Detecting nutrient deficiencies in *Eucalyptus grandis* trees using hyperspectral remote sensing and random forest

Leeth Singh[1*], Onisimo Mutanga[1], Paramu Mafongoya[1], Kabir Peerbhay[1,2], Steven Dovey[2]

[1]University of KwaZulu-Natal, Agricultural, Earth and Environmental Science, Department of Geography, P. O. Box X01, Scottsville, Pietermaritzburg, South Africa, 3209.
[2]Sappi Forests, Shaw Research Centre, P. O. Box 473, Howick, South Africa, 3290.
*singhleeth@gmail.com

1. Abstract

Nutrient deficiencies in commercial forest trees often lead to stunted growth and reduced chances of field survival, resulting in a loss of time, productivity, and trees that can become more susceptible to a host of infections. While conventional foliar analytical methods provide accurate results, they are not time and cost-effective in a high productivity environment. This study aims to test the capability of remote sensing to detect macronutrient and micronutrient deficiencies rapidly in juvenile trees. We acquired full-waveform hyperspectral data (350nm-2500nm) from 135 young trees planted in individual pots in a controlled forestry nursery environment. We quantified nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), sodium (Na), manganese (Mn), iron (Fe), copper (Cu), zinc (Zn), and boron (B) in young commercially planted forest variety. This study identified the most critical wavebands for detecting nutrient deficiencies using built-in random forest (RF) measures of variable importance. The random forest algorithm's robustness significantly reduced the dataset's noise whilst producing promising results for certain macronutrients such as P and N (0.95 and 0.89, respectively) and micronutrients such as Mn and Cu (0.90 and 0.86, respectively). We identified the red-edge, near-infrared (NIR), visible and short-wave infrared-2 (SWIR-2) regions of the electromagnetic spectrum as the most effective regions for detecting macronutrients and micronutrients in this study. We recommend testing the use of strategic portions of the electromagnetic spectrum for reducing noise and enabling faster computing time, such as portable near-infrared technology.

2. Introduction

Nutrient deficient trees present a challenge to the commercial forestry sector. Underproductive trees put pressure on economic systems and nutrient cycles. The adequate supply of macronutrients and micronutrients play a crucial role in supporting plant development, positively influencing forest productivity (Silva and Uchida, 2000). Nitrogen (N), phosphorous (P), calcium (K), iron (Fe), boron (B) and zinc (Zn) are essential macronutrients for several plant physiological functions such as photosynthesis, enzymatic reactions, respiration, ribonucleic acid (RNA) formation, tryptophan synthesis, maintaining genetic information, root development, stomatal regulation, protection against oxidative damage and creating amino acids (Silva and Uchida, 2000). However, macronutrient and

micronutrient deficiencies often result in stunted growth, chlorosis, reduction in protein content, weak stem production and can cause early maturity in some plants (Silva and Uchida, 2000).

Traditional scientists obtain foliar nutrient information using destructive sampling methods such as wet chemistry analysis, which involve ground-based periodic surveys and tedious laboratory work that is costly and time-consuming (Pullanagari *et al.*, 2016). However, researchers have made little progress using indirect spectral methods (Oliveira *et al.*, 2017). Hyperspectral remote sensing offers a rapid, non-destructive and effective approach for detecting key nutrient levels in forest trees. Hyperspectral data can provide benefits in high productivity environments, such as in younger plants in forest nurseries. The detection of foliar nutrients occurs through specific absorption features within the electromagnetic spectrum. Earlier research explains the physiological link between foliar nutrient content and remote sensing, e.g., (Dixit and Ram, 1985; Curran, 1989; Elvidge, 1990). More specifically, hyperspectral sensors (350nm–2500nm) capture detailed spectral information; however, they are often sensitive to the influence of spectral noise, which negatively impacts classification approaches (Agjee *et al.*, 2018). The presence of spectral noise can significantly impact the quality of the data acquired; hence, classification approaches' performance will deteriorate (Agjee *et al.*, 2018).

For example, Oliveira *et al.* (2017) successfully estimated the N content of 25-month *Eucalyptus* trees and compared a wide range of variable importance (VI) results. As a result, the authors obtained the best R^2's of 0.97 using inflection point position (IPP), normalized difference red-edge (reNDVI) and modified red-edge normalized difference vegetation index (mNDI) in the 400–900nm range. A later study by Oliveira and Santana (2020) estimated the full range of macronutrients and micronutrients: N, P, K, S, Ca, Mg, Mn, B, Zn, Cu, and Fe in *Eucalyptus* clones using the NIR region (400-900nm) and PLSR. As a result, the authors predicted all nutrients using the coefficient of determination of cross-validation (R_{CV}^2), with the lowest and highest estimate was Mg (0.22 R_{CV}^2) and N (0.95 R_{CV}^2), respectively. The authors found that PLSR and variable selection methods increased the accuracy of nutrient concentration estimates and suggest future studies use wavelength ranges above 900nm. Osco *et al.* (2020) tested machine learning algorithms: k-nearest neighbour (kNN), lasso regression, ridge regression, support vector machine (SVM), neural network (ANN), decision tree (DT), and random forest (RF) using a proximal hyperspectral sensor (380nm-1020nm) to predict nutrient content on a Valencia-orange orchard. The authors obtained high predictions (R^2) above 0.73 for all algorithms and found that RF was the most suitable algorithm.

In summary, many studies have predicted nutrient concentrations using hyperspectral data and a vast array of computational algorithms (Ferwerda *et al.*, 2005; Abdel-Rahman *et al.*, 2017; Wang *et al.*, 2018; Oliveira and Santana, 2020). However, few studies predicted macronutrient and micronutrients deficiency using a full-waveform hyperspectral proximal sensor and RF. Therefore, this study aimed to predict macronutrients and micronutrients in *Eucalyptus* hybrid trees using full-waveform hyperspectral data (350nm-2500nm) and the RF algorithm. Furthermore, to our knowledge, no studies have identified the most critical wavebands for detecting nutrient deficiencies in younger trees within a nursery setting. The outcomes of this study will promote the use of remote

sensing scanning systems for rapid diagnosis of macronutrient and micronutrient deficiencies in high productivity commercial forestry environments.

