Determination of Nutrient Composition of Mangroves through Combined Statistical and Physical Modeling

Alan Rex
March, 2012

DETERMINATION OF NUTRIENT COMPOSITION OF MANGROVES THROUGH COMBINED STATISTICAL AND PHYSICAL MODELING

by

Alan Rex

Thesis submitted to the International Institute for Geo-Information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Environmental Modelling and Management.

Thesis Assessment Board

Dr. Ir. C.A.M.J. de Bie (Chair)
Dr. M.F. Noomen (External Examiner)
Prof. Dr. Ing. W. Verhoef (First Supervisor)
Prof. Dr. A.K. Skidmore (Second Supervisor)

UNIVERSITY OF TWENTE.

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-Information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Abstract

A novel hybrid model method was developed by applying partial least squares regression (PLSR) to spectral residuals which were created from by subtracting simulated spectra from measured spectra. Simulated spectra were generated by the soil-leaf-canopy (SLC) radiative transfer model. The hybrid model was applied to HyVista hyperspectral images of mangrove forests in Kalimantan, Indonesia, for the purpose of modeling and mapping the concentrations of six foliar nutrients; nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg) and sodium (Na).

The model was compared to PLSR models generated from raw spectra and minimum noise fraction (MNF) transformed spectra. The models generated achieved poor to fair results. Models generated from residuals were significantly better ($p<0.01$) for N, Mg and Na. There was no significant difference between the models for P and Ca, while the raw spectrum model was significantly better for K.

There was a danger of losing information on nutrient concentrations during the subtraction process due to the correlation between chlorophyll and both N and Mg. Therefore two sets of residuals were created; one set from the complete SLC model, and one set from the SLC model where chlorophyll was kept constant. This study found that keeping chlorophyll constant in the physical model improved the model results when modeling N using spectral residuals (from $R^2=0.462$ to $R^2=0.575$ and from nRMSE=18.3% to nRMSE=16.0%) but diminished model results when modeling Mg using spectral residuals (from $R^2=0.263$ to $R^2=0.156$ and from nRMSE=21.6% to nRMSE=23.2%).

This study also created nutrient concentration maps using the models for N made with chlorophyll kept constant, to investigate a potential correlation with shrimp ponds and concentration levels. Two correlations with N concentration were found; proximity to coastline and size of forest stand, but not proximity to shrimp ponds.

Acknowledgements

I would like to first thank my supervisors, Dr. Wouter Verhoef and Dr. Andrew Skidmore for their helpful criticisms and advice. Their doors were always open.

I would also like to thank Dr. Martin Schlerf, Dr. Brittany Brand and Dr. Joshua Bandfield for their guidance as I started this journey. I am grateful to Flor Alvarez for her patience helping me with radiometry and to Lunds Universitet for a year of excellent preparation and education.

I would like to acknowledge all of my fellow GEM colleagues, especially Giles Williams, who struggled through this data with me. It has been amazing to work with you all for the last year and a half.

Finally I would like to thank my family, as well as Damir Makic, for their support.

Table of Contents

List of figures

List of tables

1 Introduction

1.1 Background

Mangrove forests provide a wide variety of unique ecosystem functions in the coastal regions they occupy throughout the tropics. Beyond providing basic habitat for many animals, including prawns, insects, sponges, mammals and birds (Holguin et al., 2001; Nagelkerken et al., 2008), there are signs that mangroves benefit the flora and fauna of neighboring ecosystems as well (Hemminga et al., 1994). Research also indicates that mangroves act as a nursery for young fish and prawns and possibly provide protection or other services in later stages of life (Nagelkerken, et al., 2008; J.H. Primavera, 1998).

Furthermore, mangroves act globally as a carbon sink, and studies suggest that the magnitude of that sink may have been previously underestimated (Bouillon et al., 2008). While the above ground biomass between mangrove species varies greatly, carbon stocks of mangroves forest include both the biomass as well as carbon sequestered in soils and detritus filled pools. These pools account for approximately 70% of the carbon stocks in mangrove systems (Kauffman et al., 2011). Per hectare, these combined carbon stocks are larger than most tropical, temperate and boreal forests (Donato et al., 2011; Kauffman, et al., 2011).

Mangroves also provide direct benefits to humans. Large mangrove forests are known to provide some level of protection to human settlements during natural disasters such as tsunamis and hurricanes (Alongi, 2008; Spalding et al., 2010) and mangroves have been used for decades to treat sewage and effluent. In addition to the trees themselves, it has been found that mangrove root systems harbor microbes which can denitrify wastewater at high rates (Corredor & Morell, 1994). As a source of fish production, healthy mangroves annually generate from $750 to $16750 USD per hectare for fishing industries (Patrik, 1999).

1.1.1 Risks to mangroves

Many mangrove forests are currently at risk from human activity; specifically pollution and intensive deforestation. Mangroves are cleared mostly for agriculture but also aquaculture, urban development and logging (Giri et al., 2008; Parks & Bonifaz, 1994; Rubin et al., 1999). In Southeast Asia, for example, over 12% of the

total mangrove forests were cleared between 1975 and 2005 (Giri, et al., 2008). This trend can be seen throughout the globe; Philippine mangroves decreased by approximately 75% from 1918 to 1994 (J.H. Primavera, 2000), mainland Caribbean mangroves decreased by 1.7% per year between 1980 and 1990 and island Caribbean mangroves decreased 0.2% per year during the same period (Ellison & Farnsworth, 1996). Globally, the total coverage of mangrove forests in Earth decreased by about 2% per year between 1980 and 1990 and about 1% per year between 1990 and 2000 (Wilkie & Fortuna, 2003).

In addition to cutting mangrove forests down, untouched mangroves are susceptible to damage by nutrients leaking in from surrounding agriculture, aquaculture or other pollution sources, as well as poor water management. While mangroves have been shown to effectively work in effluent treatment, increasing nutrients such as nitrogen and phosphorous in the ecosystem can cause mortality, root weakening and leaf loss in mangroves (Reef et al., 2010). Leaf loss and mortality can be a result of damage to N_2-fixing bacteria in the root systems (Holguin et al., 2006; Vovides et al., 2011). These N_2-fixing bacteria are susceptible to increased water temperatures, increased pH, increased salinity and higher levels of ammonium in the water, all of which can be linked to mismanagement of the water table (Vovides, et al., 2011).

