagery

Poonsak Miphokasap ^{1,*} and Wasinee Wannasiri ²

- ¹ Department of Geography, Faculty of Social Science, Kasetsart University, 50 Ngam Wong Wan Rd, Ladyaow, Chatuchak, Bangkok 10900, Thailand
- ² Geography Field of Study, Faculty of Liberal Arts, Thammasat University, 99 Moo 18, Klong Luang, Rangsit, Pathumthani 12121, Thailand; wasinee1@tu.ac.th
- * Correspondence: fsocpsm@ku.ac.th; Tel.: +66-866-216-243

Received: 5 March 2018; Accepted: 17 April 2018; Published: 20 April 2018



Abstract: This study aims to estimate the spatial variation of sugarcane Canopy Nitrogen Concentration (CNC) using spectral data, which were measured from a spaceborne hyperspectral image. Stepwise Multiple Linear Regression (SMLR) and Support Vector Regression (SVR) were applied to calibrate and validate the CNC estimation models. The raw spectral reflectance was transformed into a First-Derivative Spectrum (FDS) and absorption features to remove the spectral noise and finally used as input variables. The results indicate that the estimation models developed by non-linear SVR based Radial Basis Function (RBF) kernel yield the higher correlation coefficient with CNC compared with the models computed by SMLR. The best model shows the coefficient of determination value of 0.78 and Root Mean Square Error (*RMSE*) value of 0.035% nitrogen. The narrow sensitive spectral wavelengths for quantifying nitrogen content in the combined cultivar environments existed mainly in the electromagnetic spectrum of the visible-red, longer portion of red edge, shortwave infrared regions and far-near infrared. The most important conclusion from this experiment is that spectral signals from the space hyperspectral data contain the meaningful information for quantifying sugarcane CNC across larger geographic areas. The nutrient deficient areas could be corrected by applying suitable farm management.

Keywords: hyperspectral; nitrogen concentration; sugarcane; support vector regression

1. Introduction

Sugarcane (*Saccharum* spp. hybrid) is a tall-growing perennial crop widely cultivated in the tropical and subtropical regions of the world [1]. In Thailand, sugarcane is mostly used in the sugar and ethanol energy industries. The precise estimation of sugarcane productivity allows the establishment of deliberate policies to balance the amount of sugarcane within sugar and ethanol energy. Nitrogen is considered a key indicator of the physiological susceptibility of water availability, pests, disease and crop nutrient stress, which could potentially affect crop productivity [2]. In addition, nitrogen is one of the primary regulators of several leaf physiological processes, such as photosynthesis, respiration, and transpiration [3]. Insufficient nitrogen application leads to lower yields, lower sugarcane quality, and an increase of plant vulnerability to certain pests [4]. An increased susceptibility to pests or leaf blades turns a sugarcane leaf from light green to yellow, short, slender stalks and low productivity. To measure the nitrogen concentration using the conventional method, a great number of leaf samples from fields are needed. The process of leaf samples is laborious, lengthy and destructive [4]. Thus, the conventional method is not appropriate for large areas.

The development of hyperspectral remote sensing or imaging spectroscopy has offered possibilities for estimating and mapping the spatial distribution of biophysical and biochemical parameters [5]. Hyperspectral data offer an efficient and cost-effective solution for estimating leaf biochemical content compared to the traditional destructive sampling methods. This is mainly because the hyperspectral data provides a high number of bands less than 10 nm in width. Hyperspectral data leads to a continuous spectral profile through visible, near-infrared (NIR) and shortwave infrared (SWIR) regions [6,7]. There are a few applications of hyperspectral remote sensing for studying sugarcane physiology and chemistry [8–14]. Galvao [10] and Apan [11] attempted to discriminate sugarcane varieties using narrow spectral indices from EO-1 Hyperion. Abdel-Rahman [4] and Miphokasap [12] conducted an experiment to apply spectral indices calculated from in situ spectroscopy and regression techniques to estimate sugarcane nitrogen concentrations at the leaf and canopy levels, respectively, using in situ spectroscopy. Lebourgeois [13] used ultra-light aircraft equipped with broadband imaging sensors based on commercial digital cameras to monitor nitrogen content in the experimental sugarcane plots under the water and fertilizer supply controls. However, there are some factors that directly affect the variation of nitrogen concentration in the sugarcane canopy including the amount of fertilizer supply, environmental factors and susceptibility to pests, water stress and crop disease [15].

In the literature, there are several methods used to estimate crop nitrogen from a spectral signature [14]. Firstly, the univariate approach is the simplest way to analyze data using only one dependent and independent variable. The vegetation index, which is the computation of a narrow band vegetation index from reflectance data to draw plant characteristics, is widely applied in agricultural fields [4,8,12,14,16,17]. Partial least squares (PLS) regression has become a commonly used method to relate spectral data to biochemical concentration in plant canopies. Li [18] applied PLS and narrow band spectral indices to estimate the canopy nitrogen content of winter wheat. Wang [19] compared the performance of narrow vegetation indices from airborne hyperspectral images and PLS to estimate foliar nitrogen concentration in a mixed temperate forest. Gokkaya [20] used PLS to compute the correlation of the macronutrients of a mixed wood boreal forest with spaceborne imaging spectroscopy and LiDAR data at the canopy level. In addition, Stepwise Multiple Linear Regression (SMLR), which is one of the multivariate approaches, is a well-known technique that is applied to the estimation of biochemical content in several crop types [12,21–23].