3. Material and Methods

3.1. Study area

This research experiment was conducted under a controlled nursery environment at the Institute of Commercial Forestry Research (ICFR) nursery in Pietermaritzburg, KwaZulu Natal, South Africa (29°37'40.20"S and 30°24'13.63"E). In this study, we examined a *Eucalyptus* hybrid (*Eucalyptus grandis x Eucalyptus urophylla*). The *Eucalyptus* genus is a hardwood perennial native to Australia (Myburg *et al.*, 2014). Commercial forestry industries commonly grow *Eucalyptus* trees for their fast growth and superior wood properties (Myburg *et al.*, 2014). Hence, more than 100 countries across six continents (>20 million ha) grow *Eucalyptus* trees as a timber resource (Myburg *et al.*, 2014). The hybrid species *Eucalyptus grandis x Eucalyptus urophylla* used in this study are native to Newcastle, New South Wales to Bundaberg in Queensland and the Indonesian Archipelago Timor, respectively (Pinto *et al.*, 2014; Pajares, 2015). The shape of hybrid *Eucalyptus grandis x Eucalyptus urophylla* leaves was lanceolate with the adaxial side dark green and the abaxial slightly paler than the adaxial side.

3.2. Experimental design

We conducted a pot trial experiment to develop more explicit diagnostic indicators and measures of changes in soil nutrient status. We planted 135 hybrid seeds *Eucalyptus grandis x Eucalyptus urophylla* obtained from a commercial plantation seed orchard in KwaZulu Natal, Midlands, South Africa, in June 2014. To minimize the effect of a microclimate, we randomly arranged pots in the designated nursery environment. The pots were under an open-sided plastic cover to exclude rainfall and kept under natural sunlight. This study's soil type was Inanda soil, the predominant soil type in the Midlands, South Africa (Mucina and Rutherford, 2006). The soil texture was silty clay (56% sand & silt; 44% clay). Distilled water was added automatically via drip irrigation to maintain optimal soil moisture conditions. Drippers had an output rate of 2.2L per hour programmed to water twice a day for three minutes. Fertilizer was added once per week over four weeks and then left to acclimatize for another four weeks (Table 2). The canopy characteristics of the leaf material sampled in this study were at the sapling stage (Juvenile) of growth with leaf area was 6cm to 10cm long and 2cm to 3cm wide. The saplings grew to a height of 30cm to 60cm with a canopy width of approximately 30cm to 40cm. The root characteristics of the saplings had an elongated rooting structure. The hybrid *Eucalyptus* is designed with an extensive tap-root rooting system to anchor the trees and horizontal roots that keep the trees upright when planted in the field (Dye, 1996).

A

B

Figure 1: (A*) Eucalyptus grandis x Eucalyptus urophylla* trees planted in pots and (B) an adaxial leaf representation.

The experiment continued until each nutrient reached its depletion threshold determined through foliar and growth diagnosis (Table 1). We determined depletion thresholds suggested by Reuter and Robinson (1997). This strategy enabled calibration of extractable soil nutrient levels with tree growth and foliar nutrient diagnostics to enable improved laboratory soil data interpretation (Table 2). All samples were dried, weighed and analysed for physio-chemical properties and expressed by leaf concentration (%/dry weight or mg g-1)(Knox *et al*., 2011).

Table 2: Fertiliser compounds added to exclude specific nutrients from the pot experiment treatments.

Nutrient	Fertiliser compound added	Compound formula
N	Urea	CH_4N_2O
P	Sodium Dihydrogen Orthophosphate dihydrate	$NaH_2PO_4.H_2O$
H	Potassium Chloride	KCl
Ca	Calcium Chloride	$CaCl_2.2H_2O$
Mg	Magnesium Chloride hexahydrate	$MgCl_2.6H_2O$
Na	Sodium Sulphate anhydrous	Na_2SO_4
Micro-nutrients	Micro-Nutrient Mix	Zn, Cu, Fe, B, Mn, Mo
	Zinc Chloride	$ZnCl_2$

Table 1: Pot trial experimental design of the study.

Treatments	Nutrient
Exclusion of macronutrients	-N, -P, -K, -Ca, -Mg, -Na
Exclusion of micronutrients	-Mn, -Fe, -Cu, -Zn, -B (combined)
Control 1	No fertilizer
Control 2	Full Fertilizer

3.3. Spectral measurements

Spectral reflectance measurements were taken on the 2nd of February 2017, using a handheld field analytical spectrometer device (ASD) (FieldSpec® three spectrometers) synchronously with foliar sampling. The ASD measures at a sampling interval range of 1.4nm for 350nm-1000nm and 2nm for 1000nm-2500nm. Reflectance measurements were taken within 1m above the pot, using the fibre optic cable set at 25 degrees field of view (FOV), pointed at the nadir position. We used a white reference panel coated with a barium sulphate of known reflectivity to calibrate the sensor every ten minutes (Spectralon Labsphere, Inc., Sutton, New Hampshire). We acquired ten measurements of each plant (per pot) to derive the representative reflectance spectra for each pot for a total of 135 pots. The ASD instrument operator was positioned as far away from the area under observation to reduce interferences caused by anthropogenic shadow and reflection (Figure 2). We placed each pot in direct sunlight on a cloudless day between 10:00h and 14:00h Central African Time.