1.1.2 Monitoring mangroves

In order to protect mangrove forests it is important to monitor changes in health indicators. Since high nutrient concentrations in soil lead to high foliar chemical content in mangroves (Oxmann et al., 2010), monitoring the nutrient content of leaves can be used to assess the levels of nutrients present in an ecosystem (Townsend et al., 2007). Furthermore, the concentrations of nutrients in leaves can give indications to the rates of biological processes such as photosynthesis, respiration, and evapotranspiration, which in turn give information about the health of an ecosystem (Majeke et al., 2008). For example, both nitrogen (N) and magnesium (Mg) are constituents of chlorophyll, and therefore measuring N or Mg will give an indication of the photosynthetic capacity of the plant (Ayala-Silva & Beyl, 2005; Evans, 1989; Field & Mooney, 1986).

While N is one of the most important nutrients to plant health (M. E. Martin et al., 2008), there are several nutrients which are also known to be growth limiting nutrients. These include phosphorus (P) and potassium (K) (Mutanga et al., 2004; Pimstein et al., 2011) as well as, to a lesser extent calcium (Ca), sodium (Na) and Mg (Marschner,

1995). In addition to being important for plant health, the concentrations of several of these foliar nutrients also have a significant impact on migratory animals and local herbivores (Ferwerda & Skidmore, 2007). Unfortunately, monitoring these nutrients in mangroves in the field is a difficult and costly process, especially due to their particular inaccessibility (Green et al., 1998).

Remote sensing is a cheaper alternative to field work which has recently been used for ecosystem monitoring (Kerr & Ostrovsky, 2003). Remote sensing can completely cover vast areas at rates many orders of magnitudes faster than field measurements and requires less direct human involvement (Coppin & Bauer, 1996). Remote sensing has already been used extensively in mapping the extent and change of mangrove forests, as well as to a lesser degree modeling certain forest characteristics such as leaf area index (LAI), canopy height, biomass and species (Heumann, 2011).

Remote sensing can also be used to estimate nutrient concentrations; foliar biochemicals absorb light at certain wavelengths, boosting their valence electrons into higher energy states and creating absorption features in the spectra of light reflected off the leaf (Curran, 1989). However, most studies focus almost exclusively on grasses (Mutanga & Skidmore, 2004; Ramoelo et al., 2011) or agricultural crops (Delegido et al., 2010). Those studies that do focus on forest environments have rarely successfully measured many nutrients beyond N (Ferwerda & Skidmore, 2007; Pimstein, et al., 2011). In forests, the measurement of nutrients from absorption features has had limited success because the measured reflectance spectra also contain features from light reflected off the soil, branches/trunk and undergrowth in the forest (Asner & Martin, 2008). Mangroves are especially difficult because they often feature tidal changes in soil reflectance, a mixture of species, and gaps in the canopy crown cover (Asner, 2008; Heumann, 2011; Minchinton, 2001).

Non-homogenous forests pose a special problem for remotely sensing N. Two or more different species in a forest may have different typical LAI or canopy architecture, which in turn can influence the impact of N on spectra. This means that it may be difficult to determine if spectral variation across the canopy is due to the actual foliar N variations or simply to the different species (Asner, 2008).

1.1.3 Modeling

In order to effectively use remote sensing to map nutrients, it is important to develop a useful model. There are three basic models for biochemical features; statistical models, physical models and

hybrid models (Liang, 2005). Physical models are built on known physical relationships between spectra and biophysical parameters. Biochemical as well as physical parameters can be interpolated through a method called inversion (figure 1), where the measured reflectance spectrum is compared to simulated spectra generated by the physical model (Kimes et al., 2000). The physical parameters of the most similar simulated spectrum are then assumed to be the same as the parameters of the measured spectrum. These parameters are retrieved from a look up table (LUT), which was used to generate the simulated spectra. The benefit of a physical method is that it can be applied universally. Some physical models have recently been demonstrated to achieve a high correlation between observed and predicted nutrient levels (Zhang et al., 2008). A danger of inversion is that two or more different sets of parameters can have similar or identical simulated spectra (Combal et al., 2003). When inverting, it would then be possible that an incorrect set of physical parameters could be matched to a measured spectrum (a problem called ill-posedness).

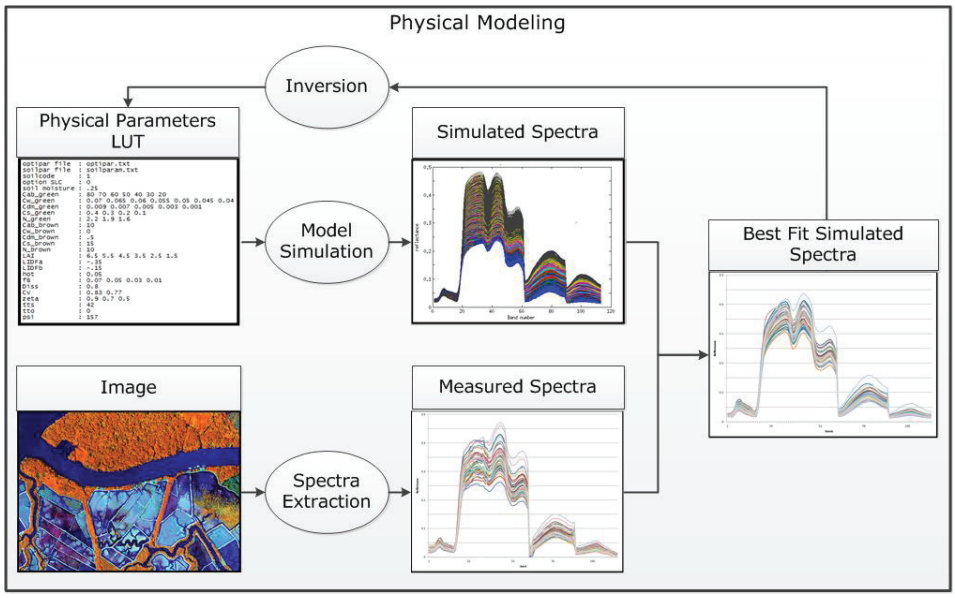

Figure 1 - Conceptual model of physical inversion modeling.