Both univariate and multivariate approaches are based on the assumption of a linear relationship between variables. Such a relationship between the vegetation properties and spectral data was linearly assumed, which might not always be the case [22,24]. The complexity of spatial and temporal characteristics of spectral characteristics derived from the fields influences the variance of quality and quantity in sugarcane and might not be explained by a linear model. A non-linear relationship usually gives more flexibility than the simple linear model and produces better estimation results [25]. Two well-known non-linear approaches applied to the explanation of the variances in crop properties are Artificial Neural Network (ANN) [26-28] and Support Vector Regression (SVR) [29,30]. In recent years, SVR is gaining popularity over ANN due to its many attractive features and its promising efficient performance [30]. In the case of SVR, the input data are first transformed by non-linear kernel and the linear model is constructed in the high dimensional feature space. When hyperspectral data are input into the SVR, two problems usually occur, including how to select the optimal feature subset from the hundreds of wavelengths, and how to define the best kernel types and parameters [31,32]. Only few research studies have focused on the applications of SVR for depicting biophysical parameters through hyperspectral data [33,34]. In addition, there are still no studies on the application of SVR for estimating nitrogen content in sugarcane from spaceborne hyperspectral data. With regard to sugarcane nutritional status, some research in the past has been undertaken to estimate foliar and canopy nitrogen concentration in the fields using a portable spectrometer, with promising results.

The use of spectral observation from spaceborne hyperspectral data allows scientists to monitor the nutrition status of sugarcane over large areas with a nondestructive method. With the appropriate scale and timing, hyperspectral data provide valuable information for the timely detection of nutrient deficiencies and precision nutrient management in sugarcane production before harvesting. The main objective of this study is to estimate the spatial variation of sugarcane Canopy Nitrogen Concentration (CNC) from an EO-1 Hyperion image in combined cultivar environments. The capability of SMLR- and SVR-based approaches for estimating CNC were compared. We also validated the performance of a transformed spectral dataset; First-Derivative Spectrum (FDS) and absorption features to correlate with nitrogen content. This study is an extension of a field spectra study, which showed the potential of FDS and SMLR techniques for predicting in situ sugarcane nitrogen at the canopy level [12]. The integration of the proposed methods with hyperspectral variables might be applicable to reliably mapping sugarcane nitrogen content in cropland.

2. Materials and Methods

2.1. Study Area

The study area is located in Sakaeo Province, in the eastern region of Thailand ($102^{\circ}15'$ E, $13^{\circ}45'$ N) as shown in Figure 1.



Figure 1. The study experimental plots in Sakaeo Province, the eastern region of Thailand.

The northern region of the study province is covered with the forested mountains in the Sankamphaeng Range and the Dangrek Mountains. The mean temperature ranges from 27.5 °C to 28.78 °C. The annual rainfall ranges from 1296 to 1539 mm. Sugarcane is the most dominant crop in this area with approximately ten different cultivars. Farmers' sugarcane plots in Wattananakorn, Wang Nam Yen and Klong Hard district, Sakaeo province are usually under the traditional treat conditions. The total study areas are of 40,230 ha. However, there are only four popular sugarcane cultivars; K84-200, LK-92-11, KK-3, and UT-8.

2.2. Research Methodology Overview

This research aims to compare the efficiency of two approaches of multivariate based statistical methods with SVR. Two approaches were applied to the estimation of CNC in sugarcane at the canopy level with combined cultivar environments. Spectral reflectance was measured from a hyperspectral satellite. The research methodologies are described below.

2.3. Data Acquisition

The single date EO-1 Hyperion of 128/50 path/row was acquired on 29 January 2012 at the peak of the regional winter season and before the sugarcane harvest. Two hundred and forty-two continuous bands, 7.5 km in width and 100 km in length, covering 400–2500 nm, were recorded. The scene center was at 13.67° N and 102.26° E, look angle was 3.59 degrees (almost nadir position) and there was no cloud cover.

2.3.1. Field Data Collection

The collections of sugarcane foliar samples were carried out on the mature sugarcane plots (9–10 months). The canopies were fully developed to avoid the effects of soil brightness. Field data collection was conducted from 4–6 February 2012, within one week of image acquisition. The image pixel size is 30 m ground resolution. Therefore, the sampling site was large enough to ensure the existence of pure pixels in the image. Half of the pixel space was extended in all directions from the considered pixel. Thus, ground measurements with the size of 60 m × 60 m (3600 m²) were randomly selected, concerning the sugarcane cultivars. A total of 70 samples were initially determined. The center coordinates of each plot were recorded. Representative samples from each plot were collected randomly by measuring sugarcane leaves. The second fully expanded leaf from the top of two random selected shoots of one tiller was collected. Sixteen random tillers from 16 subplots (15 m × 15 m) were measured. Thirty-two sampled leaves from one plot were mixed and used for analysis of the nitrogen content in the laboratory. Figure 2 shows the sampling design of field data collections.



Figure 2. Sampling design for field data collection.

2.3.2. Chemical Analysis of Foliar Samples

The second fully expanded leaves from the top of two random shoots of one tiller were collected. Thirty-two fresh leaves from one sampling site were oven-dried at 75 °C for 24 h and then ground up and oven-dried again at 75 °C for 24 h, resulting in leaf powder. The total nitrogen concentration of sugarcane foliar was measured using the Kjeldahl method. The leaf powder was digested by 98% sulfuric acid (w/w) at 380 °C until the solution was transparent. Nitrogen was measured with a Nitrogen Distillation Apparatus using KjeltecTM 2200 Auto distillation and would show the amount of nitrogen concentration in both milligrams per gram (mg/g) and a percentage of nitrogen.

2.4. Image Pre-Processing

2.4.1. Geometric Correction

Geometric corrections without the use of ground control points were performed in image level 1GST by USGS. Geo-TIFF is a raster image format that enables referencing by geodetic model or map projection. However, to achieve higher accuracy, the image must be re-rectified by using Ground Control Point (GCP) collected from the field. At least ten clearly visible GCPs were utilized to rectify the image using the first-order polynomial transformation. The *RMSE* for image rectification is less than 10 m overall.

2.4.2. Radiance Transformation

The digital number (DN) value was rescaled to get the radiance values using the Hyperion calibration coefficients. There are two different rescaling factors used for the VNIR and SWIR region of electromagnetic spectrum [35]. Radiance (L) in the unit of $W/m^2/\mu m/sr$ was computed as L = DN/40 and L = DN/80 for visible-near infrared (VNIR) and shortwave infrared (SWIR) bands, respectively.