Furthermore, pots were placed on a stable black platform in an open area to minimize bidirectional reflectance distribution function (BRDF) effects and background scattering. After averaging the spectra, all spectra were converted from radiance to reflectance using ViewSpec Pro software (ASD Inc., Boulder, Colorado, version 6.0.11). All spectra were radiometric and atmospheric corrected to reduce noise using the Natural Environment Research Council (NERC) field of spectroscopy templates (NERC, Undated).
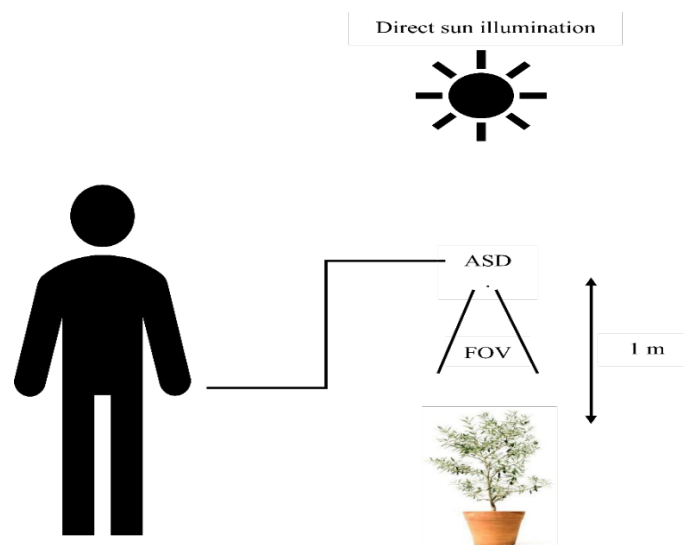

Figure 2: Schematic drawing of ASD measurements. The ASD has pointed approximately 1 meter above. *FOV: Field of view.

3.4. Reference data t-test

In this study, we measured foliar chemistry before and after nutrient depletion. We used nutrient depletion thresholds defined in Reuter and Robinson (1997). We used the foliar chemistry results to test the significant difference in means ($\bar{x}$) of foliar macronutrient and micronutrient at both high and low levels of nutrient content using a paired τ-test for nutrient deficient plants (**Equation 1**). As a

result, the paired τ-test provides a score of the significance (ρ-value) by calculating the difference in the $\bar{x}$ of two groups data (Ruxton, 2006). We used the ρ-value to calculate the statistical hypothesis test results, assuming the null hypothesis is correct (Meng, 1994). If the ρ-value is $<= 0.05$, the two datasets are significantly different, and if the ρ-value is > 0.05, the result is insignificant; hence the trees did not deplete the targeted nutrient. We calculated the τ-test using the following formula metrics:

$$t = \frac{m}{s/\sqrt{n}} \qquad\qquad\qquad [1]$$

Where τ is the paired τ-test, m and s are the mean and the standard deviation of the difference in samples, respectively. n is the size of the sample.

3.5. Random forest

This study used RF for regression based on the classification and regression trees (CARTs) to predict (Breiman, 2001). We implemented the RF ensemble using the "randomForest" package R in R statistical software (R Development Core Team). Compared to several other machine learning algorithms, the RF algorithm delivers the most consistent results, especially when using high dimensional hyperspectral data for predicting foliar nutrient data (Amirruddin *et al.*, 2020). The RF algorithm produces decision trees by drawing a subset of training samples through a replacement method known as "bagging". The bagging process refers to selecting the same sample several times while the rest of the samples can remain unselected. During the training process, approximately two-thirds of the samples from the training set are used as "in-bag" samples, while the remaining one third "out-of-bag" samples are used as an internal cross-validation technique to test the RF models performance (Belgiu and Drăguţ, 2016). The error produced through the cross-validation technique is known as the out-of-bag (OOB) error. The RF algorithm splits each node by a user-defined parameter called the m_{try} function whilst each decision tree is independently produced without pruning (Breiman, 2001). Another user-defined parameter is the n_{tree} which grows the forest whilst the algorithm creates trees that have high variance and low bias (Breiman, 2001). The model grows trees until the final classification is taken by the average class assignment probabilities using the arithmetic mean. The RF algorithm evaluates the model using the final classification and produces a new unlabelled data input against all the decision trees, and each tree votes for class membership. Here, the membership class that receives the maximum votes is the final model selected (Belgiu and Drăguţ, 2016). We parametrised the model using the m_{try} and n_{tree} functions. RF recognizes that classification accuracy is more sensitive to the m_{try} than the n_{tree} function. R sets the default n_{tree} value at 500, whereby errors stabilise as the R package grows trees (Belgiu and Drăguţ, 2016). The n_{tree} value can be optimised to find the best detection accuracies and lowest error rates (Mutanga *et al.*, 2012). After numerous runs through the RF algorithm, an n_{tree} value of 750 provided more accurate results with the data used in this study. We set the m_{try} and n_{tree} functions to 46 and 750, respectively.

3.6. Variable importance

Determining the most critical variables is essential for improving model optimisation, simplification and robustness, especially when dealing with high dimensional data in this study (Liaw and Wiener, 2002). RF has three measures of variable importance. The first measure is based on the number of times a candidate variable is selected in the model. The second measure of importance is based on the Gini impurity when a variable is chosen to split a node as proposed by Breiman (2017). Finally, the third measure is the permutation of a variable as an ensemble of variable importance (Breiman, 2001). In this study, we selected the third measure (permutation of variables) of variable importance, and we used the mean squared error (MSE) in percentage for determining variable importance. Variable importance was determined for each nutrient to generate a coherent account of the relevant variables used in the prediction models for nutrient deficiency.

3.7. Accuracy assessment

We split the final dataset into training (70%) and test (30%) (Breiman, 2017). We used the coefficient of determination (R^2) for prediction and RMSE to assess the RF algorithm's performance in determining nutrient deficiency. We calculated the coefficient of determination using predicted and observed values, values closer to one, predict better results. We calculated the RMSE using predicted and observed values, values closer to zero predict better results than values further from zero. Higher R^2 and lower RMSE values indicate a reliable model.

4. Results

4.1. Descriptive statistics and reference t-test

Table 3 summarises descriptive statistics related to *E. grandis x E. urophylla* foliar macronutrient and micronutrient at low and high levels. We took a random subset (10 pots) for each nutrient on the repetitive measured pots before and after inducing deficiency to test if nutrient deficiency had occurred. The paired τ-test revealed that there was a significant difference ($\rho \leq 0.05$) in $\bar{x}$ whereby the average $\rho - value = 0.0067$.

Table 3: Descriptive statistics of *E. grandis x E. urophylla* foliar macronutrient and micronutrient content.