Alternatively, statistical models find the relationship between foliar biochemicals and reflectance spectra using one of a variety of regression methods. While these different methods use different statistical techniques, they all seek to correlate one or many explanatory variables with a response variable. A “best fit” model is then calculated for the data points and interpolated over the test

surface. Statistical models can be used to find previously unknown relationships between biochemicals and spectra. Recent studies have demonstrated the ability of statistical regression analysis (Ge et al., 2008) and narrow band vegetation indices (Gil-Perez et al., 2010) to predict nutrient concentrations with some success. One common method of regression is partial least squares regression (PLSR). PLSR functions by looking at the latent structures of both explanatory and response variables and then finding indicators in the explanatory variables that best explains the maximum variation in the response (Haenlein & Kaplan, 2004). These indicators are derived from user defined weighted relationships. The benefit of PLSR is that it doesn't make any assumptions about the distribution of the input data, and it can effectively deal with multicollinearity.

Noise reduction techniques are often used to transform data before the application of a statistical model (Harris et al., 2006). While they often have different methods, all of these techniques aim to remove spectral features that are not directly related to the feature under study, thereby isolating the feature of interest (R. N. Clark & Roush, 1984).

Two similar and common noise reduction techniques are minimum noise fraction (MNF) and principle component analysis (PCA). Both of these techniques operate by projecting measured data into new reference frames based of maximum variance within the data. "Dimensions" or "components" with minimal variance can be considered noise and removed from the data (Kambhatla & Leen, 1997). Another recently developed approach called water removed (WR) spectra eliminated a simulated water spectrum from foliar reflectance spectra before applying regression methods to predict nitrogen and phosphorus content (Ramoelo, et al., 2011). Removing the physically modeled water spectrum improved results for both nutrients.

Finally, hybrid models attempt to combine the benefits of the two previous methods. A review of many comparative studies finds that hybrid models regularly outperform statistical models (Razi & Athappilly, 2005). Some common hybrid approaches are artificial neural networks (ANNs) and regression trees. ANNs, for example, work by connecting input parameters with respondent variables through a network of connections (or "synapses") and nodes (or "neurons"). This network can then contain user defined functions (like a physical model) and weights, which are empirically derived (like a statistical model) (Egmont-Petersen et al., 2002).

1.2 Research problem

There are drawbacks to every model. Statistical models, for example, can only be applied under conditions very similar to the ones in which they were created. This makes their usefulness extremely limited. While physical models can be applied universally, a drawback of physical models is that they can only be used when the effect of the specific biochemical or physical parameters on the reflectance spectrum is precisely known and they are incapable of discovering new relationships between parameters and outputs.

Statistical and physical models have been used less often on forest environments because of the challenging nature of the canopy setting (Ferwerda & Skidmore, 2007). While statistical models can be used to model foliar nutrients, they are hindered by the fact that other factors such as water, viewing angle and soil properties can have a much greater impact on the overall reflectance of the spectrum. The signals from nutrients are weaker and may be obscured by the "noise" which in this case is the signal of the other factors (Blackburn, 2007). Weaker signals prove more challenging to model and result in poorer models. Previous studies have demonstrated the ability of statistical models to detect nutrient concentrations at the leaf scale for some plants (Delegido, et al., 2010), however very little work has been done with mangroves.

While there are currently several physical models which incorporate the effects of biochemicals such as chlorophyll, proteins, and water content (Ganapol et al., 1999; Jacquemoud et al., 1996; Verhoef & Bach, 2007), the specific relationship between foliar nutrients and reflectance spectra is unknown in sufficient detail to include in a physical model. In terms of modeling, the physical reflectance models in use today ignore nutrients altogether.

When using the LUT/inversion method, a simulated spectrum is generated using only the physical parameters which make up the physical model. The measured spectrum is then assumed to have the same physical parameters as the simulated spectrum. However, there are always at least small differences between the simulated and measured spectra (called residuals, see figure 2). These residuals are a result of noise, as well as the parameters in the environment which are not included in the physical model. These parameters are usually ignored in order to maintain parsimonious models. However, this information could still hold valuable insights about the system being studied (figure 3).

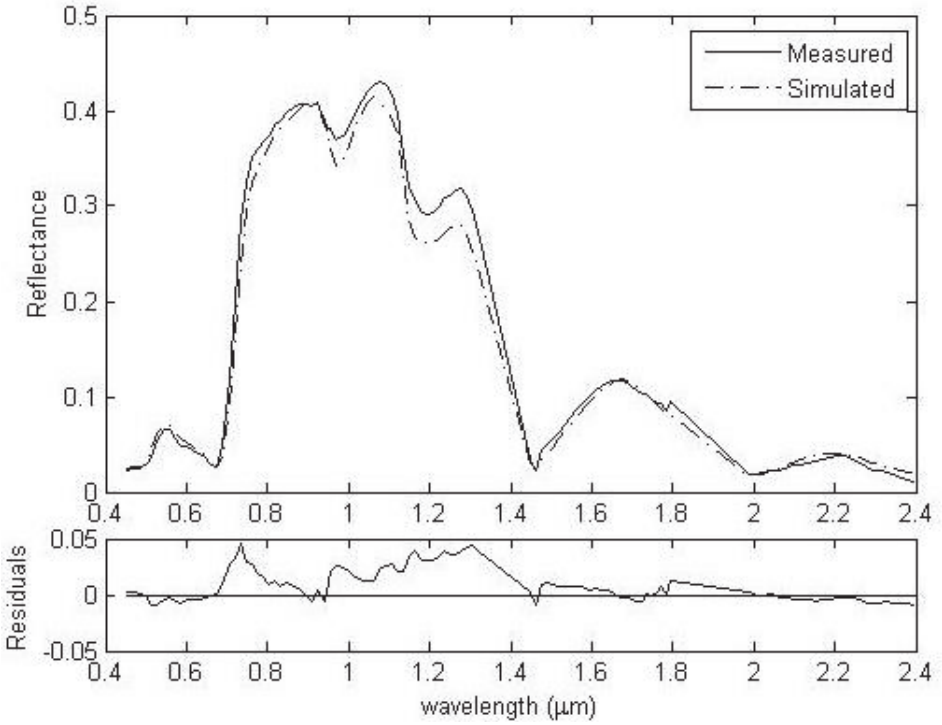

Figure 2 - Example of a residual spectrum from this study generated by subtracting a simulated mangrove reflectance spectrum from a measured mangrove reflectance spectrum.