2.4.3. Atmospheric Correction

Due to the effects of aerosol, water content, gases, digital number in remotely sensed data was considered as the reflectance at Top of Atmospheric (TOA). There are several packages for atmospheric correction widely used in hyperspectral imagery to depict the surface reflectance without atmospheric effects. Subsequently, the Radiance image was atmospherically corrected using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm built in ENVI. FLAASH uses the MODTRAN4 radiation transfer code to convert radiance at the top of the atmosphere to reflectance at the surface, pixel by pixel.

2.4.4. Band Noise Reduction

The Hyperion sensor consists of the continuous 242 bands covering 70 bands in the VNIR and 172 bands in the SWIR region of the electromagnetic spectrum. However, the analysis of Hyperion raw data illustrated that there were 44 bands that have a zero value. There are bands 1 to 7 (355.58–406.47 nm), bands 58 to 76 (902.36–935.58 nm) and bands 225 to 242 (2405.6–2577.07 nm). Thus, out of a total of 242 bands, 198 bands were selected to be calibrated and for further analysis. In addition, it was found that bands 77 and 78 in the SWIR region both have a higher level of spectral noise [36]. Therefore, there are only 196 bands that remain to be processed. Subsequently, atmospheric water vapor bands that absorbed almost all incidents and reflected solar radiation were easily identified by visual inspection of the image data and by atmospheric modeling [37]. Accepting this as a good criterion for band elimination for land surface applications produce the subsets of 176 bands listed in Table 1.

Table 1. Date of image acquisition, spatial and spectral resolution, and band subset characteristics used in this experiment.

Sensor	Acquisition Date	Spatial Resolution (m)	Spectral Resolution (nm)	Bands ¹	Wavelength (nm)
EO-1- Hyperion	29 January 2012 (winter)	30	10 (approx.)	8 to 57 79 to 120 128 to 166 179 to 223	426.8–925.4 932.6–1346.2 1426.9–1810.3 1941.5–2385.4

¹ Band 8 to band 57 cover the visible and near-infrared region; and band 79 to band 223 cover the shortwave-infrared region of electromagnetic spectrum.

2.5. Spectral Transformation

2.5.1. First-Derivative Spectrum: FDS

In this study, FDS was applied and used as a variable for implementing the estimation models. FDS is widely used to enhance absorption features that might be masked by interfering canopy background absorptions [37]. Derivative spectra can be a good noise filter since changes in the baseline have a negligible effect on derivatives. This technique is of benefit for reducing the effects of radiation scattering due to geometry and surface roughness [38]. FDS was calculated using Equation (1):

$$FDS = (R_{\lambda(i+1)} - R_{\lambda(i)}) / \Delta\lambda \tag{1}$$

where FDS is the first-derivative transformation at a specific wavelength *i* midpoint between wavebands *j* and *j* + 1. $R_{\lambda(j)}$ is the reflectance at the *j* waveband, $R_{\lambda(j+1)}$ is the reflectance at the *j* + 1 waveband and $\Delta\lambda$ is the difference in wavelength between *j* and *j* + 1.

2.5.2. Absorption Features

The features in the visible, near-infrared and shortwave-infrared regions are widely utilized to elaborate the amount of nitrogen concentration [21]. The continuum-removed reflectance, or $R'_{(\lambda i)}$, is derived by dividing the reflectance value $R_{(\lambda i)}$ of each waveband *i* in the absorption features by the reflectance level of the continuum line (convex hull) $R_{c(\lambda i)}$ at the corresponding wavelength *i*, explained in Equation (2):

$$R'_{(\lambda)} = R_{(\lambda i)} / R_{c(\lambda i)} \tag{2}$$

The first and last values of the continuum-removed spectrum are equal to 1. The FDS output ranges between 0 and 1, where the absorption pits are enhanced. In this study, four specific regions of the wavelength were focused, including R_{426} – R_{528} , R_{538} – R_{752} , R_{1104} – R_{1285} , and R_{1477} – R_{1790} , which are determined as the pigment and water content absorption zones. Two variables, examined in References [22,24] were utilized as the input variables in this study.

(i) Continuum-Removed Derivative Reflectance (CRDR) was derived using a first-derivative transformation to the continuum-removed reflectance spectrum R' as described in Equation (3).

$$CRDR = (R'_{\lambda(j+1)} - R'_{\lambda(j)})/\Delta\lambda$$
(3)

(ii) Band depth (BD) was derived by omitting the continuum-removed reflectance at specific wavelength i from 1 shown as in Equation (4)

$$BD_{(\lambda i)} = 1 - R'_{\lambda i} \tag{4}$$

2.6. Feature Dimensionality Reduction

Kernel Principal Component Analysis (KPCA) is an extension of principal component analysis (PCA) based on kernel methods. KPCA allows users to compute the principal components in a high dimensional feature space. This method is non-linearly related to the input variables.

Given a set of centered input vectors $x_t = (t = 1,..l)$, and $\sum_{t=1}^{l} x_t = 0$ each of which is of m dimension $x_t = (x_t(1), x_t(2), ..., x_t(m))^T$ (usually m < 1), the key idea of KPCA is to map the original input vector x_t into a higher dimensional feature space $\phi(x_t)$ and to calculate the linear PCA in $\phi(x_t)$. Dimension is assumed to be larger than l when mapping x_t into $\phi(x_t)$. The eigenvalue problem was solved by KPCA and illustrated in Equation (5).

$$\lambda_i \mu_i = \overline{C} \mu_i, i = 1, \dots, l \tag{5}$$

where $\overline{C} = \frac{1}{I} \sum_{i=1}^{I} J_{\varphi}(x_t) \varphi(x_t)^T$ is the sample covariance matrix of $\phi(x_t)$, λ_i is the non-zero eigenvalue, and μ_i is the corresponding eigenvector.