Nutrient	High				Low			
	Min	**Max**	**Median**	**SD**	**Min**	**Max**	**Median**	**SD**
Macronutrients (%)								
N	0.59	3.67	2.13	0.94	0.83	3.16	1.995	0.50
P	0.002	0.41	0.206	0.08	0.015	0.50	0.2575	0.06
K	0.02	2.45	1.235	0.39	0.26	1.02	0.64	0.11
Ca	0.31	2.88	1.595	0.49	0.39	1.61	1	0.25
Mg	0.003	0.88	0.441	0.11	0.21	0.59	0.4	0.08
Na	0.08	0.32	0.2	0.06	0.05	0.21	0.13	0.05
Micronutrients (ppm)								
Mn	29.96	5 137.42	2583.7	968.84	837.85	8 417.55	4627.7	1 620.46
Fe	23.11	669.57	346.34	69.98	43.04	214.85	128.95	32.92
Cu	3.42	10.88	7.15	1.44	5.51	17.77	11.64	2.57
Zn	3.44	47.17	25.305	9.24	13.19	64.64	38.915	10.73
B	23.79	166.49	95.14	39.65	5.70	98.53	52.115	22.15

4.2.   Detecting macronutrient N, P, K, Ca, Mg, Na, using RF

An important step was to determine the nutrient content of each tree. Furthermore, it was essential to test trees prediction accuracy with combined low and high nutrient concentrations. The RF model ($n_{tree} = 750$ and $m_{try} = 46$) successfully predicted macronutrients of trees with low and high levels of nutrient concentration. Figure 3 shows predicted versus observed distribution plots for all macronutrients. As a result, the RF algorithm successfully predicted macronutrients N, P, K, Ca, Mg, and Na with R^2's of 0.88, 0.95, 0.73, 0.88, 0.70, 0.75 and RMSE's of 0.15, 0.13, 0.14, 0.10, 0.06, 0.04, respectively (Figure 3).

Nitrogen

Phosphorous
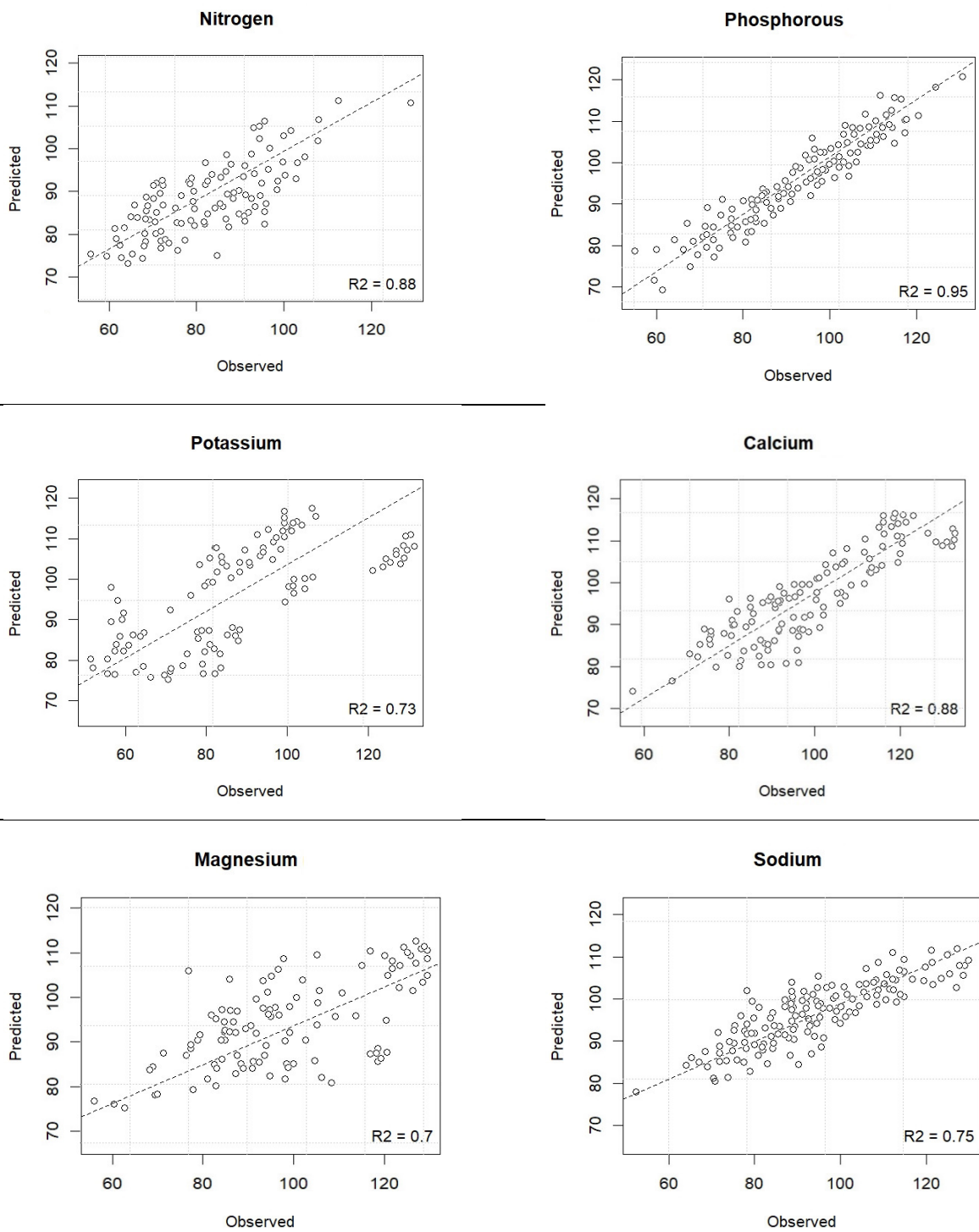
Potassium

Calcium

Magnesium

Sodium

Figure 3: The one-to-one relationship between predicted versus observed macronutrients: N, P, K, Ca, Mg, Na using RF.

