Estimating leaf nitrogen concentration from similarities in fresh and dry leaf spectral bands using a model population analysis framework

Cecilia Masemola[a,b,*] and Moses Azong Choa,[c]

[a]Earth Observation Research Group, Natural Resources and Environment, Council for Scientific and Industrial Research, Pretoria, South Africa;

[b]College of Agriculture and Environmental Science, University of South Africa, Pretoria, South Africa;

[c]Forest Science Postgraduate Programme, Department of Plant and Soil Science, University of Pretoria, South Africa

*Correspondence to: Cecilia Masemola; Earth Observation Research Group, Natural Resources and Environment, Council for Scientific and Industrial Research, P O Box 395, Pretoria 0001, South Africa. Email: Cecilia.mulukwane@gmail.com

Abstract

Fresh leaf spectral reflectance is primarily influenced by leaf water content and structural aspects such as the inter-cellular spaces within the spongy mesophyll, which also interfere with the estimation of the leaf nitrogen content. It is therefore essential to identify spectral bands that are least affected by the above perturbing factors for improving leaf nitrogen estimation for fresh leaves across any landscape. Wavelengths selection plays a vital role in identifying the best spectral features for assessing leaf nitrogen concentration from hyperspectral data of dry and fresh leaves. The primary objective of this study was to determine typical optimal bands for leaf nitrogen estimation from spectra (400-2500 nm) of whole fresh and dry leaves for the same specimens of *Eucalyptus grandis*. This was achieved via the use of competitive adaptive re-weighted sampling (CARS), and Monte Carlo Cross-validation-competitive adaptive re-weighted sampling (MCCV-CARS) band selection approaches. Bands selected (931 nm, 1003 nm, 1027 nm, 1036 nm, 1177 nm, and 1180 nm) via the MCCV-CARS approach yielded the highest estimation accuracy for both fresh predicted coefficient of determination (R2cal) = 0.82 and predicted root mean square error (RMSEP) =0.14) and dry leaves (R2P = 0.88 and RMSEP = 0.13) when compared to CARS (2044 nm, 2107 nm, and 2188 nm) only. The identified spectral features could be relevant for assessing leaf nitrogen concentration for different seasons, for example, wet to dry season.

1. Introduction

Leaf nitrogen (LN) is one of the essential macro-nutrient required for chlorophyll (Chl) production and building of proteins in plants (Muñoz-Huerta *et al.* 2013). As a result, LN plays an important role in leaf physiological processes such as photosynthesis, respiration, and transpiration (Wang 2016). In ecology studies, LN is perceived as a vital indicator of productivity and nutrient efficiency (Ramoelo *et al*. 2011). Hence, the Convention on Biological Diversity (CBD, 2010) has added LN as one of the essential biodiversity variables that would be used to monitoring global biodiversity change (Pereira *et al*. 2013).

Remote sensing is widely considered a convenient tool for the quantification of plants structural (e.g. Leaf area index) and biochemical traits (e.g., Chl, LN and lignin) when compared to the conventional ground-based method at the landscape level. However, the estimation of LN from remote sensing data is a challenge because a significant content of it is allocated in the different organelles of leaf and only a small amount exists as free components (Wang *et al*. 2017). The spectral absorption features of LN are also confounded with those of other leaf biochemical. For instance, Curran (1989) and Fourty *et al*. (1996) have found that LN share absorption features with cellulose, lignin and starch. Consequently, it is a challenge to estimate LN content with broadband remote sensing data (e.g., Landsat imagery). Fewer broad bands tend to masks absorption features of LN (Vorovencii 2009; Wang 2016).

On the other hand, hyperspectral remote sensing has been used to estimate LN with high accuracy (Curran 1989; Kokaly and Clark 1999; Blackburn 2007; Gökkaya *et al*. 2015; Wang 2016). The data

has the potential to detect subtle absorption features of LN, lignin and cellulose (Kokaly and Clark 1999). Serrano *et al*. (2002), predicted LN and lignin from Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data in Mediterranean vegetation. Ramoelo *et al*. (2011) estimated LN concentration of savannah grass using field spectroradiometer. However, hyperspectral data analysis is challenging due to the high dimensionality of the data (Curran 1989; Fourty *et al*. 1996). In addition, overlapping features of other properties amplifies the challenge of using these sensors given that LN contents are found in small amounts in leaves (Wang 2016). Other than the issue of overlapping biochemical features, most absorption features of LN are located in the shortwave (SWIR) region which is also strongly influenced by leaf water content (Ramoelo et al 2011). Consequently, the fresh leaf reflectance spectra data have limitations for estimating LN throughout the reflectance spectrum (Kokaly and Clark 1999; Wang 2016).

Several techniques have been used to minimize the problem of leaf water. Commonly used techniques include continuum removal (CR) (Huang and Asner 2009; Schlerf *et al*. 2010; Ramoelo *et al*. 2011), wavelet analysis (Ferwerda and Jones 2006; Blackburn 2007) and partial least square regression (PLSR). Kokaly and Clark (1999) used CR to determine absorption features that have strong correlation with LN. The method did not only prove to enhance prediction accuracy of LN from dry ground leaves but also to reduce the impact of water from fresh leaf spectra (Mutanga *et al*. 2003, Ramoelo *et al*. 2012). Other methods such as orthogonal transformation methods have also proved to be useful for dimensionality reduction and optimum waveband selection.

Generally, continuum removal (CR) (Huang and Asner 2009; Schlerf *et al*. 2010; Ramoelo *et al*. 2011) and wavelet analysis (Ferwerda and Jones 2006; Blackburn 2007) are the most used methods to circumvent the multiple-collinearity problem in vegetation studies. In Kokaly and Clark (1999) study, CR determined absorption features centred at 2054 nm and 2172 nm to have a strong correlation with LN concentration. The method did not only prove to enhance prediction accuracy of LN from dry ground leaves but also to reduce the impact of water from fresh leaf spectral (Mutanga *et al*. 2003, Ramoelo *et al*. 2012). Other methods such as orthogonal transformation methods have also proved to be useful for dimensionality and variable selection strategies.

The orthogonal transformation methods include principal component analysis (PCA) (Wold *et al*. 2001) and Partial least square regression (PLS) (Wold *et al*. 2001). However, PCA is disadvantageous because principal components (PCs) are determined without using information about the property to be predicted. Contrary to PCA, PLS find factors by projecting predictors (wavelengths) (X) through the variable of interest (e.g. physiochemical properties) and use weights (W) to uphold orthogonal scores. Tran *et al*. (2014) highlighted the complexity of interpreting the models when latent variables (LVs) are used instead of the original features. They argued that, due to the nature of analytical data which are inclined to multiple sources of variability, it becomes challenging to directly interpret the model parameters only based on the weights and the loadings. Consequently, Tran *et al*. (2014) proposed significant Multivariate Correlation (sMC) PLS based model which not only use a latent vector to select variables but also correct autocorrelation factor.

Other studies have shown that the dimensionality of the data can be significantly improved by fusing wavelength selection algorithms (Li *et al*. 2009; Fan *et al*. 2012; Deng *et al*. 2015). The methods have also shown to improve the generalisation and the robustness of the

prediction models (Deng *et al*. 2015). The utility of Model-population analysis (MPA) methods have been demonstrated on spectral data (Deng *et al*. 2015). The most effective methods are competitive adaptive reweighted sampling (CARS) (Li *et al*. 2009), Monte-Carlo-Uninformative-Variable-Elimination (MC-UVE) (Li *et al*. 2009; Fan *et al*. 2012) and Monte Carlo-Cross-Validation (MCCV) (Xu and Liang 2001).

These techniques have been used in various studies to select significant predictor variables which are then used to construct the calibration model. However, the concept of MPA is commonly used within the chemo-metrics community (Yun *et al*. 2014; Deng *et al*. 2015; Wang 2016) and has been proven to enhance predictive accuracy and robustness in multi-variable models. To our knowledge, few studies have explored the MPA based strategy on vegetation, e.g. Crops (Yun *et al*. 2014; Wang *et al*. 2017). Yun *et al*. (2014) predicted moisture content of corn using near-infrared (NIR) spectral region. Recently, Vohland *et al*. (2016) and Kawamura *et al*. (2017) used MPA to predict available nitrogen in soils.