Then, KPCA was applied to the training dataset. The extracted non-linear components were projected to the datasets to produce the training and testing data for SVR. There are two well-known types of kernels—Polynomial and Gaussian. The Gaussian equation is shown in Equation (6).

Gaussian kernel (k_{ij} = exp
$$\left(-\frac{\parallel x_i - x_j \parallel^2}{2\sigma}\right)$$
, (6)

KPCA was applied to reduce the 176 reflectance wavelengths of the sugarcane canopy. The KPCA algorithm was run using Matlab environments developed in Reference [39]. The Gaussian function is selected as the kernel function of KPCA because the Gaussian kernel tends to give a good performance under general smoothness assumptions [40].

2.7. Mapping of Sugarcane CNC Using Multivariate Based Statistical Methods

Stepwise Multiple Linear Regression (SMLR) was applied to estimate the CNC in sugarcane from the pixel reflectance spectra and its transformations. The statistical package software SPSS was used for creating the regression model. Estimation models were calibrated using three different datasets—FDS, CRDR and BD. SMLR started running the regression equation with no variable. At each step, the most statistically significant wavelength with the highest or lowest *p*-value was added to the equation [41]. Wavelength variables were sorted. The variable was retained when the variables were added. However, all other wavelengths in the proposed model were re-tested to check if they were still contributing to the capability of the model.

In the case of a wavelength that no longer contributed to the model significantly, the selected variables were removed. In this study, *p*-values to enter and remove wavelengths were set at 0.01 and 0.02, respectively. The maximum number of selected wavelengths was limited to five to avoid an over fitting problem. The multicolinearity of variables was validated using a variance inflation factor (VIF) that must be lower than three [17].

2.8. Mapping of Sugarcane CNC Using SVR Based Machine Learning Approach

In the process of SVR training, the users must select an appropriate kernel and a set of parameter values. The kernel function transforms the non-linear data boundaries in an original data space into linear ones in the higher dimensional feature space. In this study, linear and Radial Basis Function (RBF) were used as the tested kernel function. The optimal values of free SVR parameters were determined based on the Leave One Out (LOO) cross validation, including a penalty factor (*C*), the epsilon value of Vapnik's insensitive loss function (ε) and the RBF kernel parameter (γ).

In addition, the best value for the kernel parameter was chosen based on LOO cross validation. The number of used PCs was increased one by one, related to the eigenvalue sorted. The raw data in terms of x, y and z values were transformed into the higher dimensional dot product feature space by the function of Φ . This transformation is done by using a kernel function. In this experiment, a linear kernel (K(x, x_i) = (x x x_i)) and an RBF kernel (K(x, x_i) = exp($-\gamma \mid |x-x_i| \mid 2$), $\gamma > 0$) were used. This step was run using Matlab software (MathWorks, Natick, MA, United States).

2.9. Model Validation

The accuracy of the estimated models was summarized and shown in terms of the coefficient of determination (R^2) and the Root Mean Square Error of the Estimate (*RMSE*) as illustrated in Equation (9) [42]. The Leave One Out (LOO) method was adopted to validate the proposed estimation models (n = 70).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_{i} - y_{i})^{2}}{n}}$$
(7)

where \hat{y}_i and y_i are the estimated and measured crop variables, respectively, and n is the number of samples.

3. Results

3.1. Descriptive Statistics of Measured Nitrogen Concentration

Table 2 shows the variation of sugarcane foliar nitrogen concentrations, which were collected from the field plots. The Kolmogorov-Smirnov method was applied to test the distribution normality of sugarcane nitrogen concentration at the 95% confidence level (p = 0.05). The null hypothesis tested was that the data follow a normal distribution (Ho: p > 0.05) versus the alternative hypothesis (Ha: p < 0.05) that the data did not follow a normal distribution. The results showed that the data used in this experiment were normally distributed (p > 0.05). However, the ranges of nitrogen content in the leaf samples are quite small. This is mainly because sugarcane plots in the study areas are in the mature stage (9–10 months). Most of the sugarcane plots of less than 3 ha in size were cultivated in the rainfed zone. In addition, the local farmers usually supply a low amount of N-fertilizer to their plots due to the limitation of costs. Thus, mean sugarcane productivity in Thailand is quite low when compared to the productivity in Brazil or South Africa.

Table 2. Descriptive statistics of the sugarcane nitrogen concentration in leaf samples (un	it: % nitrogen).

Data Set	Min	Max	Mean	Std Deviation
K84-200	0.508	0.663	0.591	0.065
KK-3	0.626	0.843	0.728	0.090
LK92-11	0.597	0.835	0.720	0.083
UT-8	0.513	0.644	0.586	0.047
Pooled	0.508	0.843	0.669	0.067

3.2. Estimation of Sugarcane CNC from Hyperion Satellite Image using a Multiple Linear Regression Approach

The number of spectral wavelengths was selected to estimate the CNC using a SMLR ranging from two to four. The results demonstrated that the best model yielded the correlation coefficient with nitrogen content values of 0.74 and *RMSE* values of 0.038% nitrogen. The sensitive wavelengths are listed in Table 3. The proposed models were validated using the LOO method. The regression equation is $Y = 10.75x_1 - 17.54x_2 - 11.75x_3 + 7.366x_4 + 0.69$. The scatter plots in Figure 3 show the relationships between the measured and estimated CNC in each sugarcane cultivar. Sugarcane cultivars influence the measured CNC. The percentage of measured canopy nitrogen of KK-3 is higher than that of LK92-11, UT-8 and K84-200, respectively. The spatial distribution of the sugarcane CNC, which was estimated by the SMLR data, are depicted in Figure 4. The CRDR is the best estimator developed by SMLR for explaining the variations of sugarcane CNC. The regression equation is $Y = 10.75x_1 - 17.54x_2 - 11.75x_3 + 7.366x_4 + 0.69$.