4.3. Detecting micronutrients Fe, Mn, Cu, Zn, B using RF

We used the same n_{tree} and m_{try} values for predicting micronutrients: Fe, Mn, Cu, Zn, B in the RF prediction model. Figure 4 shows predicted versus observed distribution plots for all micronutrients. As a result, the RF algorithm successfully predicted low and high levels of

micronutrients concentrations Fe, Mn, Cu, Zn, and B with R^2's of 0.79, 0.90, 0.86, 0.69, 0.66 and RMSE's of 11.86, 745.01, 1.70, 15.09, 2.54, respectively (Figure 4).
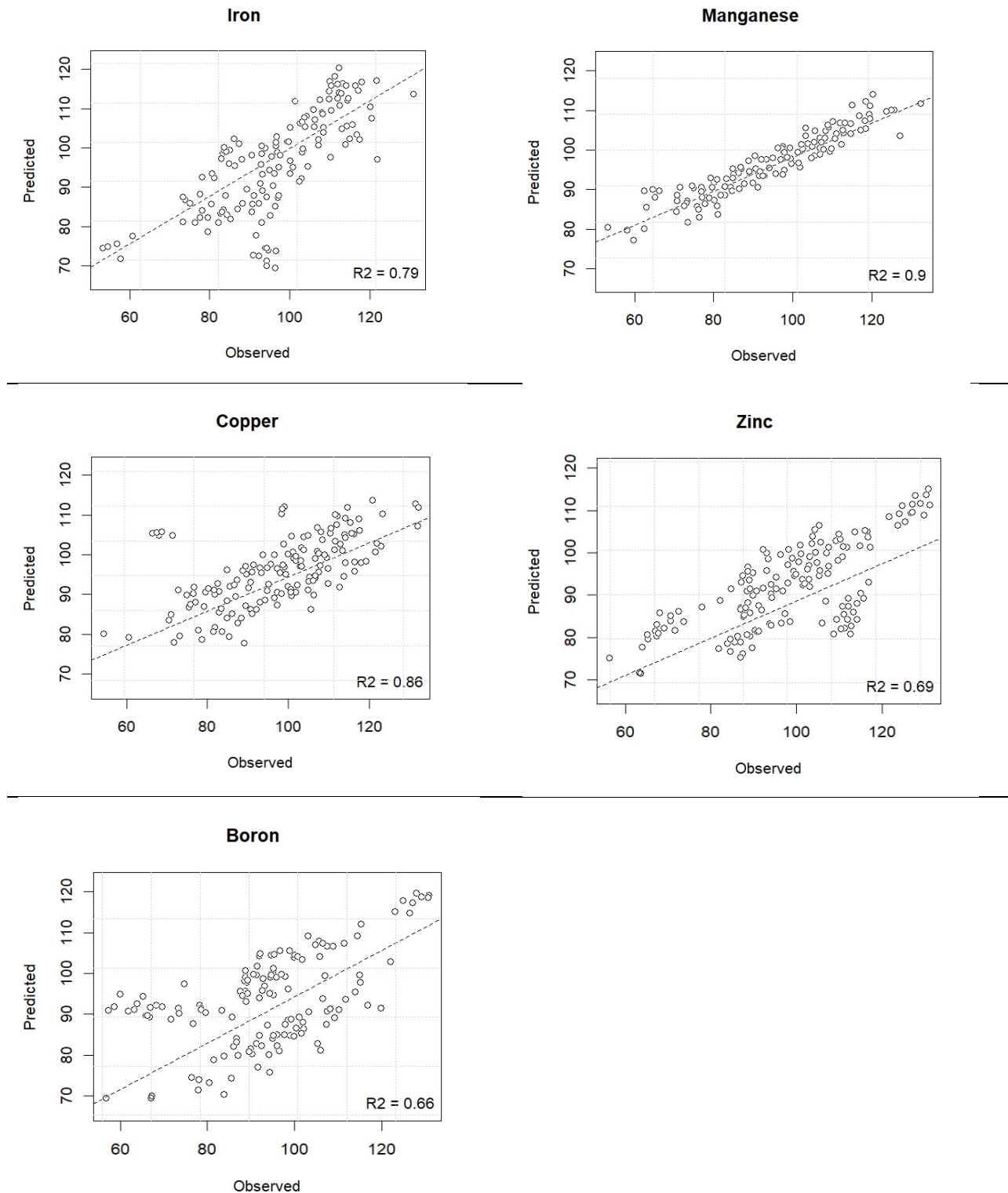

Figure 4: The one-to-one relationship between predicted versus observed micronutrients: Fe, Mn, Cu, Zn, and B using RF.

4.4.   Variable importance of macronutrients

Figure 5 shows the most effective wavebands for detecting macronutrients at low and high concentrations used in the prediction model. The RF ensemble measures variable importance to obtain the most important wavebands. In this study, we considered the top ten most important wavebands for each macronutrient. To avoid region generalisations, we partitioned the short-wave infrared (SWIR) region into SWIR-1 (1300nm to 1900nm) and SWIR-2 (1900nm to 2500nm). For N content and Na concentrations, the most important wavebands are (1949nm, 675nm, 722nm, 892nm, 908nm, 918nm, 930nm, 947nm, 950nm, 955nm) and (901nm, 931nm, 955nm, 1131nm, 1371nm, 1420nm, 1960nm, 902nm, 907nm, 908nm) found in the NIR and red-edge regions of the spectrum, respectively. For P, Ca, and Mg concentrations, the important wavebands are (429nm, 353nm, 360nm, 367nm, 407nm, 413nm, 438nm, 444nm, 452nm, 453nm), (357nm, 383nm, 354nm, 355nm, 362nm, 389nm, 408nm, 422nm, 507nm, 672nm) and (1295nm, 1302nm, 1980nm, 1296nm, 429nm, 442nm, 444nm, 449nm, 450nm, 453nm) found in the visible (VIS) region, respectively. Whilst, for K concentration detection, the most important wavebands are (1912nm, 389nm, 350nm, 390nm, 2045nm, 2124nm, 2139nm, 2494nm, 406nm, 2098nm) found in the VIS (40 %) and SWIR-2 region (60 %), respectively (Figure 5).
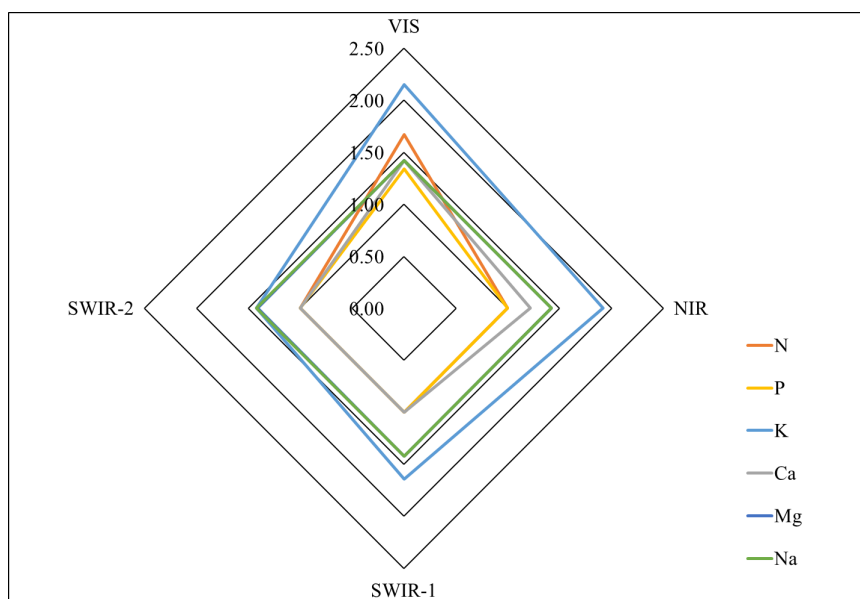