The confounding effect of leaf water content on remote sensing of LN is still unresolved. Ramoelo *et al.* (2011) successfully used water-removed spectra to estimate LN. The water removed spectra were computed by a non-linear least-squares spectral matching method, deriving fresh leaf spectrum as a non-linear combination of a leaf water spectrum and a dry matter spectrum (Schlerf *et al.* 2010; Ramoelo *et al.* 2011). No study has directly compared the spectral features of intact fresh and dry leaves of the same sample for LN absorption features. Therefore, the question is, are there any common spectral bands for estimating LN in both fresh and dry leaf?

i. What are the common absorption features for quantifying LN concentration based on both dry and fresh leaf spectral?

ii. To compare the accuracy and robustness of the proposed strategies for identifying common spectral features for estimating LN concentration from spectral data fresh and dry leaves of the sample samples.

2. Material and Methods

2.1 Study Area

The *Eucalyptus grandis* (*E.grandis*) leaf sampling was conducted in the plantation site located in KwaZulu-Natal Midlands, South Africa (29°49 ′S 30º17 ′E). Plantation forestry in

the region is significant land use. The main species grown for commercial forestry include Pines, *Eucalyptus spp* and Australia native acacia species such as Black wattle. The average rainfall is 1000 mm p.a. While temperatures range from 24 ºC to 26 ºC in summer and drops to 5 ºC and 14 ºC in winter. The area has an altitude rising from 800 m to 1400 m above sea level. The landscape generally consists of undulating plains with steep river valleys. The geology of an area is mainly sandstone and clay formations (Cho et al. 2010).

2.2 Collection of reflectance spectra

2.2.1 Fresh leaf reflectance spectra

E. grandis leaves were collected at a plantation in the Richmond area of KwaZulu-Natal, South Africa (29o 49' S, 30o 17' E). To ensure variability in LN, the leaves were harvested from trees growing under three different site qualities (poor, medium and good) (Cho et al. 2010). The classification of the site quality is based on the total available water. To further increase variability on the leaf samples, the specimens consisted of leaves at different stages of development e.g. from the young and newly formed leaves at the tip to old and previous season's leaves at the base of the branch. In total 53 samples (consisting of 5-10 leaves) were collected, placed in sampling plastic sampled bags and stored in a cooler box.

Leaf reflectance of the adaxial (upper) surfaces were made using the ASD spectroradiometer (FieldSpec3 Pro FR, Analytical Spectral Device, Inc, USA) plant probe. The ASD radiometer covers the spectral range from 350 – 2500 nm with varying resolutions in different regions e.g. 3 nm in the 350-1050 nm range, 12 nm in the 1050-2500 nm region. However, the results are interpolated by the

ASD software to a resolution of 1 nm. We resampled the data to 3 nm. Two measurements were made per leaf at one-third and two-third of the distance from the leaf top. Two sets of spectral measurements were made. First, spectral measurements were conducted on the fresh leaf samples using the ASD leaf probe. The samples were then oven-dried at 60 oC for 24 hours after which leaf spectra were made on the intact dry leaves using the ASD leaf probe. The spectra of the fresh and dry leaves are shown in figure 1.

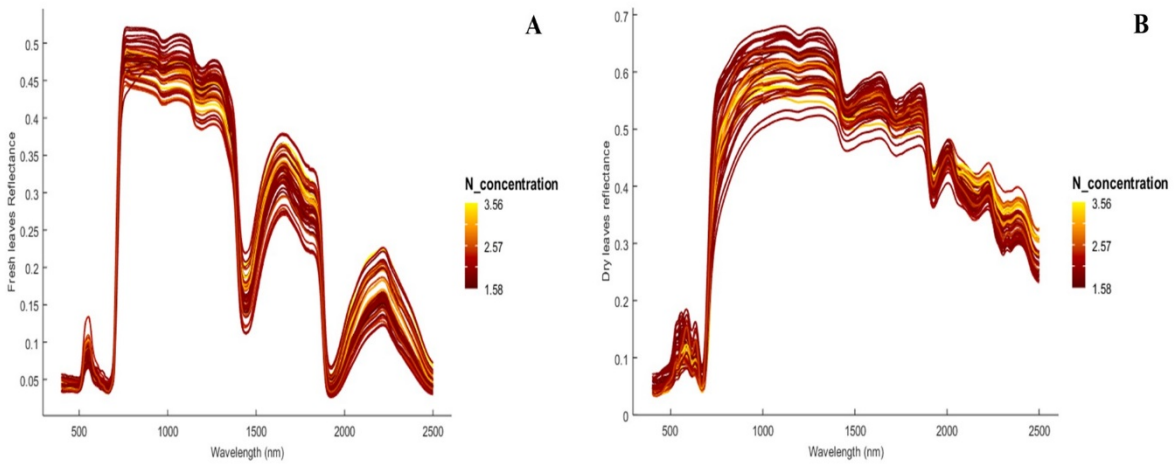

Figure 1. Demonstrate the original spectral and measured LN concentration of the two data sets (a) fresh leaf spectral reflectance and (b) dry leaf spectral reflectance.

2.2.2 Dry leaf reflectance spectral data and nitrogen measurements

Subsequent fresh leaf spectral reflectance measurements, the same leaf samples were oven dried for 24 hours at 60 °C and then milled in preparation for LN analysis. As outlined in Cho *et al*. (2010) LN concentration was measured using a Leco FP528 nitrogen analyser (Horneck and Miller 1998). The FP528 nitrogen analyser is based on rapid incineration of the dry leaf

sample followed by quantification of the thermal conductivity of the resulting LN oxides (NO2) and N2 gases for LN (Cho *et al*. 2010) (Fig. 1).

2.3 Spectral pre-processing

Spectral data pre-processing was conducted using the "Field Spectroscopy Facility (FSF) Post Processing Toolbox which is a MATLAB toolbox" (Robinson and Arthur 2012). The toolbox was used (*i*) for the exclusion of outliers caused by measurement errors and atmospheric interference (*ii*) the exclusion of 350–399 nm bands, thus limiting the spectral range to the traditional Visible (Vis) to shortwave infrared (SWIR) (400 nm to 2500 nm); (*iii*) eliminating sensor noise using a moving Savitzky-Golay filter (Savitzky and Golay 1964) with a nine-point window size and second polynomial order. Furthermore, we explored various spectral transformation algorithms to evaluate the impact of spectral transformation on species discrimination

.

2.4. Spectral transformation.

Various spectral transformation algorithms were implemented to assess the influence of spectral transformation on the calibration model. We considered the following methods: multiplicative scatters

correction (MSC) (Naes *et al*. 1990), standard normal variation (SNV) (Barnes *et al*. 1993) and Savitzky-Golay smoothing convolution (SG). The transformed spectral data were compared with un-transformed spectral data for estimating *E. grandis* LN concentration. Fig. 2, shows spectral data measured from fresh leaves after transformation. Both transformed and untransformed spectral data were used for the selection of common informative wavelengths for estimating LN with fresh and dry leaf spectral reflectance of *E. grandis*.

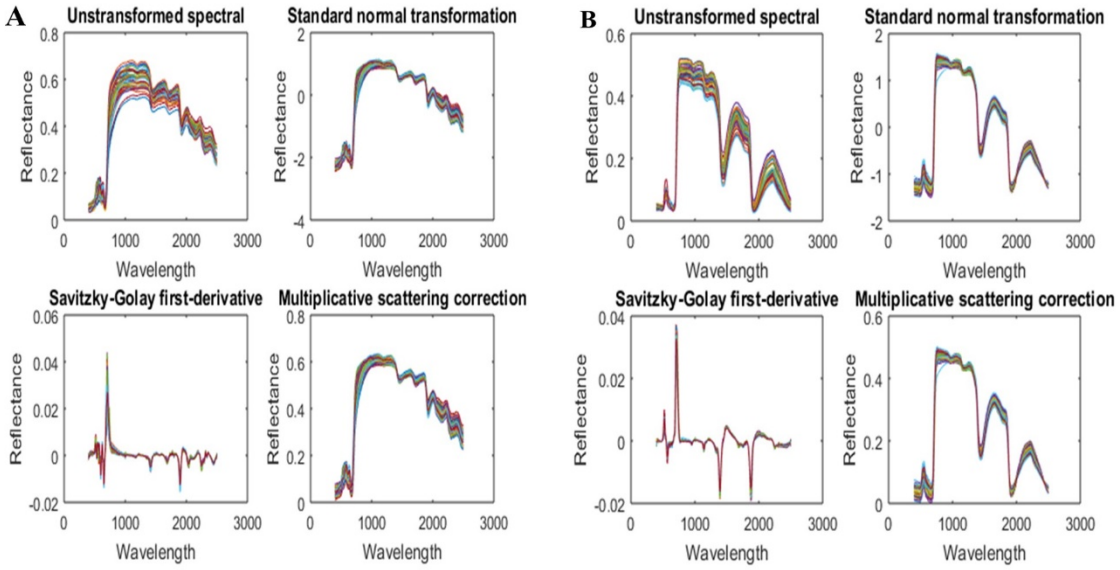

Figure 2. Spectral transformation results for both dry and fresh leaves: (*a*) denote transformed fresh leaf reflectance, and (*b*) represent transformed dry leaf reflectance of *E. grandis*. (*i*) depict untransformed leaf spectra, (*ii*) Standard normal transformation (*iii*) Savitzky-Golay smoothing convolution (SG) and (*iv*) multiplicative scatters correction (MSC).