Table 3. Performances of Stepwise Multiple Linear Regression (SMLR) in determining sugarcane Canopy Nitrogen Concentration (CNC) with the mixed cultivar data sets (unit: % nitrogen).

		Pooled Data Set		
Variable	wavelengths λ (nm) –	R^2_{cv}	<i>RMSE</i> _{cv}	
FDS	660/721/1134/1265	0.73	0.039	
CRDR	721/1205/1265/1769	0.74	0.038	
BD	721/742	0.60	0.042	

 $RMSE_{cv}$: root mean square error of validation using LOO technique, expressed as %N; R^2_{cv} : relative coefficient of determination; Fit between estimated and observed values at the 0.01 level was considered highly significant.



Figure 3. Measured versus estimated CNC in each sugarcane cultivar; the model was developed by a Stepwise Multiple Linear Regression (SMLR) approach with three different datasets: (**a**) First derivative spectrum (FDS); (**b**) Continuum-Removed Derivative Reflectance (CRDR); and (**c**) Band depth (BD).



Figure 4. Spatial distribution of canopy nitrogen concentration (% nitrogen) in sugarcane using a Stepwise Multiple Linear Regression approach (SMLR) calculated from CRDR data sets.

3.3. Estimation of Sugarcane CNC from Hyperion Satellite Image Using Support Vector Regression

The high dimensionality of the orbital Hyperion image was reduced using the KPCA method. The Gaussian kernel with δ value of 1 was applied to the FDS, CRDR and BD data sets, respectively. With the spectral data of 176 bands, 76.84 percentage of the variability of FDS was explained by the first 20 Principal Components (PCs). In the case of BD, the first 10 PCs contributed more than 90 percent of the 68 absorption features. It can be concluded that an eigenvalue of the first PCs computed from the BD data set is rather higher than those computed from the CRDR and FDS datasets, respectively.

Table 4 summarizes the best estimation models developed from the FDS, CRDR and BD datasets with their optimal number of PCs and SVR parameters. The result shows that an estimation model generated from BD yields a higher accuracy than those computed from CRDR and FDS. Eleven components of BD-KPCA could produce the robust estimation model with the R^2 values of 0.78 and *RMSE* values of 0.035%, validated with the LOO method. Eleven components of BD-PCA contained 92.12 percent of the spectral information.

Table 4. Performance of the best estimation models computed from the FDS, CRDR and BD datasets using SVR method, (unit: % nitrogen).

Estimation Model	Data Set	No. of PCs	Optimal SVR Parameters	$R^2_{\rm cv}$	<i>RMSE</i> _{cv}
SVR/Linear	FDS	15	$C = 0.82, \epsilon = 0.03$	0.63	0.043
SVR/RBF	FDS	17	$C = 0.47, \varepsilon = 0.03, \gamma = 0.03$	0.65	0.041
SVR/Linear	CRDR	11	$C = 0.65, \epsilon = 0.04$	0.66	0.042
SVR/RBF	CRDR	13	$C = 0.8, \varepsilon = 0.02, \gamma = 0.06$	0.74	0.038
SVR/Linear	BD	10	$C = 0.42, \varepsilon = 0.04$	0.73	0.039
SVR/RBF	BD	11	$C = 1.3, \varepsilon = 0.01, \gamma = 0.04$	0.78	0.035

C, ε is the common parameter for SVR and γ is the parameter for RBF kernel.

Figure 5 depicts the scatter plots of measured CNC versus estimated CNC in each sugarcane cultivar developed by the SVR method for model calibration and KPCA for reducing the feature dimensions. The optimal estimation model based on the SVR-RBF method was applied to map the spatial distribution of CNC in the cropland illustrated in Figure 6.



Figure 5. Cont.



Figure 5. Measured versus estimated CNC in each sugarcane cultivar; the model was developed with the SVR/KPCA approach with three different datasets; (**a**) FDS; (**b**) CRDR and (**c**) BD.



Figure 6. Spatial distribution of canopy nitrogen concentration (% nitrogen) in sugarcane using Support Vector Regression (SVR) based RBF Kernel calculated from 11 KPCA components of BD data set, SVR free optimal parameters are C = 1.3, $\varepsilon = 0.01$, $\gamma = 0.04$.

4. Discussion

In this experiment, three predictive methods were applied to develop estimation models for sugarcane CNC in the mature stage (11–12 months). The transformed spectral information derived from spaceborne hyperspectral data was employed. The used technical methods for estimating CN consist of SMLR and SVR. The optimal model generated by the SVR-RBF kernel method using the BD dataset as the independent variable yields the highest coefficient of determination value of 0.78 and the lowest *RMSE* values of 0.035% nitrogen. Our study proves that spectral signals derived from hyperspectral data can possibly be applied to map the sugarcane nitrogen concentration in cropland with the combined cultivar environment. With the proposed SVR predictive model, it is

practical to monitor the nutrition status of sugarcane over large areas with the spectral signature and a nondestructive method.

The results indicate that the estimation models developed by SVR based on a non-linear kernel (RBF) yields a higher correlation coefficient with sugarcane CNC compared to the other models. The predictive model developed by the SVR method possibly improves by approximately 4% of the CNC predictive accuracy when compared with the SMLR model. This is because the information about the relationship between dependent and independent variables is not linear. In addition, the noise redundancy of the spectral data, which were observed from the space was removed before calibrating the model using the KPCA. Subsequently, the raw reflectance was transformed into FDS, CRDR and BD. Such techniques were performed to enhance the absorption regions, which were usually affected by canopy backgrounds [37]. It could be concluded from the study that BD data is the best estimator for the estimation model generated by SVR. BD was calculated from continuum-removed spectra centered in the regions of visible, red-edge and shortwave infrared. Such spectral regions are widely known as the pigment and water content absorption zones. BD is meaningful data, particularly in the case of limited reflectance value due to the saturation problem, especially in the mature stages of sugarcane. This conclusion is the same as those of previous studies [12,43,44].