Figure 5: Radar plot showing important spectral regions for detecting deficient macronutrients: N, P, K, Ca, Mg, and Na. Illustration of the electromagnetic regions are in the following colours: VIS (yellow), NIR (green), SWIR-1 (blue) and SWIR-2 (red) in the bar graph.

4.5.   Variable importance of micronutrients

Figure 6 shows the most effective wavebands for detecting micronutrients at the low and high concentrations used in the prediction model. Fe, Zn and B produced correlations with the NIR region. The most critical wavebands for detecting Fe, Zn and B in this study were (1123nm, 1128nm, 894nm, 907nm, 919nm, 922nm, 941nm, 958nm, 961nm, 969nm), (788nm, 1186nm, 949nm, 1564nm, 445nm,

469nm, 636nm, 714nm, 742nm, 744nm), (784nm, 896nm, 904nm, 1263nm, 897nm, 1069nm, 1073nm, 1170nm, 1307nm, 1318nm), respectively. Mn and Cu showed strong correlations with the SWIR-2 region. The most critical wavebands for detecting Mn and Cu in this study were (2496nm, 1909nm, 2418nm, 2495nm, 2451nm, 1800nm, 1877nm, 1222nm, 1260nm, 1291nm) and (1906nm, 1921nm, 1914nm, 1923nm, 1922nm, 1909nm, 1929nm, 2483nm, 405nm, 1225nm) found in the SWIR-2 region, respectively (**Figure 6**).
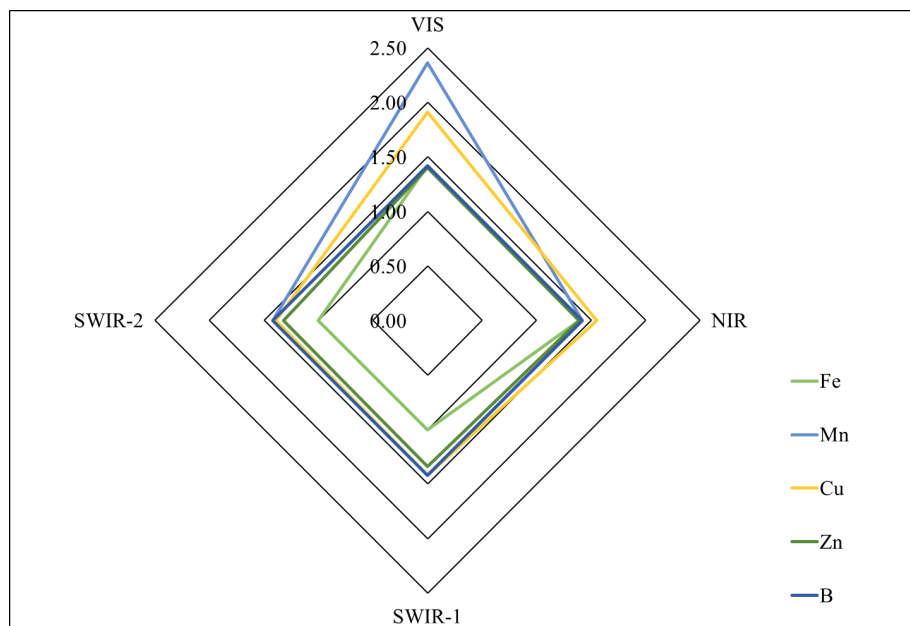

Figure 6: Radar plot showing important spectral regions for detecting deficient micronutrients: Fe, Mn, Cu, Zn, and B. Illustration of the electromagnetic regions are in the following colours: VIS (yellow), NIR (green), SWIR-1 (blue) and SWIR-2 (red) in the bar graph.

5. Discussion

In commercial forestry, the supply of nutrients from root to shoot is vital for plant growth and forest productivity. Quantifying nutrient deficient trees in a compartment remains unworkable and could prove challenging when dealing with many younger plants in the nursery. Hence new methods are needed to be adaptable early, either at the nursery before planting, to provide rapid detections or out in the field (Quentin *et al.*, 2017). The early detection of potential nutrient depletion at the nursery level could guide the optimisation of future forest management practices and lead to a more robust approach to nutrient measurement and assessment (Garcia *et al.*, 2018). A recent study by Acevedo *et al.* (2020) suggests that the use of nutrient loading at the nursery level will improve seedling nutritional status, morphological attributes, and the growth of new roots. The authors suggest the need for modelling growth responses to improve their understanding of physiological processes further.