3. Data analysis

3.1. Datasets and data pre-processing

The dataset used consists of three sets of transformed spectral reflectance of the *E. grandis* species measured from 400 nm to 2500 nm spectrum regions and corresponding LN as the response variable. The spectral data were measured with 3 nm increments on the wavelength scale. Consequently, a total of 702 dry and fresh leaf wavelengths were used. The models used have an embedded pre-processing method, namely, mean centring ('mean'), auto-scaling, MSC plus mean centring, and MSC plus autoscaling. The pre-processing allowed us to perform mean centring before data calibration.

3.2. Development of calibration models

Different multivariate calibration models have been employed to investigate the common wavelengths set that best predict LN concentration from both fresh and dry spectral reflectance data of *E. grandis* (Fig. 3). In particular, we employed various models based on model population analysis. MPA based framework provides an integrated environment for PLS regression (Li *et al*. 2014). For example, the framework uses various resampling and data reduction algorithms to harness the efficacy of the PLS calibration model. The advantage of the MPA framework includes the integrated pre-treatment of the data, detection of outlier and variable selection, model assessment and PLS regression modelling that facilitates calibration procedure (Li *et al*. 2014). The algorithms explored are Competitive adaptive reweighted sampling (CARS) and its extended version called Monte Carlo Cross Validation (MCCV)-

CARS and significant multivariate correlation (sMC)-PLS. The sMC–PLS using full spectral data was applied to evaluate the performance and parsimoniousness of the CARS-PLS and MCCVCARS-PLS regression models (Fig. 3).

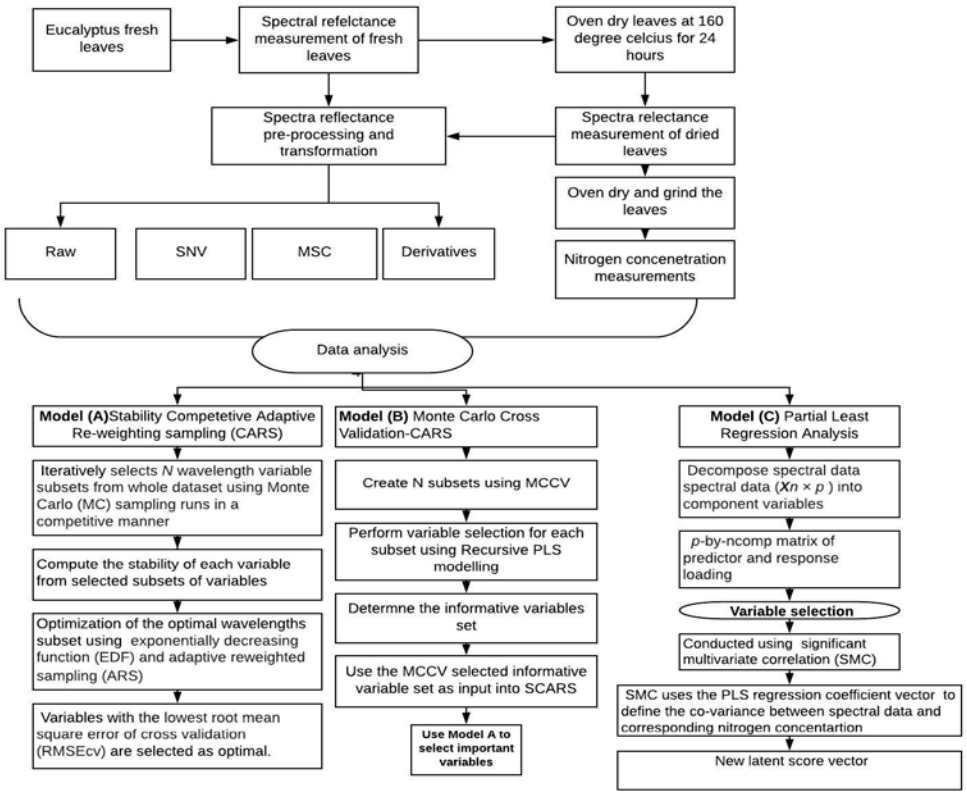

Figure 3. An overview of methods used for common wavelengths between dry and fresh leaf spectral data of E. grandis. In general, the models consist of spectral transformation, data analysis and accuracy assessment.

3.2.1 Monte Carlo Cross-Validation PLS

MCCV (Xu 2001) is a simple and effective sampling algorithm used to find the variety of sub-variable sets in multivariate data analysis. Regarding the best model, MCCV sampling method has been reported to be asymptotically consistent and to have a more substantial probability when compared to leave-one-out-cross-validation (LOO-CV) (Shao 1993). The basic idea behind the MCCV is to randomly split data into fractions using bootstrap sampling

strategy to form the training set and the test set. The process is repeated into a specified number of times and is defined as in Eq.1. Unlike conventional cross-validation, the MCCV based sampling is constructed in such a way that the partitions are done independently for each run, as such the same point can appear in the test set multiple times. In general, MCCV reduce the computational complexity of the model by allowing fewer samples to be used in model calibration.

$$MCCV_{(nv)}(k) = \frac{1}{Nnv}\sum_{i=1}^{N}\left(y_{sv(i)} - y_{k*_{v(i)}}\right)^2 \quad (1)$$

Where $y_{k*_{v(i)}}$ is the predicted nitrogen contents of Eucalyptus leaves in the form of a vector calculated using PLS model with k components based on calibration sub-sets without variable selection, *N* denote number of repeats MCCV is to perform.

Dry and fresh leaf spectral data together with the corresponding LN concentration were randomly partitioned into calibration (c) and validation (v) subsets. The subsets were then used to build a calibration model and calculate the corresponding mean square error cross-validation (MSECV) metrics using the validation dataset. The procedure was repeated until the model yielded stable calibration results regarding MSECV, root mean square error cross-validation (RMSECV) and optimal latent variables (LVs). The performance of the MCCV model depends on the number of iterations (*n*itr) and some information held by the variables left out for validation (nv). According to the rule of thumb, the number of iterations should be set to *n*itr = *m*2. In this study, the suitable number of the repeat was determined by performing MCCV based on different scenarios, i.e. *n*itr = 50, 100, 500 and 1000, respectively. The variables selected based on the MSE after cross-validation of each sub-models were used to calculate RMSECV for the predictive validity of the model, which was subsequently used to decide the optimal number of LVs.

Moreover, the corresponding RMSECV in the selected LVs number was also calculated. The comparison of the RMSECV, RMSEcal and LVs number at all MCCV repeat scenarios indicated that all statistical metrics became stable after 500 ni, as such this study used 500 repeats.

3.2.2 Competitive adaptive reweighted sampling (CARS)-PLS

CARS is an independent variable selection algorithm proposed by Li *et al*. (2009). In combination with PLS linear model, CARS has been effectively used for predictive analysis. As illustrated in Fig. 3 CARS partition data in *n* number of subsets of variables based on iterative sampling strategies. Each iteration is conducted in four successive steps, namely, (*i*) Monte Carlo model sampling, (*ii*) enforced wavelength reduction by EDF, (*iii*) competitive wavelength reduction by adaptive reweighted sampling (ARS) and (*iv*) RMSECV calculation for each subset. The combination of EDF and ARS is used for variables reduction and selection. According to Li *et al*. (2009), CARS employs a Darwin-like "survival of the fittest" and realises to some extent the selection of an optimal subset of predictors, in this case, fresh and dry leaf spectral data.