1.3 Research objectives

1.3.1 General objective

The objective of this research is to develop a novel hybrid model method which will apply a statistical model to the information (residuals) discarded when inverting a physical model and investigate if those residuals are linked to foliar nutrients. That model will then be used to map nutrient concentrations in mangrove forests.

1.3.2 Specific objectives

- Assess the ability of this hybrid approach to predict nutrient concentrations using spectral residuals, in comparison to regression using an unmodified spectrum and a spectrum modified with the MNF transformation.

- Investigate if the correlations of chlorophyll with N and Mg impact the effectiveness of the hybrid model.

- Create nutrient maps showing the spatial distribution of the nutrients N, P, K, Ca, Na and Mg.

- Use the nutrient map to investigate a possible spatial correlation between shrimp ponds and mangrove nitrification.

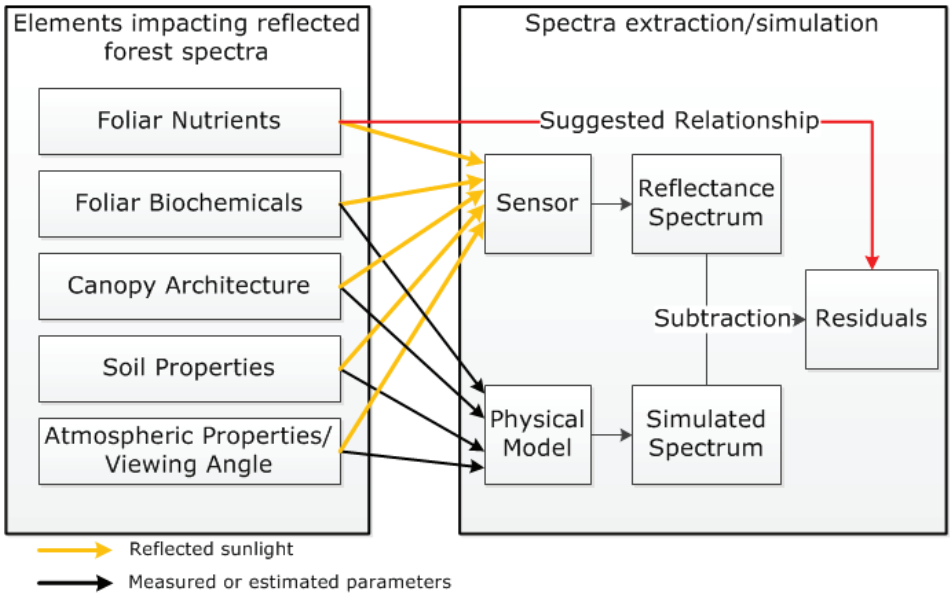

Figure 3 - Conceptual model of the system. Red arrow highlights the main subject of this study. This image demonstrates the idea that there should be a correlation between the information which generates simulated and reflectance spectra and the residuals between the two spectra.

1.3 Research questions

- Does regression of physical model residuals result in a better predictive model than regression of untransformed or MNF transformed spectra?

- Considering the known correlation between chlorophyll and N, as well as chlorophyll and Mg, will the inclusion of chlorophyll as a variable in the physical model mean that information about N concentrations are lost when the physical model is inverted, or should chlorophyll be kept constant?

1.4 Hypotheses

General Null Hypothesis

- There will be no significant difference between the various methods and types of regression.

Alternative Hypothesis 1

- There will be a statistically significant improvement in the results of the model using physical model residuals instead of unmodified or MNF transformed spectra.

Alternative Hypothesis 2

- There will be a statistically significant improvement in the results of the model if chlorophyll is kept constant in the physical model.

2 Materials and methods

2.1 Study area

The study areas for this research were mangrove forests in two river deltas on the province of Kalimantan, Indonesia (figure 4); the Mahakam delta which has highly developed shrimp aquaculture and the Berau delta which is relatively untouched by human influence.

2.1.1 Mahakam

The Mahakam River is the largest in the province of Kalimantan, and ends in an 1800 km^2 delta comprised of 46 islands. The area for this study was conducted in the northeastern section of the delta between 0°29'42"S – 0°35'45"S latitude and 117°27'59"E – 117°35'03"E longitude. While the Mahakam delta was once home to a large and diverse population of mangroves, 73% of the original forests have been converted into shrimp ponds (Zwieten et al., 2006). The Mahakam delta also suffers from increased urbanization and is at threat from pollution from the second largest fossil fuel drilling operation in Indonesia (Powell & Osbeck, 2010).

The delta is mostly comprised of two mangrove species; *Nypa fruticans* and *Rhizofora mucronata*. These two species have fairly distinct distributions; with *Rhizofora mucronata* generally being found near the coasts and *Nypa fruticans* more inland (Wandera, 2011). *Avicennia spp.*, *Brugueira spp.* and *Sonneratia spp.* are also present.

2.1.2 Berau

The Berau delta, unlike Mahakam, is relatively untouched by human interference. The coastal waters around the Berau delta are known for high levels of marine biodiversity (Kreb & Budiono, 2005). The delta is several hundred kilometers north of the Mahakam, and is substantially smaller. The study area is located in the south end of the delta, between 1°57'02"N – 2°04'31"N and 117°44'45"E – 117°54'17"E. While there are a few settlements nearby, the delta itself is mostly undisturbed and shrimp farming, while present, is significantly less developed than Mahakam. The delta is mostly comprised of *Rhizofora spp.*, but also contains *Nypa fruticans*, *Xylocarpus spp.* and *Brugueira spp.* with some *Avicennia spp.* along the saltier, more inundated coastline (Axelsson, 2011).