The results of the study indicate that there is variability of the sugarcane nitrogen content between the eastern and western sites of the study area. This is mainly due to the structure of the sugarcane canopy and water stress. In fact, soil type, topography, planting date and fertilizer practices in the study area are not different. Sugarcane planting cultivars in the western site are mostly KK3 and LK92-11 and in the eastern site are mostly K84-200 and UT-8. At the edge of the eastern region is a rainfed agriculture that relies on natural rainfall. Structures of sugarcane canopy are explained by leaf orientation, which has a significant bearing on its spectral signature. The canopy structure of existing cultivars in the study areas could possibly be classified into planophile (KK-3 and LK92-11) and erectophile structure (K84-200 and UT-8) [10]. In review, sugarcane leaves with a planophile structure contains a higher nitrogen concentration than leaves with an erectophile structure [12]. This is consistent with the findings from the different crop types, such as grassland, winter wheat, cotton, paddy rice [43]. The planophile structure (30 degrees) contributes to significantly greater reflectance in the near infrared and to a greater absorption in the red due to the larger leaf areas facing towards the sun light compared with the erectophile structure (65 degree) [10,45]. The small portions of the leaf area and water stress, due to the variation in rainfall, results in the low rate of photosynthesis, chlorophyll and nitrogen content.

The narrow sensitive spectral wavelengths extracted from hyperspectral image were utilized to estimate sugarcane CNC. The selected wavelengths are centered in the visible-red, longer portion of red edge, far-near infrared and shortwave infrared regions of the electromagnetic spectrum. The red-edge region, which centered at 680–780 nm, has been indicated to be more sensitive to water content, chlorophyll and nitrogen [46]. This is mainly due to the low reflectance in the red region because of chlorophyll absorption and high reflectance in the near-infrared region.

The two most frequently occurring hyperspectral bands in calibrating the estimation model were centered at 660 nm and 721 nm. The narrow band centered at 660 nm is known to be more sensitive to changes in chlorophyll content [5]. The changes in biomass, genetic cultivar, canopy structure, moisture, and nitrogen concentration affect the absorption in the red visible region [47]. The maximum change in the slope of the reflectance spectra on the red-edge region usually occurs at 720 nm [48]. It could be concluded that the variation in crop growth, vegetation stress, leaf area index, chlorophyll and nitrogen status in plants could be detectable by wavelength 720 nm [49]. The wavelengths, centered in the far-near infrared and shortwave infrared are sensitive to the plant canopy moisture. The centered wavelength at 1205 nm and 1265 nm, which related to the leaf nitrogen concentration in sugarcane at the field level [12], yields a high correlation with the nitrogen content.

Most of the previous research attempts to explain the spatial variation of nitrogen content in paddy rice, wheat, cotton and natural grass using the spectral signal both at the foliage and the canopy [28],

but a little in sugarcane. There are a few experiments that were conducted, both at a leaf level [4,12] and at a space level [50] by applying the Random Forest (RF) algorithm used for feature selection and regression in hyperspectral data necessary for predicting sugarcane leaf nitrogen concentration. The results show that sugarcane leaf N concentration can be predicted using non-linear RF regression ($R^2 = 0.67$) with root mean square error of validation 0.15% derived from the first-order derivative of reflectance.

The results of the study indicate the high potential of SVR for estimating sugarcane nitrogen concentration from hyperspectral data in the combined cultivar environment. The efficiency of the proposed estimation models is more stable and reliable than those in previous studies. With the success in mapping nitrogen content across larger geographic areas from spaceborne hyperspectral data, it allows cane growers to monitor the nutritional status of the sugarcane earlier so that the nutrient deficient areas could be corrected by applying suitable management. However, the proposed method for nitrogen estimation needs to be validated at other sites. This study was conducted at one site representing the sugarcane plots. It remains to be seen whether sugarcane in different physical and climatic environmental conditions affects the spectral response and nitrogen content. To combine hyperspectral data for a precise agriculture concept requires the availability of spatially and temporally continuous imaging spectroscopy. In Precision agriculture studies from space images, date differences between image acquisition and field surveys results in some estimated errors for estimating sugarcane nitrogen.

5. Conclusions

The main objective of this research was to estimate the sugarcane Canopy Nitrogen Concentration (CNC) from the spaceborne hyperspectral data. The capabilities of two well-known methods were evaluated and applied to describe the spatial variations of CNC. The methods used to estimate nitrogen concentration consist of Stepwise Multiple Linear Regression (SMLR) and Support Vector Regression (SVR). Derivative spectra and transformed absorption features were calculated and used as an independent variable in the estimation models. The most important conclusion from this experiment is that spectral signals from the spaceborne hyperspectral data contain meaningful information for quantifying sugarcane CNC across larger geographic areas. The nutrient deficient areas could be corrected in time by applying suitable farm management. This is a benefit that can reduce intensive field visiting.

The variability of the sugarcane canopy nitrogen content in the study area is mainly due to canopy structure and water availability. Our results showed that sugarcane CNC can be accurately estimated by the proposed method. The best estimation model for sugarcane CNC was developed by non-linear transformed Radial Basis Function (RBF) kernel based SVR using Principal Components (PCs) of Band Depth data (BD). Such a model could explain 78% of CNC variation in cropland with a Root Mean Square Error (*RMSE*) value of 0.035% nitrogen. The SVR method has a higher performance compared to the SMLR method for estimating the sugarcane CNC at the mature stage. One of the research targets is to compare the performance of BD, First Derivative Spectrum (FDS) and Continuum-Removed Derivative Reflectance (CRDR) to enhance the absorption regions of the raw spectral data. The results of the study indicate that BD is the best independent variable for calibrating and validating the estimation model compared to FDS and CRDR. The high potential of hyperspectral wavelengths that correlate with the concentration of nitrogen at the sugarcane canopies are centered in the visible-red, longer portion of red edge, far-near infrared and shortwave infrared regions of the electromagnetic spectrum. The proposed estimation model was developed under the combined sugarcane cultivars and measured at the mature stage. The narrow wavelengths selected for this experiment are closely related to the changes in chlorophyll, nitrogen, biomass, and canopy moisture. The upcoming developments in hyperspectral sensors and of estimation methods allow us to monitor nutrient deficient areas, water stress and to predict crop production.