3.2.3 Significant Multivariate Correlation (sMC)-PLS

In order to comprehensively validate the performance of the models above, MCCV-PLS and MCCV-CARS-PLS, PLS is also employed. PLS is a multivariate regression that converts multi-collinearity variables into uncorrelated variables. The method selects the most widespread variables and detects the irrelevant variables using decomposing and filtering approach. In this study, we explore the Significant Multivariate Correlation (sMC) PLS based approach developed by Tran *et al*. (2014). The method follows under the filter variable selection criterion. According to Tran *et al*. (2014), sMC can correct the rotation effect that tends to introduce biases in the variability estimates used by methods such as selectivity ratio and target projection during variable importance selection. In this study, sMC was applied on PLS in which the methods were applied to each regression coefficient vector yielded by response variable. The sMC estimates the source of variability for each variable using regression

variance and residual variance resulting from the PLS regression coefficient vector. By so doing, sMC statistically determines the variable importance concerning the regression model. As described in Tran *et al*. (2014) the sMC-PLS start by defining the co-variance between *E. grandis* leaf spectral data and the corresponding measured LN concentration in combination with predicted LN concentration. Also, the models' algorithm defines a new latent score vector of the PLS model. PLS regression coefficients are projections of the ordinary least squares regression coefficients into space spanned by a matrix of leaf spectral data of the target species along a direction orthogonal to the loadings.

On the other hand, the sMC loadings are the normalised coefficients resulting from PLS regression which is then used to formulate the sMC score vector (Tran *et al*. 2014) and are proportional to the predicted LN concentration values. *F*-test is then used to evaluate the significance of the variables concerning their association to measured response values (*y*) using Eq. 2 to Eq.4.

$$\hat{b}_{pls} = RP'\left(X'X\right)^{-1}X'y = RP'b_{OLS}, R = \left(P'W\right)^{-1} \quad (2)$$

$$P_{sMC} = \frac{\hat{b}_{PLS}}{\left\|\hat{b}_{PLS}\right\|} \quad (3)$$

$$t_{sMC} = X\,p_{sMC} = X\frac{\hat{b}_{PLS}}{\left\|\hat{b}_{PLS}\right\|} = \frac{\hat{y}}{\left\|\hat{b}_{pls}\right\|} \quad (4)$$

$$X = t_{sMC}P'_{sMC} + E_{sMC} = \frac{\left(\hat{y}\hat{b}'_{PLS}\right)}{\left\|\hat{b}_{PLS}\right\|^2} + E_{sMC} \quad (5)$$

$$SS_{i,PLS_{regression}} = \left\|P_{sMCi}t_{sMC}\right\|^2 = \frac{\left\|\hat{y}b'_{plsi}\right\|^2}{\left\|b'_{plsi}\right\|^4} \quad (6)$$

$$ss_{i,PLSresidual} = \left\|e_{sMCi}\right\|^2 = \left\|x_i - \frac{(\hat{y}\hat{b}'_{plsi})}{\left\|\hat{b}_{plsi}\right\|^2}\right\|^2 \quad (7)$$

$$MS_{i,PLSregression} = SS_{i,PLSregression}/1 \quad (8)$$

$$MS_{i,PLSresidual} = SS_{i,PLSresidual}/(n-2) \quad (9)$$

$$F_{sMC} = \frac{MS_{i,PLSregression}}{MS_{i,PLSresidual}} = \frac{\left\|\frac{\left(\hat{y}\hat{b}'_{plsi}\right)}{\left\|\hat{b}_{pls}\right\|^2}\right\|^2}{\left\|x_i - \frac{\hat{y}\hat{b}'_{plsi}}{\left\|\hat{b}_{plsi}\right\|^2}\right\|^2 \Big/ (n-2)} \square F(1-\alpha,1,n-2) \quad (10)$$

where SS is a sums of squares, MS*i*, PLSregression calculated the explained variance, while MS*i*, PLSresidual find the residual variance.

4. Models Evaluation Metrics

The spectral data together with the corresponding LN concentration were randomly divided into two sets (training and prediction). The training set was used as a cross-validation set and the remaining data as an independent prediction set. To evaluate the reliability of the calibration model we used RMSE (Eq. 5) and coefficient of determination (*R*2) (Eq. 6). *R*2 compares the inconsistency of the estimation

errors with that of the measured values by computing the coefficient of determination value from measured and predicted data. On the other hand, RMSECV and RMSEP were applied to estimate the accuracy yielded between the predicted and the measured LN concentration based on the selected variables. The smaller the RMSE and the higher the *R*2, the better the performance of the wavelengths data employed in the model for predicting the LN concentration of the *E. grandis* species. For a good model, the RMSE for training and test sets has to be similar. The reverse means there was an over-fit of the data.

$$RMSECV\ /\ RMSEP = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y(\chi_i) - \hat{Y}(\chi_i)^2\right.} \quad (11)$$

In this case, n is the number of predictors (number of leaf spectra wavelengths), $Y(x_i)$ are the measured nitrogen concentrations of the species and $\hat{Y}(\chi_i)$ present the predicted nitrogen concentrations from calibration set (RMSECV), and that from independent validation test is shown by (RMSEP), respectively.

$$CV\ R^2 \Big/ P R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - y_m\right)^2} \quad (12)$$

In this study, R^2 is defined as the total variations described by the created regression models or the total amount of the total variation in nitrogen concentration explained by the leaf optical information. Where n represent, MSE and

5. Results

5.1 Full spectra data analysis

Fig. 4 presents the optimal number of PLS component to estimate LN concentration in each spectral region (Vis, red-edge, NIR, ESWIR, MSWIR and FSWIR) of the leaf full spectrum. The italic number on each spectral region depict the optimal number of PLS component in that region. Moreover, the dotted line is the RMSECV of the global model with six latent variables. From Fig. 4 it is evident that FSWIR region performed better for fresh spectral data. Whereas, MSWIR and FSWIR performed better for the dry leaf spectral data. In both dry and fresh spectral data red-edge regions poorly performed when compared to other regions with RMSECV which is above that of the global model. Regarding NIR, dry data seems to estimate LN concentration of *E. grandis* leaf better than that of Fresh leaf reflectance. On the other hands, Vis spectral performed the same in both data sets. The prediction performance of the best spectral regions was compared with CARS and MCCV-CARS based models.