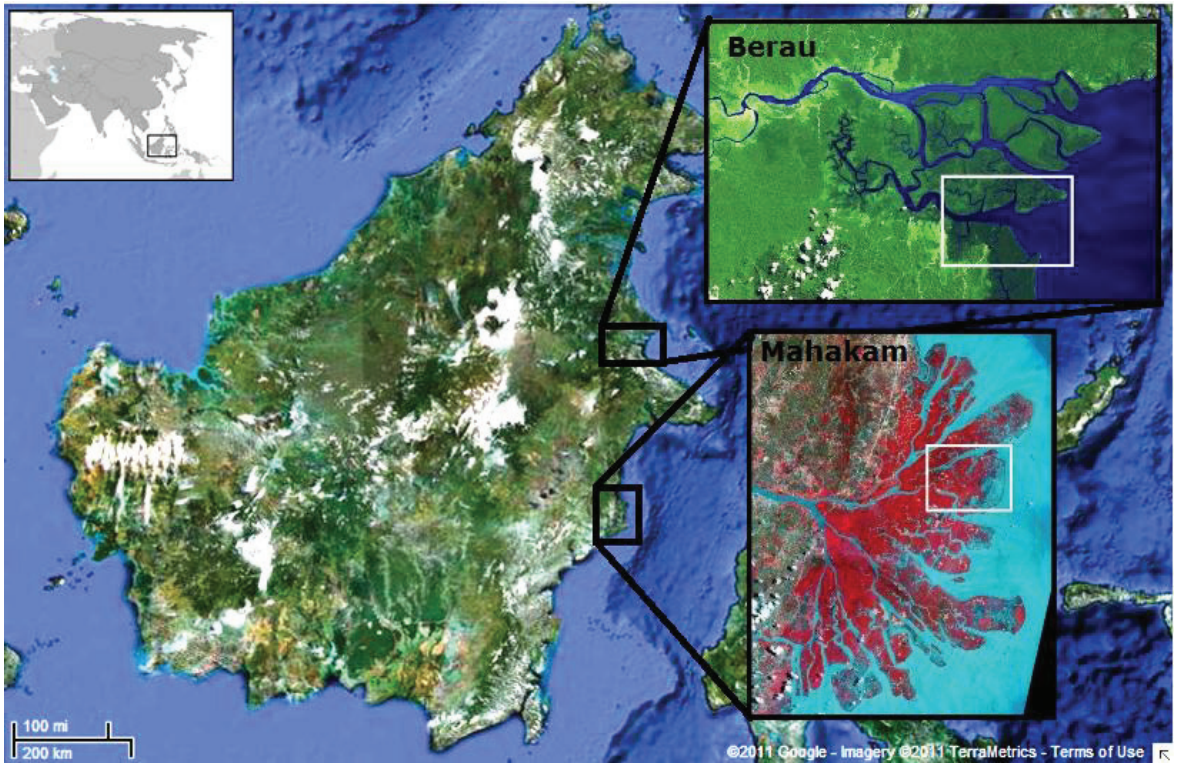

Figure 4 - Map of the study areas on the island of Kalimantan. Background image was taken from Google Maps. Expanded areas are a true color satellite image of Berau (Buschman, 2010) and a false color image of Mahakam (Darman et al., 1999). White/beige human settlements can be seen upriver in Berau while dark blue patches within the Mahakam delta are shrimp ponds. White spots in the lower left of both images are clouds. White boxes represent the extent of the HyMap images used in this study.

2.2 Image data

Aerial photographs were taken on October 16th (Mahakam) and 18th (Berau) of 2009 with a HyMap camera, which has proven capable of measuring foliar biochemicals (Mutanga & Skidmore, 2004). The resulting composite hyperspectral images have 126 channels which range from 420 to 2490 nm and have a resolution of 3.1 m.

Preliminary image processing was completed by the HyVista company of Sydney, Australia. Raw images were transformed into reflectance using the Hycorr atmospheric correction software. The images were geocoded using ground control points (GCP) at locations such as bridges and roads. Positional accuracy of the GPS units taking GCPs was between 6 and 9 meters. All of the images for each study area were collected along west-east flight lines and mosaics were created from the image strips (8 images for Mahakam and 9 for Berau). Clouds and cloud shadows were eliminated from the images by manually creating masks. This image processing was completed by two previous studies using the same data (Axelsson, 2011; Wandera, 2011).

Previous work on the Berau image found there to be errors in the atmospheric correction. Reflectance measurements in the visible wavelengths from mangrove forest pixels were considerably higher than the range of expected values. Therefore, the raw data was re-corrected using the FLAASH module within the ENVI software program. This work was completed by Flor Alvarez of the Universidad de Leon. Furthermore, 12 bands were removed from images of both study areas due to large amounts of noise and one band was removed because it was outside of the range of optical parameters contained in the soil leaf canopy (SLC) physical model.

Since the error for the GPS units was more than twice that of the image resolution, spectra were extracted at field sample points from an average of the four closest pixels (Mary E. Martin & Aber, 1997). Care was taken to use homogenous pixels, and avoid pixels with water cover, roots or soil. If four contiguous homogenous pixels were not available, then only two or three pixels were used depending on availability. A MNF transformation was also applied to both mosaics. Spectra were extracted from both untransformed and MNF transformed images, for comparison. The forward and inverse MNF transformations were both completed using the ENVI software. The entire image was transformed based on components of maximum variance. Components with the largest variance (the largest eigenvalues) were then inverted back into the original bands of the

image, and components with small eigenvalues were discarded. For both images, the ten largest components were inverted, which contained over 99% of the variance in the data.

2.3 Field data

Ground truth data was collected in October 2009 (Mahakam) and Sept/Oct 2010 (Berau) by previous research groups (Axelsson, 2011; Wandera, 2011). A total of 138 leaf samples were collected along forest transects, which were later analyzed for concentrations of N, P, K, Ca, Mg and Na. Sample points were predefined through a random representative sampling method which ensured that the samples would include wide biochemical variation. Hyperspectral images were used in advance to determine this variation. Due to the high inaccessibility of mangrove forests (Green, et al., 1998), areas far inland were excluded from the random sample. Samples were collected every 50 meters along transects, which were made at least 400 meters from the coastline, and ranged from 250 to 350 meters long.