This study successfully determined foliar macronutrients in *E. grandis x E. urophylla* at low and high nutrient concentration levels using hyperspectral data and RF. The prediction results were high throughout all macronutrients, which is vital for improving remote sensing efficacy, particularly in

deficient samples. The prediction results (R^2) of the most limiting growth nutrients N, P, K in this study explain the findings of previous studies (Adams *et al.*, 2000; Axelsson *et al.*, 2013; Özyiğit and Bilgen, 2013), which detected similar R^2 accuracies in foliar wheat and grass samples. The use of the RF algorithm significantly improved detection accuracy when compared to previous studies that used VI's and partial least squares (PLS) to detect N, P, K and Na in wheat samples (Pimstein *et al.*, 2011; Mahajan *et al.*, 2014; Oliveira *et al.*, 2017). For example, P predicted considerably better than previous studies (Özyiğit and Bilgen, 2013; Mahajan *et al.*, 2014) with R^2's above 0.90, while previous studies produced R^2's below 0.50. The RF algorithm has built-in parameter fine-tuning, which permits the optimisation of the n_{tree} function, providing more robust results than generalised VI's used in previous studies. RMSE values remained low; however, we found N samples to be considerably higher than the rest but still permissible. Our findings confirm that foliar micronutrients can be detected in *E. grandis x E. urophylla* at both deficient and not deficient levels using hyperspectral data and RF. While many studies generally predict macronutrients N, P, K, this study predicted a wide range of micronutrients. Furthermore, this study predicted micronutrients: Ca, Mg, Na, Mg at low and high concentrations. This study could predict micronutrients better with prediction accuracies ranging from 0.66 to 0.90. While other studies used VI's (Adams *et al.*, 2000; Özyiğit and Bilgen, 2013; Oliveira *et al.*, 2017), this study achieved higher accuracies when using the RF algorithm. These results validate and promote the robustness and effectiveness of the RF ensemble to discriminate each micronutrient, especially when using high dimensional data.

An important step was to determine which waveband regions correlate with the deficient nutrients. To our knowledge, our study is the first to examine the critical wavebands for detecting macronutrient and micronutrient deficiencies in foliar tree material. Hence, we could not directly compare this study's results to previous studies, which mainly examined N deficiency in heterogeneous environments (Blackmer *et al.*, 1996; Goel *et al.*, 2003). However, we could correlate our results with other corresponding regions of the electromagnetic spectrum associated with general reflectance markers in foliar material. For example, this study found that most macronutrient deficiencies correlate with wavebands in the VIS (P, Ca, Mg) and NIR (N, Na) regions. Blackmer *et al.* (1996) found similar correlations in the VIS region when examining N deficiency in corn using a portable spectroradiometer (350-1100nm). Their study could not examine the NIR edge and SWIR regions in the absence of the latest instrumentation, which was essential for determining N deficiency in our study.

Furthermore, According to Goel *et al.* (2003), wavebands 498nm and 671nm in the VIS region correlated with N stress in foliar corn material. Similarly, in this study, we found waveband 675nm closely related to their study (Figure 5). However, in this study higher correlations were found for N deficiencies in the red-edge and NIR region. Like N, most micronutrient deficiencies correlated with the NIR (Fe, Zn, B) regions related to Liew *et al.* (2008). Mn and Cu deficiencies were more closely related to the SWIR-2 region, which is a new finding. We emphasised obtaining lower RMSE values than higher R^2 values, which are highly crucial for generating more robust models for each nutrient during modelling. Lower RMSE values improve the technology's efficacy and provide confidence to

the user (forester, nursery manager or technical staff), particularly during the system's implementation into commercial forestry nurseries. The RF algorithm helped provide variable importance measures, essential for identifying the most critical wavebands in the prediction amongst highly dimensional data. Deriving reference paired τ-test results formed an essential component of this study for deciphering between nutrient deficient trees and not nutrient deficient. The results from the reference paired τ-test showed that there was a significant difference ($\rho \leq 0.05$) in samples with deficient and not deficient for all macronutrients and micronutrients.

This study provides a framework for proactive decision making about the nutrient health status of a tree. Nurseries could use this method for quality control and risk assessment purposes. Rapid spectroscopy is cost-effective, time-efficient and requires fewer resources for the chemical processing of samples. Furthermore, future studies should upscale this assessment to a live standing compartment. This study will help foresters, land managers, and commercial timber industries rapidly assess each tree's health status within a compartment. Upscaling to a satellite would be beneficial; however, problems of resolution accuracy may hinder successful detection.

6. Summary and Conclusion

To our knowledge, this is the first study that has explored remote sensing of a full range of tree macronutrients (N, P, K, Ca, Mg, Na) deficiencies and micronutrients (Fe, Mn, Cu, Zn, B) deficiencies, using full-waveform hyperspectral data (350nm-2500nm) and the RF algorithm. From this study, we conclude that:

1. The study successfully predicted N, P, K, Ca, Mg, S, Fe, Mn, Cu, Zn, and B in *E. grandis x E. urophylla* using hyperspectral data and RF analysis.

2. Variable importance results predicted wavebands for detecting nutrient deficiencies in *E. grandis x E. urophylla*.

3. The results improve the efficacy of using remote sensing methods for nutrient analysis in a high productivity forestry nursery environment.

We recommend using this study as a framework for rapid plant nutrient analysis in commercial forestry nurseries. Future research could upscale the results from this study from nursery to field level using remote sensing imagery.

7. Acknowledgements

References

Abdel-Rahman, E. M., Mutanga, O., Odindi, J., Adam, E., Odindo, A. & Ismail, R. 2017. Estimating Swiss chard foliar macro-and micronutrient concentrations under different irrigation water sources using ground-based hyperspectral data and four partial least squares (PLS)-based (PLS1, PLS2, SPLS1 and SPLS2) regression algorithms. *Computers and Electronics in Agriculture,* 132**,** 21-33.

Acevedo, M., Rubilar, R., Dumroese, R. K., Ovalle, J. F., Sandoval, S. & Chassin-Trubert, R. 2020. Nitrogen loading of Eucalyptus globulus seedlings: nutritional dynamics and influence on morphology and root growth potential. *New Forests***,** 1-16.

Adams, M. L., Ns, W. A., Philpot, W. D. & Peverly, J. H. 2000. Spectral detection of micronutrient deficiency in 'Bragg'soybean. *Agronomy Journal,* 92**,** 261-268.