Acknowledgments: This research was supported by the National Center for Genetic Engineering and Biotechnology (BIOTEC), 111 Thailand Science Park (TSP), Phahonyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand, Ministry of Science and Technology, Thailand. We are grateful to the kind assistance of the ES Research and Development Co. LTD., and the Agricultural System and Engineering Laboratory, Asian Institute of Technology.

Author Contributions: Poonsak Miphokasap and Wasinee Wanasiri developed the research plan, and designed the experiments. Both authors analyzed the data and wrote the paper. Both authors equally contributed to the editing of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNC	Canopy Nitrogen	Concentration

- SVR Support Vector Regression
- RBF Radial Basis Function
- SMLR Stepwise Multiple Linear Regression
- KPCA Kernel Principal Component Analysis
- FDS First Derivative Spectrum
- CRDR Continuum-Removed Derivative Reflectance
- BD Band Depth
- LOO Leave One Out cross validation

References

- 1. Grof, C.P.L.; Campbell, J.A. Sugarcane sucrose metabolism: Scope for molecular manipulation. *Aust. J. Plant Physiol.* **2001**, *28*, 1–12. [CrossRef]
- Zarco-Tejada, P.J.; Ustin, S.L.; Whiting, M.L. Temporal and spatial relationship between within field yield variability in cotton and high spatial hyperspectral remote sensing imagery. *Agron. J.* 2005, *97*, 641–652. [CrossRef]
- 3. Field, C.; Mooney, H.A. The photosynthesis-nitrogen relationship in wild plants. In *On the Economy of Plant Form and Function*; Givnish, T.J., Ed.; Cambridge University Press: Cambridge, UK, 1986; pp. 25–55.
- 4. Abdel-Rahman, E.M.; Ahmed, F.B.; van den Berg, M. Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12S*, S552–S557. [CrossRef]
- 5. Mohan, B.K.; Porwal, A. Hyperspectral image processing and analysis. *Curr. Sci.* **2015**, *108*, 833–841.
- 6. Kokaly, R.F.; Despain, D.G.; Clark, R.N.; Livo, K.E. Mapping vegetation in Yellow stone National Park using spectral feature analysis of AVIRIS data. *Remote Sens. Environ.* **2003**, *84*, 437–456. [CrossRef]
- 7. Schmidt, K.S.; Skidmore, A.K. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sens. Environ.* **2003**, *85*, 92–108. [CrossRef]
- 8. Johnson, R.M.; Richard, E.P., Jr. Prediction of sugarcane sucrose content with high resolution, hyperspectral leaf reflectance measurements. *Int. Sugar J.* **2011**, *113*, 48–55.
- 9. Mokhele, A.; Ahmed, F.B. Estimation of leaf nitrogen and silicon using hyperspectral remote sensing. *J. Appl. Remote Sens.* **2010**, *4*, 1–18.
- 10. Galvao, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of Sugarcane Varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sens. Environ.* **2005**, *94*, 523–534. [CrossRef]
- Apan, A.; Held, A.; Phinn, S.; Markley, J. Spectral Discrimination and Classification of Sugarcane Varieties using EO-1 Hyperion Hyperspectral Imagery. In Proceedings of the 25th Asian Conference on Remote Sensing, Chiang Mai, Thailand, 22–26 November 2004; Available online: www.gisdevelopment.net/aars/ 2004/hyper/acrs2004_a1001.asp (accessed on 15 May 2014).
- 12. Miphokasap, P.; Honda, K.; Vaiphasa, C.; Souris, M.; Nagai, M. Estimating canopy nitrogen concentration in sugarcane using field imaging spectroscopy. *Remote Sens.* **2012**, *4*, 1651–1670. [CrossRef]
- 13. Lebourgeois, V.; Bégué, A.; Labbé, S.; Mallavan, B.; Prévot, L.; Roux, B. Can Commercial Digital Cameras Be Used as Multispectral Sensors? A Crop Monitoring Test. *Sensors* **2008**, *8*, 7300–7322. [CrossRef] [PubMed]