When predetermined locations of sample sites were located in the field, effort was made to find areas with homogenous species cover, to avoid the risk of matching pixels with the wrong species. Areas with canopy gaps were avoided. Samples consisted of ten mature, non-senescent leaves, which were clipped from near the top of the canopy. Leaves were then sealed and shipped to the laboratory at Mulawarman University in Samarinda, Indonesia for chemical analysis. Leaves were dried and N, P, K, Ca, Mg and Na were measured as a percent of the dry matter (%dm). N was processed via the Kjeldahl method, P via the BioMate UV-visible spectrophotometer and K, Ca, Mg and Na with a BioMate atomic absorption spectrophotometer. Several sample points were removed after field work if they were covered by clouds during the imaging process or due to pixel mismatch, leaving 121 points total (47 in Mahakam and 74 in Berau). See sample points summary in table 1.

2.4 SLC model

2.4.1 Generate simulated spectra

The first step towards generating spectral residuals was to create simulated spectra for each sample point. The soil-leaf-canopy (SLC) physical reflectance model which was originally developed by Verhoef & Bach (2007). To be useful, a reflectance model must incorporate as much of the media through which the light passes as possible. Since this study is attempting to isolate the signal from foliar

nutrients, it is important to use a physical model which models as many parameters as possible. The SLC model was chosen because it incorporates the parameters of three previous physical models; PROSPECT, Hapke soil BRDF (bidirectional reflectance distribution function), and 4SAIL2 which model the radiative transfer through vegetation, soil reflectance, and atmosphere/canopy, respectively.

Table 1 - Measurement statistics for the nutrient concentration ground samples.

Chemical	Study Area	Mean (%dm)	Range (%dm)	Standard Deviation (%dm)
N	Mahakam	1.02	0.60-1.45	20.3
	Berau	1.09	0.66-2.30	26.6
P	Mahakam	0.10	0.07-0.16	20.7
	Berau	0.09	0.02-0.29	64.0
K	Mahakam	0.50	0.36-0.61	12.3
	Berau	0.67	0.24-1.45	35.5
Ca	Mahakam	0.75	0.28-1.48	50.2
	Berau	2.07	0.03-6.47	66.2
Mg	Mahakam	0.29	0.03-0.39	31.1
	Berau	0.29	0.10-0.38	23.8
Na	Mahakam	0.77	0.23-1.33	39.4
	Berau	0.74	0.23-0.90	22.8

While the SLC model is still subject to the danger of ill-posedness, this study should not suffer because an incorrect simulated spectrum will be virtually identical to the correct spectra. This study is not investigating any of the parameters of the physical model, so the only danger if incorrect simulated spectra are selected is that a small part of the foliar nutrient signal could be lost. To ameliorate this danger, the LUT in this study was developed to be as large as computationally possible. Minimizing the gaps between simulated spectra will minimize the potential error from this problem.

With the physical model chosen, input parameters where chosen to simulate spectra. Sensitivity analysis was performed to determine the impact that each variable had on the simulated reflectance spectra. Previous work had already determined that the reflectance was sensitive to chlorophyll concentrations, water and dry matter content, leaf structure, LAI and fraction of brown leaves. This study then determined if these sensitivities were uniform or varied.

The first step was to determine the possible range for each variable. When possible, these ranges were determined from the range of values measured in the field. Otherwise, ranges were determined

from literature, or by manually testing with the SLC physical model software (SLCdemo).

Two parameters (LAI and chlorophyll a and b concentration) were found to have a variable impact on spectra (see figure 5 for LAI example). However, analysis of the sum of squares of the residuals between measured and simulated spectra found that the non-uniform impact on reflectance from chlorophyll concentrations had no significant impact on the accuracy of the model. Therefore, LAI was the only parameter to be distributed non-uniformly in the LUT.

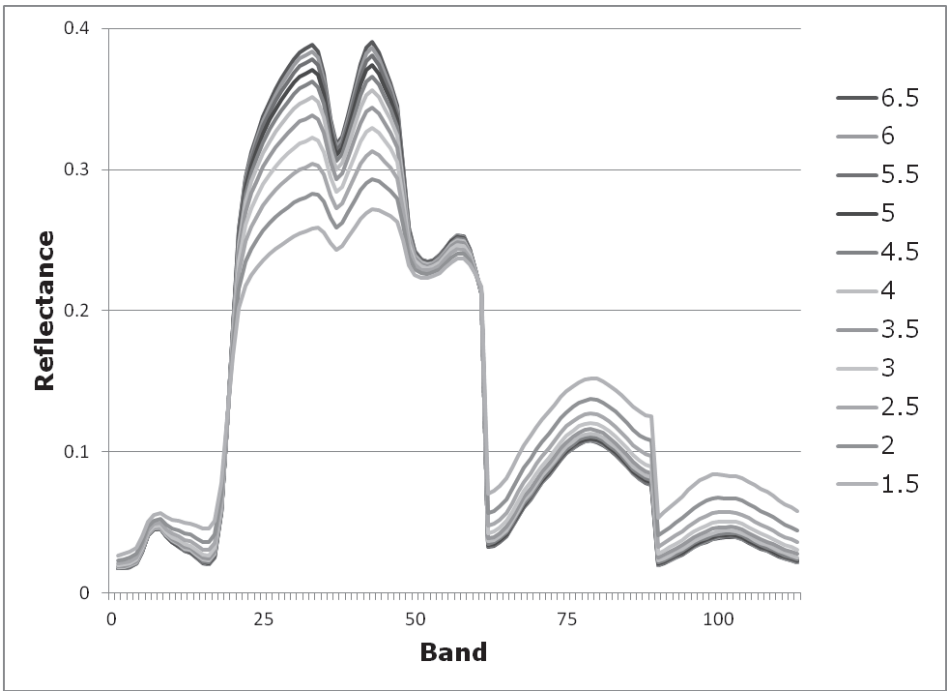

Figure 5 - Simulated spectra with LAI ranging from 1.5 to 6.5, all other variables constant. The figure shows that as LAI increases, is has a decreasing impact on reflectance spectra.