Agjee, N. E. H., Mutanga, O., Peerbhay, K. & Ismail, R. 2018. The impact of simulated spectral noise on random forest and oblique random forest classification performance. *Journal of Spectroscopy,* 2018.

Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Ismail, M. F., Tan, N. P. & Karam, D. S. 2020. Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest. *Computers and Electronics in Agriculture,* 169**,** 105221.

Axelsson, C., Skidmore, A. K., Schlerf, M., Fauzi, A. & Verhoef, W. 2013. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *International Journal of Remote Sensing,* 34**,** 1724-1743.

Belgiu, M. & Drăguţ, L. 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing,* 114**,** 24-31.

Blackmer, T. M., Schepers, J. S., Varvel, G. E. & Walter-Shea, E. A. 1996. Nitrogen deficiency detection using reflected short-wave radiation from irrigated corn canopies. *Agronomy Journal,* 88**,** 1-5.

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

Breiman, L. 2017. *Classification and regression trees*, Routledge.

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

Dixit, L. & Ram, S. 1985. Quantitative analysis by derivative electronic spectroscopy. *Applied Spectroscopy Reviews,* 21**,** 311-418.

Dye, P. 1996. Response of Eucalyptus grandis trees to soil water deficits. *Tree Physiology,* 16**,** 233-238.

Elvidge, C. D. 1990. Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing,* 11**,** 1775-1795.

Ferwerda, J. G., Skidmore, A. K. & Mutanga, O. 2005. Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of Remote Sensing,* 26**,** 4083-4095.

Garcia, W., Amann, T. & Hartmann, J. 2018. Increasing biomass demand enlarges negative forest nutrient budget areas in wood export regions. *Scientific Reports,* 8**,** 1-7.

Goel, P. K., Prasher, S. O., Landry, J.-A., Patel, R. M., Bonnell, R., Viau, A. A. & Miller, J. 2003. Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn. *Computers and Electronics in Agriculture,* 38**,** 99-124.

Knox, N. M., Skidmore, A. K., Prins, H. H., Asner, G. P., van der Werff, H. M., De Boer, W. F., Van Der Waal, C., De Knegt, H. J., Kohi, E. M. & Slotow, R. 2011. Dry season mapping of savanna forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. *Remote sensing of environment,* 115**,** 1478-1488.

Liaw, A. & Wiener, M. 2002. Classification and regression by randomForest. *R News,* 2**,** 18-22.

Liew, O. W., Chong, P. C. J., Li, B. & Asundi, A. K. 2008. Signature optical cues: emerging technologies for monitoring plant health. *Sensors,* 8**,** 3205-3239.

Mahajan, G., Sahoo, R., Pandey, R., Gupta, V. & Kumar, D. 2014. Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.). *Precision Agriculture,* 15**,** 499-522.

Meng, X.-L. 1994. Posterior predictive $ p $-values. *The Annals of Statistics,* 22**,** 1142-1160.

Mucina, L. & Rutherford, M. 2006. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia 19.,(South African National Biodiversity Institute: Pretoria, South Africa). *Memoirs of the Botanical Survey of South Africa.*

Mutanga, O., Adam, E. & Cho, M. A. 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation,* 18**,** 399-406.

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H. & Bauer, D. 2014. The genome of Eucalyptus grandis. *Nature,* 510**,** 356-362.

Nerc. Undated. *Excel Post Processing Templates and User Guides* [Online]. Available: https://fsf.nerc.ac.uk/resources/post-processing/Excel.shtml [Accessed June 2018].

Oliveira, L. F. R. D., Oliveira, M. L. R. D., Gomes, F. S. & Santana, R. C. 2017. Estimating foliar nitrogen in Eucalyptus using vegetation indexes. *Scientia Agricola,* 74**,** 142-147.

Oliveira, L. F. R. D. & Santana, R. C. 2020. Estimation of leaf nutrient concentration from hyperspectral reflectance in Eucalyptus using partial least squares regression. *Scientia Agricola,* 77.

Osco, L. P., Ramos, A. P. M., Faita Pinheiro, M. M., Moriya, É. A. S., Imai, N. N., Estrabis, N., Ianczyk, F., Araújo, F. F. D., Liesenberg, V. & Jorge, L. A. D. C. 2020. A Machine Learning Framework to Predict Nutrient Content in Valencia-Orange Leaf Hyperspectral Measurements. *Remote Sensing,* 12**,** 906.

Özyiğit, Y. & Bilgen, M. 2013. Use of spectral reflectance values for determining nitrogen, phosphorus, and potassium contents of rangeland plants. *Journal of Agricultural Science and Technology,* 15**,** 1537-1545.

Pajares, G. 2015. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Engineering & Remote Sensing,* 81**,** 281-330.

Pimstein, A., Karnieli, A., Bansal, S. K. & Bonfil, D. J. 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *Field Crops Research,* 121**,** 125-135.

Pinto, D. S., Resende, R. T., Mesquita, A., Rosado, A. M. & Cruz, C. D. 2014. Early selection in tests for growth traits of Eucalyptus urophylla clones test. *Scientia Forestalis,* 42**,** 251-257.

Pullanagari, R., Kereszturi, G. & Yule, I. 2016. Mapping of macro and micro nutrients of mixed pastures using airborne AisaFENIX hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 117**,** 1-10.

Quentin, A. G., Rodemann, T., Doutreleau, M. F., Moreau, M. & Davies, N. W. 2017. Application of near-infrared spectroscopy for estimation of non-structural carbohydrates in foliar samples of Eucalyptus globulus Labilladiere. *Tree Physiology,* 37**,** 131-141.

Reuter, D. & Robinson, J. B. 1997. *Plant Analysis: An Interpretation Manual*, CSIRO publishing.

Ruxton, G. D. 2006. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology,* 17**,** 688-690.

Silva, J. A. & Uchida, R. 2000. Plant Nutrient Management in Hawaii's Soils, Approaches for Tropical and Subtropical Agriculture. *Plant Nutrient Management in Hawaii's Soils*, 31-55.

Wang, Z., Skidmore, A. K., Darvishzadeh, R. & Wang, T. 2018. Mapping forest canopy nitrogen content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. *Agricultural and Forest Meteorology,* 253**,** 247-260.