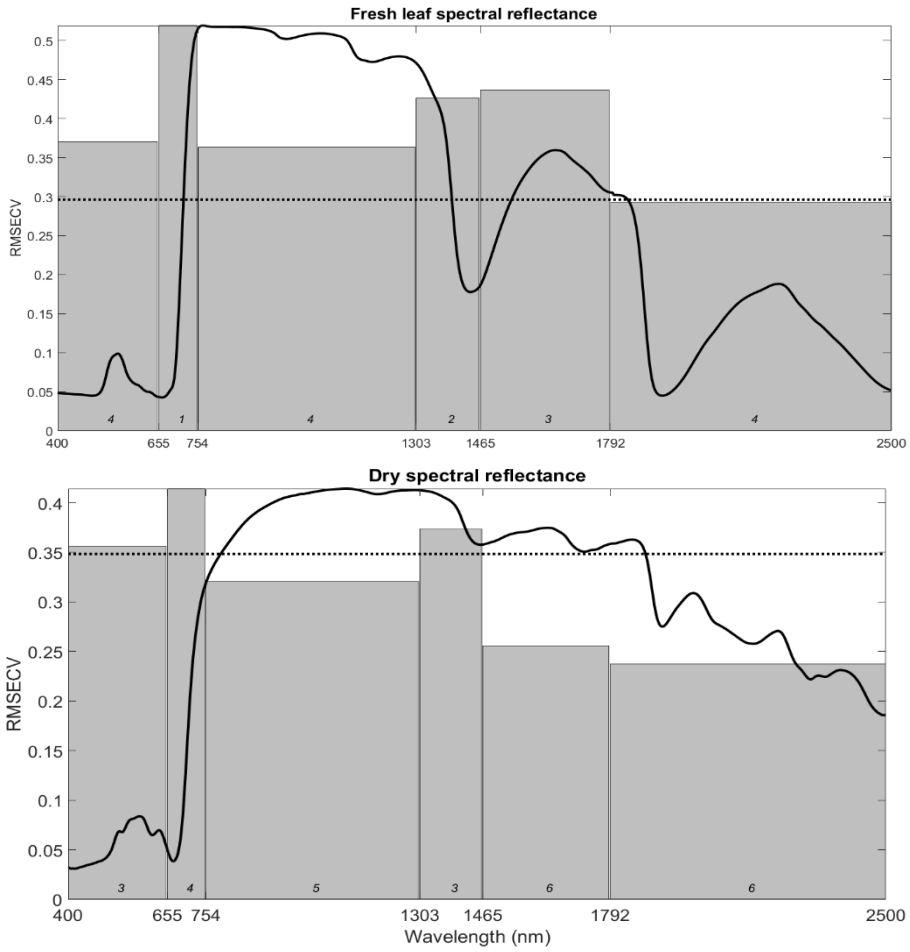

Figure 4. Optimal spectral regions used to develop PLS model for the estimation of leaf nitrogen.4 (*a*) and 4 (*b*) depict spectral regions for fresh and dry leaf reflectance, respectively. The italic number on each bar indicates the optimal number of PLS components in each spectral region and the dotted line is the RMSECV of the global model with four PLS components. Spectral regions -Vis (400- 655 nm), red-edge (657-754 nm), NIR (755-1303 nm), ESWIR (1301-1465 nm), MSWIR (1466-1792 nm) and FSWIR (1793-2500 nm).

5.2 Selection of important variables based on MCCV-CARS and CARS methods

From Fig. 5, it is evident that the ten-fold RMSECV value varied throughout CARS ran with fresh lead spectral and showed the least variations from dry leaf spectra data. With regards to the fresh spectral dataset, RMSECV increased after 30 runs which are attributed to the loss of informative wavelengths. RMSECV is a value of the wavelengths subset selected as optimal for the estimation of the LN content of the leaves. The subset with the lowest value is the most important in estimating LN concentration of the species. On the other hand, the informative variables were lost after 25 runs when using dry leaf spectra data. When MCCV-CARS-PLS was performed, the lowest RMSECV was reached at the 10th CARS runs with fresh spectra, whereas with dry leaf it was reached after 30 runs. The selected optimal wavelengths and their performance in estimating LN concentration are shown in Table 2.

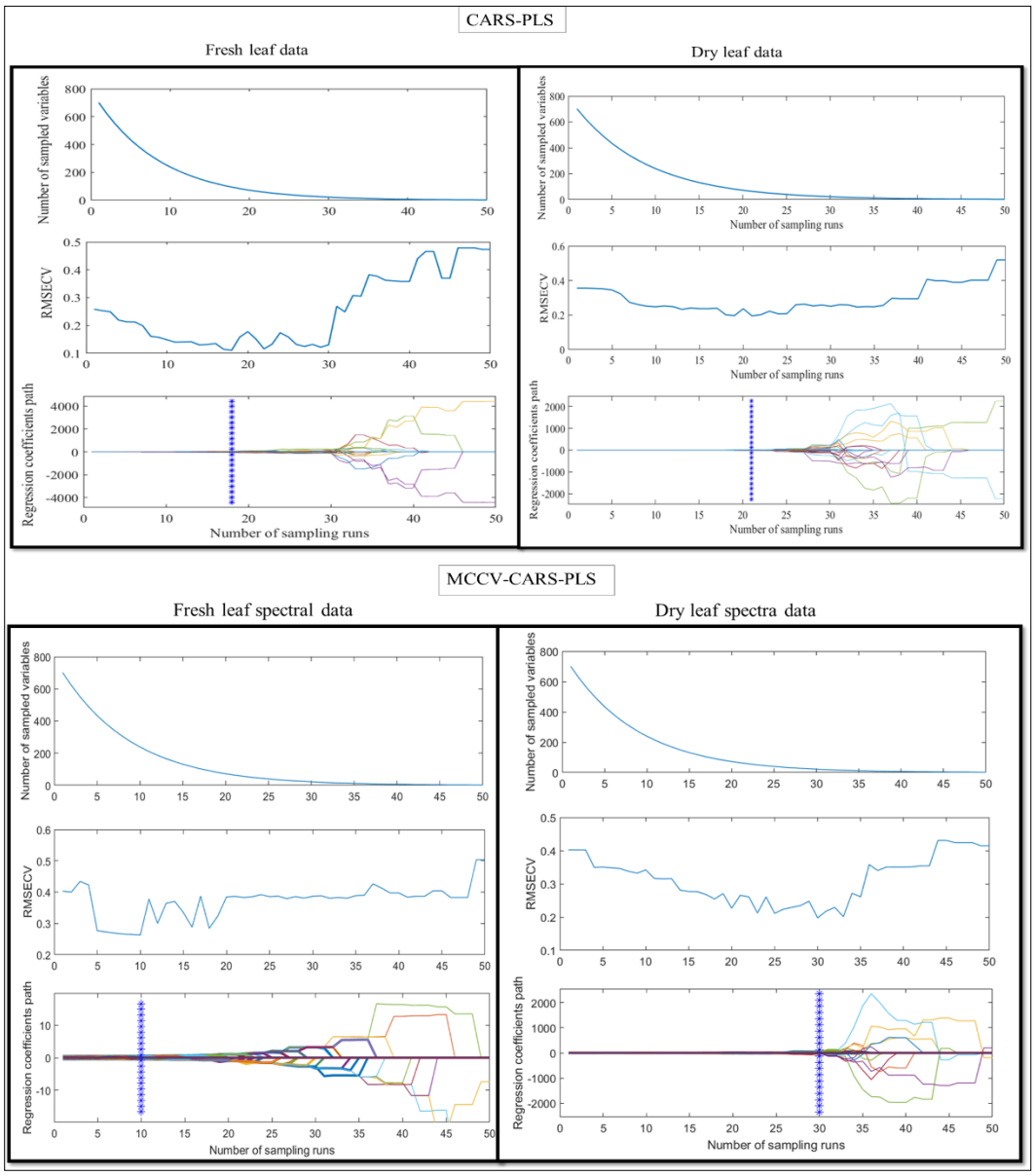

Figure 5. The predictor variables selection result of CARS-PLS and MCCV-CARS-PLS regression analysis: Number of sampled variables present the number of variables as a function of iterations; RMSECV is the ten-fold cross-validated errors at each MC sampling run, and regression coefficients are the coefficient path of each variable in which each line signifies the path of a variable. 5(*a*) and 5(*b*) represent the results of fresh and dry based CARS modeling, respectively. 5(*c*) and 5(*d*) demonstrate the results of fresh and dry based CARS modeling. The vertical dotted line denotes the wavelengths set with the lowest ten-fold.

5.3 The wavelengths selected from a different method

The results presented are based on the SNV spectral transformation, CARS and MCCV-CARS variable selection methods. The SNV indicated high LN concentration prediction when compared to MSC and derivative. Table 1 showed the wavelengths selected by the different method (MCCV-CARS and CARS) for both dry and fresh hyperspectral leaf data. On average 38 and 25 features were selected by MCCV-CARS and CARS, respectively. As shown in Table 1 features selected by CARS are between 2000 - 2300 regions for both dry and fresh leaf spectral, that is, (2041, 2044, 2047, 2107, 2113, 2119, 2167, 2182, 2185, 2188, 2200, 2326, 2380, 2419 nm) and (2044, 2107, 2110, 2179, 2188, 2449 nm) for fresh and dry leaf, respectively. The selection patterns showed by CARS is consistent in both fresh and dry spectral. Conversely, MCCV-CARS selected informative features across the spectrum (Vis-NIR-SWIR) (Table 1). Overall, MCCV-CARS selected wavelengths in the Visible and near-infrared (731-

1400 nm) which are associated with leaf pigments, whereas CARS selected wavelengths which are vital for quantifying nutrients and proteins information of the plants.

5.4 Common spectral bands for predicting leaf nitrogen using dry and fresh leaf reflectance data of E. grandis.

There are common spectral bands for estimating leaf nitrogen from dry and fresh leaf spectral of *E. grandis* Table 1. CARS selected wavelengths between 2000 nm and 2200 nm for all datasets (Table 1). The features selected by CARS irrespective of transformation methods used are related to known LN absorption features (2130 nm, 2180 nm, 2060 nm, 2240 nm, and 2300 nm). MCCV-CARS selected mostly wavelengths in the visible and mid-near-infrared. Also from Table 1, we found that some of the features selected by MCC-CARS are close to those reported in Curran (1989) e.g. 1020 nm, 1510 nm, 1730 nm, 1930 nm. In the present study, SNV-MCCV-CARS selected 1027 nm and 1519 nm which on predicting LN concentration, the PLS model was built on the selected wavelengths and the full spectrum of both dry and fresh spectral.