With parameters and ranges chosen, the LUT could be generated. However, due to limitations of processing power and RAM available for Matlab functions, the upper limit of the size of the LUT was approximately 500,000 simulated spectra. With nine variables included in the physical model, it was necessary to limit the number of possible values for each parameter. The magnitude of the impact of each variable was then assessed by testing the sensitivity of each parameter when all others were constant. Variables with low impact on the spectra over their range of possible values were then given fewer “steps” in the LUT. Due to the possible combined influence of

variables on the spectra, this test was also verified manually in the SLCdemo program.

Once parameters were chosen, the two LUTs were generated. For complete list of parameters, see table 2. The LUT with chlorophyll as a variable contained 364,880 simulated spectra. For the second LUT, chlorophyll was kept constant at the average measured value for the two study areas combined (60 µg/cm^2). This LUT contained 60,480 simulated spectra (figure 6).

The simulated spectra were then compared to the measured spectra, and a "best fit" was calculated. This process was completed for each LUT separately.

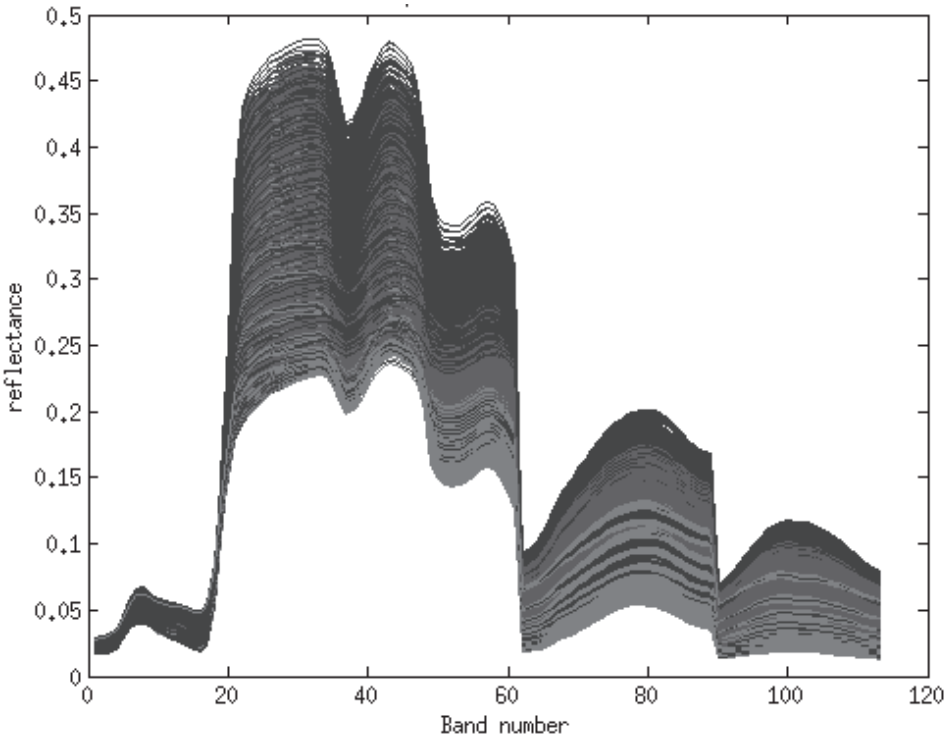

Figure 6 – 60,480 simulated spectra generated from LUT with chlorophyll kept constant.

Table 2 - Parameters used for the generation of the LUT when chlorophyll (CHL) was considered a variable. The second LUT contained the same parameters, with chlorophyll kept constant at 60.

Code	Parameter	Values
LIDFa	Average leaf slope	-0.35
LIDFb	Leaf inclination distribution function bimodality	-0.15
LAI	Leaf Area Index	5.5 5.0 4.5 3.5 2.5 1.5
Cab_green	CHL content (green leaves)	80 70 60 50 40 30 20
Cw_green	Water content (green leaves)	0.1 0.08 0.06 0.04 0.02
Cdm_green	Dry matter content (green leaves)	0.009 0.007 0.005 0.003 0.001
Cs_green	Senescence (green leaves)	0.4 0.3 0.2 0.1
N_green	Structure (green leaves)	2.2 1.9 1.7
Cab_brown	CHL content (brown leaves)	10
Cw_brown	Water content (brown leaves)	0
Cdm_brown	Dry matter content (brown leaves)	0.5
Cs_brown	Senescence (brown leaves)	15
N_brown	Structure (brown leaves)	10
hot	hot spot parameter	0.05
fB	Fraction of brown leaves	0.07 0.05 0.03
Diss	Canopy dissociation factor	0.8
Cv	Crown clumping	0.8
zeta	Crown cover	1 0.7 0.5
tts	Solar zenith angle	42
tto	Observation zenith angle	0
psi	Relative azimuth angle	157

2.4.2 Subtract simulated spectra from measured spectra

Each measured spectrum from the ground sample points were then matched to the simulated spectra with the lowest possible RMSE. This process was automated with a Matlab script. Each simulated spectrum was then subtracted from the corresponding measured spectrum, leaving residuals that contain all of the information not encompassed by the physical model (see example in figure 2).