- Bappel, E.; Begue, A.; Despinoy, M.; Buchon, Y.; Siegmund, B. Spectral indices as bio-indicators of sugarcane crop condition from hyperspectral CASI data. In Proceedings of the IEEE International Geosciences and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003; pp. 516–563.
- 15. Matson, P.; Johnson, L.; Billow, C.; Milier, J.; Pu, R. Seasonal patterns and remote spectral estimation of canopy chemistry across the Oregon transect. *Ecol. Appl.* **1994**, *4*, 280–298. [CrossRef]
- Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C.; Corsi, F.; Cho, M. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS J. Photogramm. Remote Sens.* 2008, 63, 409–426. [CrossRef]
- 17. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182. [CrossRef]
- 18. Li, F.; Miao, Y.; Chen, X.; Zhang, H.; Jia, L.; Bareth, G. Estimating Winter Wheat Biomass and Nitrogen Status using an Active Crop Sensor. *Intell. Autom. Soft Comput.* **2010**, *16*, 1219–1228.
- Wang, Z.; Wang, T.; Darvishzadeh, R.; Skidmore, A.K.; Jones, S.; Suarez, L.; Woodgate, W.; Heiden, U.; Heurich, M.; Hearne, J. Vegetation Indices for Mapping Canopy Foliar Nitrogen in a Mixed Temperate Forest. *Remote Sens.* 2016, *8*, 491. [CrossRef]
- 20. Gökkaya, K.; Valerie Thomas, V.; Noland, T.L.; McCaughey, H.; Morrison, I.; Treitz, P. Prediction of Macronutrients at the Canopy Level Using Spaceborne Imaging Spectroscopy and LiDAR Data in a Mixedwood Boreal Forest. *Remote Sens.* **2015**, *7*, 9045–9069. [CrossRef]
- 21. Curran, P.J.; Dungan, J.L.; Peterson, D.L. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies. *Remote Sens. Environ.* **2001**, *76*, 349–359. [CrossRef]
- Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 1999, 67, 267–287. [CrossRef]
- 23. Martin, M.E.; Aber, J.D. Estimation of forest canopy lignin and nitrogen concentration and ecosystem processes by high spectral resolution remote sensing. *Ecol. Appl.* **1997**, *7*, 431–443. [CrossRef]
- 24. Mutanga, O.; Skidmore, A.K. Continuum-removed absorption features estimate tropical savanna grass quality in situ. In Proceedings of the 3rd EARSeL Workshop on Imaging spectroscopy, EARSeL, Hersching, Germany, 13–16 May 2003; pp. 542–558.
- 25. Motulsky, H.J.; Christopoulos, A. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, 2nd ed.; GraphPad Software Inc.: San Diego, CA, USA, 2003.
- 26. Foody, G.M. Land cover classification by an artificial neural network with ancillary information. *Int. J. Geogr. Inf. Syst.* **1995**, *9*, 527–542. [CrossRef]
- 27. Keiner, L.E.; Yan, X.H. A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery. *Remote Sens. Environ.* **1998**, *66*, 153–165. [CrossRef]
- 28. Mutanga, O.; Skidmore, A.K. Integrating imaging spectrometry and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sens. Environ.* **2004**, *90*, 104–115. [CrossRef]
- 29. Cortes, C.; Vapnik, V. *Support-Vector Networks, Machine Learning*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995; pp. 273–297.
- 30. Vapnik, V. Structure of Statistical Learning Theory. In *Computational Learning and Probabilistic Reasoning*; Gammerman, A., Ed.; John Wiley and Sons: Hoboken, NJ, USA, 1996; pp. 33–41.
- 31. Huang, C.L.; Wang, C.J. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* 2006, *31*, 231–240. [CrossRef]
- 32. Cherkassky, V.; Ma, Y. Comparison of model selection for regression. *Neural Comput.* **2003**, *15*, 1691–1714. [CrossRef] [PubMed]
- 33. Durbha, S.S.; King, R.L.; Younan, N.H. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sens. Environ.* **2007**, *107*, 348–361. [CrossRef]
- 34. Camp-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Vila-Frances, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97. [CrossRef]
- 35. Thenkabail, P.S.; Enclona, E.A.; Ashton, M.S.; van der Meer, V. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ.* **2004**, *91*, 354–376. [CrossRef]

- Datt, B.; McVicar, T.R.; Van Niel, T.G.; Jupp, D.L.B.; Pearlman, J.S. Preprocessing EO-1 Hyperion hyperspectral Data Support the Application of agricultural indexes. *IEEE Trans. Geosci. Remote Sens.* 2003, 41, 1246–1259. [CrossRef]
- 37. Dawson, T.P.; Curran, P.J. A new technique for interpolating the reflectance red edge position. *Int. J. Remote Sens.* **1998**, *19*, 2133–2139. [CrossRef]
- 38. De Jong, S.M. Imaging spectrometry for monitoring tree damage caused by volcanic activity in the long valley caldera, California. *Int. J. Appl. Earth Obs. Geoinf.* **1998**, *1*, 1–10.
- 39. Van der Maaten, L.; Postma, E.; Van Den Herik, H. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 1–41.
- Cao, J.J.; Lee, S.C.; Ho, K.F.; Zhang, X.Y.; Zou, S.C.; Fung, K.K.; Chow, J.C.; Watson, J.G. Characteristics of carbonaceous aerosol in Pearl River Delta region, China during 2001 winter period. *Atmos. Environ.* 2003, 37, 1451–1460. [CrossRef]
- Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C.; Corsi, F.; Cho, M. Estimation of leaf area index and chlorophyll for a Mediterranean grassland using hyperspectral data. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, China, 3–11 July 2008; pp. 409–426.
- 42. Yi, C.H.; Sogah, D.K.; Boyce, M.; Degterev, A.; Christofferson, D.E.; Yuan, J. A genome-wide RNAi screen reveals multiple regulators of caspase activation. *J. Cell Biol.* **2007**, *179*, 619–626. [CrossRef] [PubMed]
- 43. Mutanga, M.; Skidmore, A. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrusciliaris) measured under controlled laboratory conditions. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 87–96. [CrossRef]
- 44. Yao, X.; Huang, Y.; Shang, G.; Zhou, C.; Cheng, T.; Tian, Y.; Cao, W.; Zhu, Y. Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration. *Remote Sens.* **2015**, *7*, 14939–14966. [CrossRef]
- 45. Thenkabail, P.S. Optimal hyperspectralnarrowbands for discriminating agricultural crops. *Remote Sens. Rev.* **2002**, *20*, 257–291. [CrossRef]
- 46. Li, F.; Miao, Y.; Feng, G.; Yuan, F.; Yue, S.; Gao, X.; Liu, Y.; Liu, B.; Ustin, S.L.; Chen, X. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Res.* **2014**, *157*, 111–123. [CrossRef]
- 47. Blackburn, G.A. Quantifying chlorophylls and carateniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sens. Environ.* **1998**, *66*, 273–285. [CrossRef]
- 48. Mutanga, O.; Skidmore, A.K. Red edge shift and the quality of tropical grass canopies. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 34–42. [CrossRef]
- 49. Elvidge, C.D.; Chen, Z. Comparison of broadband and narrow-band red and near-infrared vegetation indices. *Remote Sens. Environ.* **1995**, *54*, 38–48. [CrossRef]
- Abdel-Rahman, E.M.; Ahmed, F.B.; Ismail, R. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *Int. J. Remote Sens.* 2013, 34, 712–728. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).