Table 1. Common wavebands selected for predicting leaf nitrogen content from the spectral reflectance of fresh and dry leaves using different band selection methods: MCCV = Monte Carlo Cross-Validation, CARS = Competitive adaptive re-weighted sampling, NIR = near-infrared, SWIR = shortwave-infrared.

Variable selection methods	VIS	Red edge (680 nm to 730)	NIR (731-1400 nm)	SWIR	
				Water bands	Non-water bands
MSC transformed spectral reflectance					
CARS					2059,2107
MCCV-CARS	-		1012, 1189, 1342 1345,	1519, 1522, 1615, 1642	
SNV **transformed spectral reflectance**					
CARS	-	-	-	-	2044, 2107, 2188
MCCV-CARS	433,451, 669	706 694,717	931, 1003, 1027, 1036, 1177, 1180	-	-
Un-transformed leaf spectra					
MCCV-CARS	406,415, 433,451, 484,517, 586,673	709	778,796,874,931,946,1009,1012,1024,1027,1039,1048,1054,1057,1150, 1159,1168,1174,1177	-	-
1st derivatives transformed spectral reflectance					

CARS	-	-	-	-	2062, 2065, 2068, 2146, 2152, 2155, 2158
MCCV-CARS	433,651, 459		787,1024, 1159, 1180, 1195, 1342,1348		

5.5 . Performance of SMC-PLSCV, CARS-PLS and MCCV-CARS-PLS

Amongst the variable selection methods, common fresh and dry leaf spectral wavelengths selected by MCCV-CARS provided the highest *R*2cal and *R*2P with leaf nitrogen using PLS (Table 2). The dry leaf spectral data estimated LN with higher accuracy (*R*2P =0.92 and RMSEP=0.12) when compared to fresh leaf dataset (*R*2P =0.88 and RMSEP=0.12) (Table 2). Figure 6 showed leaf nitrogen prediction accuracies yielded with the calibration dataset. Overall results show that dry and fresh leaf spectra predict LN concentration of *E. grandis* differently even from wavelengths bands. The high accuracy from CARS and MCCV-CARS selected variables reveals that the methods are helpful for prediction of LN concentration.

Moreover, the results indicate that hyperspectral data, when combined with variable selection strategies, can accurately predict the LN concentration of *E. gradins* leaves without over-fitting.

Table 2. Regression evaluation results from dry and fresh leaf spectra reflectance using CARS, MCCV-CARS and sMC variable selection methods. The results highlight essential variables for the estimation of LN concentration and their prediction accuracies.

Variable Selection Method	Prediction Assessment Parameters			
	Cal-R^2	*P-R^2*	*Cal-RMSE*	*P-RMSE*
Standard Normal Transformation				
Dry leaf spectra				
PLSCV	0.72	0.70	0.33	0.11
PLS-CARS	0.83	0.80	0.26	0.16
PLSMCCV-CARS	0.88	0.84	0.12	0.13
Fresh leaf spectra				
PLSCV	0.72	0.76	0.36	0.11
PLS-CARS	0.72	0.83	0.29	0.13
PLSMCCV-CARS	0.76	0.82	0.34	0.14
Un-transformed leaf spectra				
Dry leaf spectra				
PLSCV	0.72	0.63	0.12	0.24
PLS-CARS	0.81	0.78	0.17	0.26
PLSMCCV-CARS	0.79	0.68	0.11	0.23
Fresh leaf spectra				
PLSCV	0.82	0.35	0.21	0.34
PLS-CARS	0.71	0.64	0.72	0.13
PLSMCCV-CARS	0.98	0.28	1.27	0.36
PLSCV	0.82	0.35	0.21	0.34

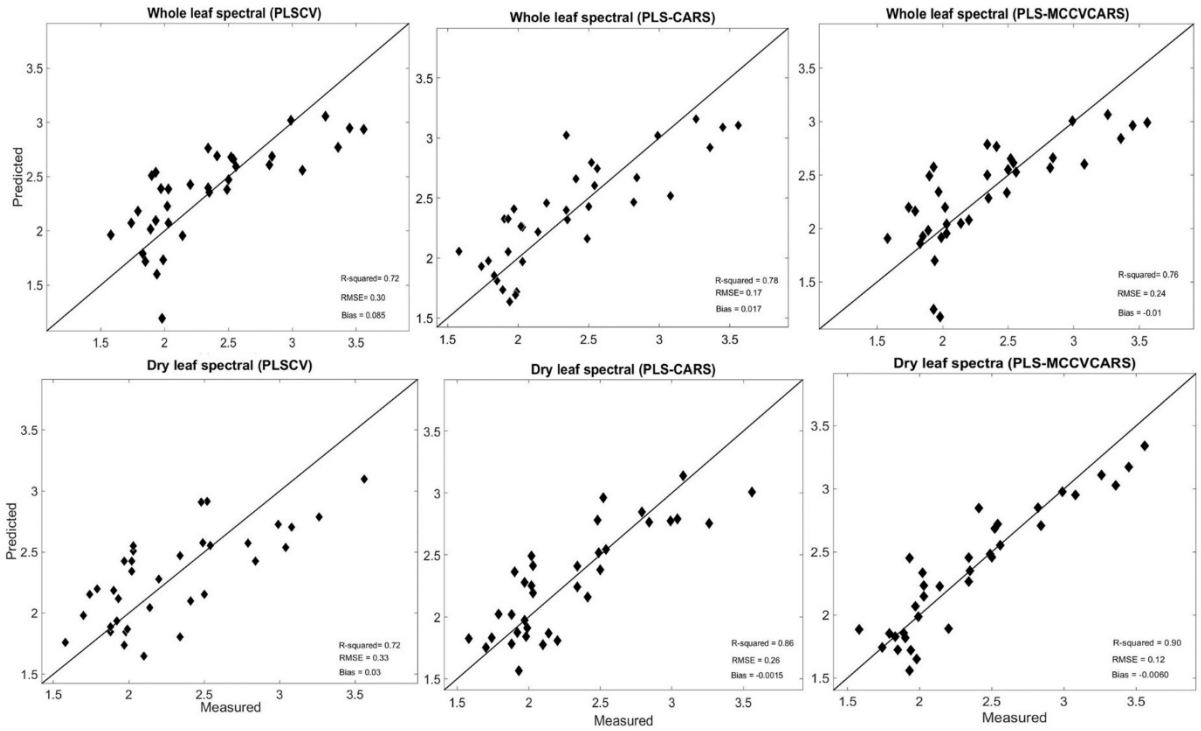

Figure 6. The scatter plot of predicted and measured LN concentration based on SNV transformed leaf spectral data. Dry leaf spectral data (*a*)=SMC-PLS, (*b*)= PLS-CARS and (*c*)= PLS-MCCV-CARS. Fresh leaf spectral data (*d*)=SMC-PLS, (*e*)= PLS-CARS and (*f*)= PLS-MCCV-CARS.

6. Discussion

We sought to determine the common absorption features for quantifying LN concentration from dry and fresh leaf spectra of *E. grandis*. Use of model population analysis (MPA) framework to select the common spectral bands for predicting leaf nitrogen was also explored. The results showed that the MPA framework can be used to select useful bands for predicting LN from the spectral data of fresh and dry leaves. However, the ability of the MPA framework to select optimal spectral bands depends on the variable selection algorithm. MCCV-CARS and CARS-PLSR algorithms showed a higher predictive performance when compared to sMC-PLSR. Although PLSR has been extensively used to reduce dimensionality and multi-collinearity of hyperspectral data, it was not valuable in selecting optimal predictive bands for LN in this study. The performance of PLSR slightly improved when useful bands were selected using the combination of sMC-PLSR. Furthermore, the study indicated that selection of the optimal bands was also enhanced by transformation of the spectral reflectance.

Comparing predictive accuracy: MSC, SNV and SG produced high prediction accuracy when compared to un-transformed data. However, among the transformation methods, SNV predicted LN with the highest accuracy for all the leaf spectral datasets. Strictly following SNV, is the MSC spectral, while un-transformed spectral was the least performing method. The same pattern of performance was also observed in other studies (Huang *et al*. 2004; Mutanga and Skidmore 2004; Rameolo *et al*. 2011) and Mitchell *et al*. 2012). For example,

Huang *et al*. (2004), observed an increase in the coefficient of determination from 0.65 to 0.85 when using standard derivative analysis and CR spectral transformations algorithms, respectively, for estimating foliar LN concentration of *Eucalyptus* tree canopies in open woodland. Likewise, Mitchell *et al*. (2012) observed the higher prediction accuracy of LN from normalised band depth using transformed high-resolution spectrometer data. Qu *et al*. (2016), strongly predicted alcohol

concentrations using SNV transformed NIR spectra as compared with un-transformed spectra. The same prediction pattern reported in Qu *et al*. (2016) study was also observed in the present study.

Both dry and fresh leaf spectral exhibited higher accuracy from SNV transformed data, followed by MSC and SG, while raw data were the least accurate. The slight difference between SNV and MSC observed could be attributed to the fact that MSC and SNV, primarily use the same strategy to transform spectral signals. MSC compensates the influence of non-uniform scattering induced by different particle sizes using the linear correlation between each spectrum and the average spectrum. On the hand, SNV transforms the spectrum by eliminating the deviation caused by the particle size and scattering. During the process, the values of the spectrum from the original spectrum are averaged, subsequently divided by the standard deviation. As observed in this study, Shi *et al*. (2011), also confirmed that the correction capability of SNV is stronger than that of MSC. Therefore, the spectral transformation, particularly of fresh leaf spectral, could be of great use in predicting LN of *E. grandins* forest.

Through the proposed variable selection models (MCCV-CARS and CARS) we were able to ascertain common spectral signals which predicted LN concentration with high accuracy as compared to the conventional PLS regression model. Noticeably, there was a substantial selection of wavelengths at the Vis, red-edge and SWIR using MCCV-CARS, whereas CARS selected mainly wavelengths located at the SWIR region of the spectrum. Most of the wavelengths bands selected with MCCV-CARS are highly correlated with known leafchl concentration absorption features. In particular, the wavelengths selected from both fresh and dry leaf spectral data range close to the well-known absorption spectra of leafchl. In both cases, we also observed a selection of the known LN absorption band (555 nm). The high correlation between leafchl and LN was also expressed with the selection of wavelengths near 705 nm. This band is within the red-edge region of the spectrum which has been reported to be significant for estimating leafchl. In general, the selected wavelengths in Vis

and red-edge spectrum regions could be attributed to the high correlation between LN and leafchl concentration.

The prediction accuracy of regression model produced from Vis and red-edge features demonstrate the importance of absorption features of leafchl located at the Vis and red-edge regions for LN detection contents with both fresh and dry leaf spectral data. In contrast to fresh leaf spectral, both variables selection models selected more wavelengths at the SWIR region with dry leaf spectra data. Moreover, most of the selected wavelengths bands are either known absorption wavelength regions or within 10 nm of known absorption features. This validates the impact of SWIR located water absorption features to the LN estimations as indicated in Ramoelo *et al*. (2012). Most of the absorptions found in the SWIR are characterised by the presence of LN, lignin, and cellulose and leaf water. However, the presence of water in the leaf tends to mask those small peaks of the LN because as mentioned before, LN it is distributed into various organelles of leaf, and only a small amount occurs as free constituents (Wang *et al*. 2017).

There was an inconsistency between CARS and MCCV-CARS regarding feature selection. CARS consistently selected bands in the SWIR region (2044, 2107, and 2188) for both dry and fresh leaf spectral. Interestingly, most of the features selected by both approaches are comparable to those illustrated in Curran (1989) as absorption features of LN (2060nm, 2180 nm, 2240 nm, and 2300 nm) and Kokaly and Clark (1999) study (2054 and 2172 nm). Similar to Mitchell *et al*. (2012), most of the wavelengths selected with CARS occurred within 20 nm of known LN absorption features located in the NIR and SWIR spectrum regions (1016 nm, 2043 nm, 2062 nm, 2117 nm, 2293 nm, 2360 nm, and 2376 nm). However, we observed that most of the bands selected by MCCV-CARS are located in the NIR regions (931 nm, 1003 nm, 1027 nm, 1036 nm, 1177 nm, 1180 nm). It should be noted that most

of the NIR selected bands have not been reported as LN sensitive bands. However, they indicated to the reliable for predicting LN contents. The bands were able to predict LN with high accuracy.

Comparing the performance of fresh and dry leaf spectral using identified common wavelengths, dried spectral predicted LN with high accuracy. The high prediction accuracy of LN content shown by dried leaf spectra is in agreement with the results observed in Mitchell *et al*. (2012). Their study observed high accuracy from sagebrush dried leaf spectra than from fresh leaf spectral. Regarding prediction accuracy of the LN, coefficient of determination (*R*2) increased from dried leaf spectral in PLS calibration. These results suggest that the spectra of

fresh leaves are primarily influenced by leaf water content and structural aspects including the inter-cellular spaces within the spongy mesophyll. These internal leaf components are known to perturb factors to LN estimation. However, they appear to have a significant impact on the LN prediction accuracy as illustrated in dried leaf results. The high accuracy from dried leaf spectral is inconsistent with the finding of Ramoelo *et al.* (2012). The study demonstrated that by removing water from the leaf enhanced N estimation of grass in the savanna landscape. The identified spectral features could be relevant for assessing LN concentration for different seasons, such as wet to dry season.

7. Conclusion

The outcome of the present study showed that: There are common spectral bands for estimating leaf nitrogen from dry and fresh leaf spectral data. Estimation of leaf nitrogen from dry leaf reflectance is higher than with fresh leaf spectral data. The prediction performance of the PLSR models improves when linked with variable selection methods. The combination of MCCV and CARS for variable selection enhances the performance of the prediction models for leaf nitrogen estimation.

Acknowledgements

The authors would like to thank Mondi Business Paper and the Council for Scientific and Industrial Research for respectively providing logistical and funding support for this project. They would also like to thank Dr Mark Norris-Rogers and Mr Marius du Plessis (Mondi BP) for their assistance during the field data collection. Also, we will like to acknowledge Matlab codes CARS and MCCV-CARS (copyright Hongdong Li, 2011); http://code.google.com/p/carspls.

References

Afanador, N.L., Tran, T.N., Blanchet, L. and Buydens, L.M.C., 2014. Variable importance in PLS in the presence of autocorrelated data—Case studies in manufacturing processes. *Chemometrics and Intelligent Laboratory Systems*, *139*, pp.139-145.

Barnes, R., M. Dhanoa, and S. Lister, 1993, Correction to the description of standard normal variate (SNV) and de-trend (DT) transformations in Practical Spectroscopy with Applications in Food and Beverage Analysis—2nd edition. *Journal of Near Infrared Spectroscopy,* 1, pp. 185-186.

Blackburn, G.A., 2007, Hyperspectral remote sensing of plant pigments. *Journal of experimental botany,* 58, pp. 855-867.

Cai, W., Y. Li, and X. Shao, 2008, A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems,* 90, pp. 188-194.

Centner, V., D. Massart, O.E. de Noord, S. de Jong, B.M. Vandeginste, and C. Sterna, 1996, Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry,* 68, pp. 3851-3858.

Cho, M.A., A. Skidmore, F. Corsi, S.E. Van Wieren, and I. Sobhan, 2007, Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation,* 9, pp. 414-424.

Cho, M.A., J. Van Aardt, R. Main, and B. Majeke, 2010, Evaluating variations of physiology-based hyperspectral features along with a soil water gradient in a *Eucalyptus grandis* plantation. *International Journal of Remote Sensing,* 31, pp. 3143-3159.

Curran, P.J., 1989, Remote sensing of foliar chemistry. *Remote Sensing of Environment,* 30, pp. 271-278.

Deng, B., Y. Yun, P. Ma, C. Lin, D. Ren, and Y. Liang, 2015, A new method for wavelength interval selection that intelligently optimises the locations, widths and combinations of the intervals. *Analyst,* 140, pp. 1876-1885.

Du, L., W. Gong, S. Shi, J. Yang, J. Sun, B. Zhu, and S. Song, 2016, Estimation of rice leaf N contents based on hyperspectral LIDAR. *International Journal of Applied Earth Observation and Geoinformation,* 44, pp. 136-143.

Fan, W., Y. Shan, G. Li, H. Lv, H. Li, and Y. Liang, 2012, Application of competitive adaptive reweighted sampling method to determine useful wavelengths for prediction of total acid of vinegar. *Food Analytical Methods,* 5, pp. 585-590.

Ferwerda, J.G., and S.D. Jones, 2006, Continuous wavelet transformations for hyperspectral feature detection. In *Progress in spatial data handling,* Anonymous pp. 167-178 Springer, 2006).

Fourty, T., F. Baret, S. Jacquemoud, G. Schmuck, and J. Verdebout, 1996, Leaf optical properties with an explicit description of its biochemical composition: direct and inverse problems. *Remote Sensing of Environment,* 56, pp. 104-117.

Gökkaya, K., V. Thomas, T. Noland, H. McCaughey, I. Morrison, and P. Treitz, 2015, Mapping continuous forest type variation by means of correlating remotely sensed metrics to canopy N: P ratio in a boreal mixedwood forest. *Applied vegetation science,* 18, pp. 143-157.

Han, Q., H. Wu, C. Cai, L. Xu, and R. Yu, 2008, An ensemble of Monte Carlo uninformative variable elimination for wavelength selection. *Analytica Chimica Acta,* 612, pp. 121-125.

Huang, C., and G.P. Asner, 2009, Applications of remote sensing to alien invasive plant studies. *Sensors,* 9, pp. 4869-4889.

Huang, Z., B.J. Turner, S.J. Dury, I.R. Wallis, and W.J. Foley, 2004, Estimating foliage N concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment,* 93, pp. 18-29.

Ilari, J., H. Martens, and T. Isaksson, 1988, Determination of particle size in powders by scatter correction in diffuse near-infrared reflectance. *Applied Spectroscopy,* 42, pp. 722-728.

Jiang, J., R.J. Berry, H.W. Siesler, and Y. Ozaki, 2002, Wavelength interval selection in the multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Analytical Chemistry,* 74, pp. 3555-3565.

Kokaly, R.F., and R.N. Clark, 1999, Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment,* 67, pp. 267-287.

Li, H., Y. Liang, Q. Xu, and D. Cao, 2009, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta,* 648, pp. 77-84.

Martens, H., and M. Martens, 2000, Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food quality and preference,* 11, pp. 5-16.

Mehmood, T., K.H. Liland, L. Snipen, and S. Sæbø, 2012, A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems,* 118, pp. 62-69.

NHruska, 2012, Spectroscopic detection of N concentrations in sagebrush. *Remote sensing letters,* 3, pp. 285-294.

Muñoz-Huerta, R.F., R.G. Guevara-Gonzalez, L.M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez, and R.V. Ocampo-Velazquez, 2013, A review of methods for sensing the N status in plants: advantages, disadvantages and recent advances. *Sensors,* 13, pp. 10823-10843.

Mutanga, O., and A.K. Skidmore, 2004, Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *International journal of applied earth observation and Geoinformation,* 5, pp. 87-96.

Mutanga, O., A.K. Skidmore, and S. van Wieren, 2003, Discriminating tropical grass (Cenchrus ciliaris) canopies grown under different N treatments using spectroradiometry. *ISPRS Journal of Photogrammetry and Remote Sensing,* 57, pp. 263-272.

Naes, T., T. Isaksson, and B. Kowalski, 1990, Locally weighted regression and scatter correction for near-infrared reflectance data. *Analytical Chemistry,* 62, pp. 664-673.

NøRgaard, M., N.K. Poulsen, and O. Ravn, 2000, New developments in state estimation for nonlinear systems. *Automatica,* 36, pp. 1627-1638.

Pellissier, P.A., S.V. Ollinger, L.C. Lepine, M.W. Palace and W.H. McDowell, 2015, Remote sensing of foliar N in cultivated grasslands of human-dominated landscapes. *Remote Sensing of Environment*, 167, pp. 88-97.

Pereira, H.M., S. Ferrier, M. Walters, G.N. Geller, R.H. Jongman, R.J. Scholes, M.W. Bruford, N. Brummitt, S.H. Butchart, A.C. Cardoso, N.C. Coops, E. Dulloo, D.P. Faith, J. Freyhof, R.D. Gregory, C. Heip, R. Hoft, G. Hurtt, W. Jetz, D.S. Karp, M.A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J.P. Scharlemann, S.N. Stuart, E. Turak, M. Walpole, and M. Wegmann, 2013, Ecology. Essential biodiversity variables. *Science (New York, N.Y.),* 339, pp. 277-278.

Peterson, E., 2005, Estimating cover of an invasive grass (Bromus tectorum) using tobit regression and phenology derived from two dates of Landsat ETM data. *International Journal of Remote Sensing,* 26, pp. 2491-2507.

Qu, F., D. Ren, J. Wang, Z. Zhang, N. Lu, and L. Meng, 2016, An Ensemble Successive Project Algorithm for Liquor Detection Using Near Infrared Sensor. *Sensors,* 16, pp. 89.

Ramoelo, A., A.K. Skidmore, M. Schlerf, R. Mathieu, and I.M. Heitkönig, 2011, Water-removed spectra increase the retrieval accuracy when estimating savanna grass N and phosphorus concentrations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, pp. 408-417.

Schlerf, M., C. Atzberger, J. Hill, H. Buddenbaum, W. Werner, and G. Schüler, 2010, Retrieval ofChl and N in Norway spruce (Picea abies L. Karst.) using imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation,* 12, pp. 17-26.

Serrano, L., J. Penuelas, and S.L. Ustin, 2002, Remote sensing of N and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment,* 81, pp. 355-364.

Shi, B., L. Zhao, H. Wang, and D. Zhu, 2011, Signal optimisation approaches on the prediction of apples firmness by near-infrared spectroscopy. *Sensor Letters,* 9, pp. 1062-1068.

Singh, A., S.P. Serbin, B.E. McNeil, C.C. Kingdon, and P.A. Townsend, 2015, Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecological Applications,* 25, pp. 2180-2197.

Skidmore, A.K., J.G. Ferwerda, O. Mutanga, S.E. Van Wieren, M. Peel, R.C. Grant, H.H. Prins, F.B. Balcik, and V. Venus, 2010, Forage quality of savannas—Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment,* 114, pp. 64-72.

Tran, T.N., N.L. Afanador, L.M. Buydens, and L. Blanchet, 2014, Interpretation of variable importance in partial least squares with significance multivariate correlation (sMC). *Chemometrics and Intelligent Laboratory Systems,* 138, pp. 153-160.

Vohland, M., M. Ludwig, M. Harbich, C. Emmerling, and S. Thiele-Bruhn, 2016, Using variable selection and wavelets to exploit the full potential of Visible–near-infrared spectra for predicting soil properties. *Journal of Near Infrared Spectroscopy,* 24, pp. 255-269.

Vorovencii, I., 2009, The hyperspectral sensors used in satellite and aerial remote sensing. *Bulletin of the Transilvania University of Braşov• Vol,* 2, pp. 51.

Wang, Z., 2016, Mapping spatial variation of foliar N using hyperspectral remote sensing.

Wang, B.J., Chen, J.M., Ju, W., Qiu, F., Zhang, Q., Fang, M. and Chen, F., 2017. Limited effects of water absorption on reducing the accuracy of leaf N estimation. Remote Sensing, 9(3), p.291.

Wessman, C.A., J.D. Aber, D.L. Peterson, and J.M. Melillo, 1988, Remote sensing of canopy chemistry and N cycling in temperate forest ecosystems.

Wise, B.M., 2004, *Properties of partial least squares (PLS) regression, and differences between algorithms,*

Wold, H., Estimation of principal components and related models by iterative least squares. Multivariate Analysis. Edited by: Krishnaiaah PR. 1966.

Wold, S., M. Sjöström, and L. Eriksson, 2001, PLS-regression: a primary tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems,* 58, pp. 109-130.

Xu, Q., and Y. Liang, 2001, Monte Carlo cross-validation. *Chemometrics and Intelligent Laboratory Systems,* 56, pp. 1-11.

Yun, Y., W. Wang, M. Tan, Y. Liang, H. Li, D. Cao, H. Lu, and Q. Xu, 2014, A strategy that iteratively retains informative variables for selecting an optimal variable subset in multivariate calibration. *Analytica Chimica Acta,* 807, pp. 36-43.