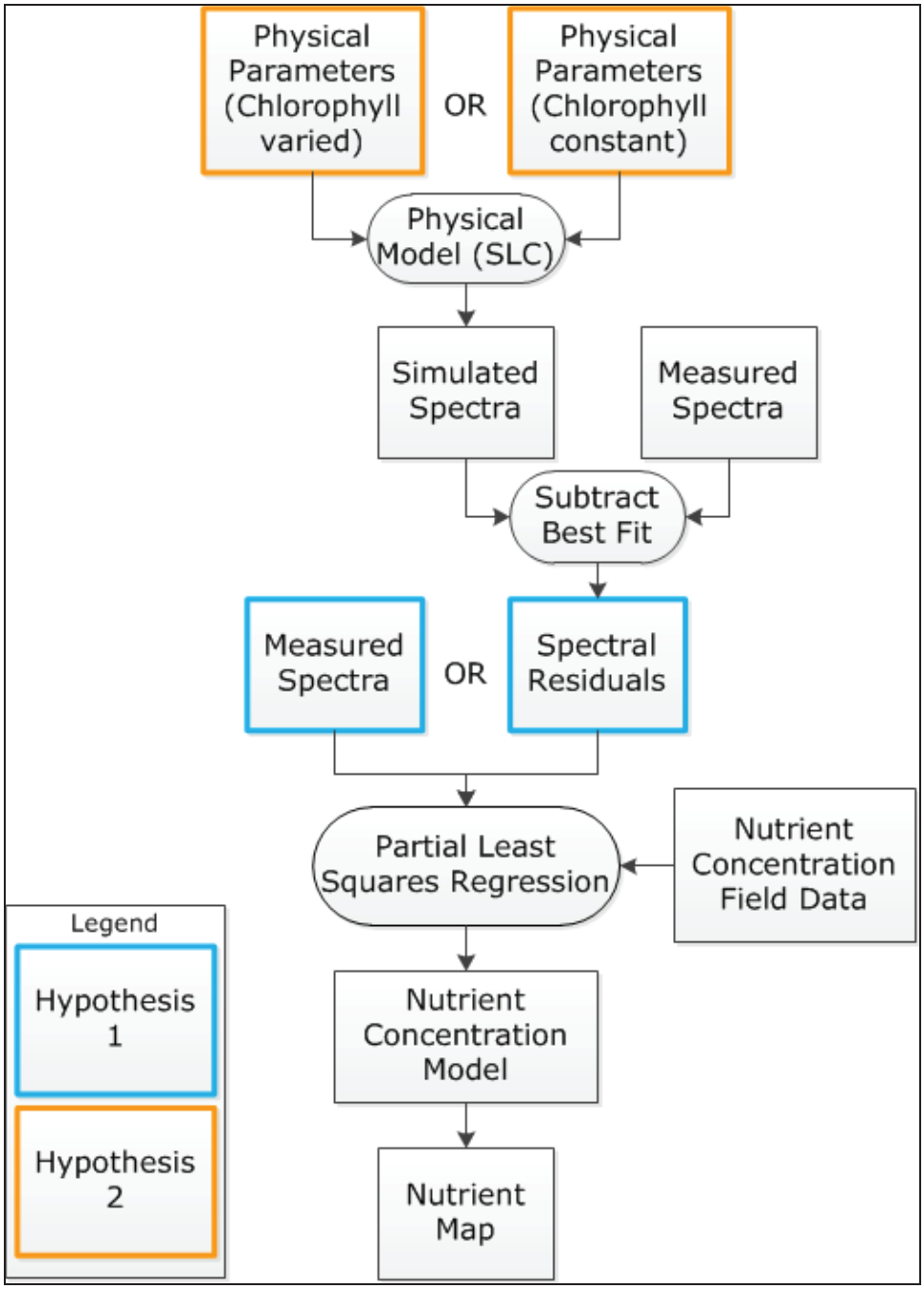

Figure 7 - Method flowchart. Matching colored boxes show alternative hypotheses being tested.

2.5 Partial least squares regression model

PLSR was then performed to find the relationship between the spectral residuals and the foliar nutrients collected in the field (N, P, K, Ca, Mg and Na). PLSR was also applied to the raw spectra and to spectra to which an MNF transformation had been applied.

Because of the limited number of observations compared to variables (in this case 113, as each spectral band is a variable), there is a danger of model instability. An unstable model is one where a small change in an input parameter could cause a large, unpredictable change in the model output. A common rule of thumb when using a regression model is to have 10 to 20 times more observations than variables (Skidmore et al., 1997). In order to achieve this ratio, it would be necessary to reduce the hyperspectral images to approximately five components or bands, which defeats the purpose of having a hyperspectral image in the first place. To avoid this problem, this study employed the use of PSLR, which does not require more observations than variables (Haenlein & Kaplan, 2004). This method also has the added benefit of reducing the impact of multicollinearity (Wold et al., 2001). Furthermore, PLSR is one of the most common regression methods in ecosystem modeling (Keithley et al., 2009) which will enhance the relevance of the results.

PLSR works very similar to the PCA/MNF methods discussed in section 1.1.3. However, instead of finding new reference frames based on maximum variance, PLSR reprojects both the input variables and the respondent variables and finds the maximum covariance between the two using linear regression. The resulting model contains a PLS regression coefficient for each input variable, which describes the weight or importance that variable has on the output.

Cross-validation was used to validate the model because of the limited amount of ground observations. Permanently splitting the data into trial and validation data sets would reduce the sample size of the trial set, which is already relatively small compared to the number of variables, and could risk overfitting the model. One common method of cross validation is “leave-one-out” cross validation, where a single sample point is set aside as the test set while the rest of the data is used to train the model. This process is then repeated for the each sample point. However, this method is not recommended because of fundamental statistical problems (Shao, 1993). Instead, this study employed 10-fold cross validation because it has been shown to provide the best balance of bias versus variability in small samples (Braga-Neto & Dougherty, 2004; Kohavi, 1995). Since the sets are randomly selected, each model could

theoretically be different from the next. Therefore, the 10-fold cross validation was repeated with 100 iterations. Results were then averaged. Comparisons between the results were based on the coefficient of determination (R^2) as a measure of variability, and normalized root mean squared error (nRMSE) as a measure of precision.

Before running a PLSR, two parts of the model must be calibrated to the data. Since PLSR can operate with multiple respondent (Y) variables at once (Wold, et al., 2001), one must determine whether the Y variables are correlated or not. Second, it is essential to determine the number of PLS components to use in the regression.

2.5.1 Model calibration: Y variables

PLSR can analyze models with multiple Y variables separately or together. When the Y variables are correlated with each other, it is important to analyze them together, otherwise information might be lost. However, if the variables are independent of one another, analyzing them separately will make interpretation of each variable simpler (Wold, et al., 2001).

Measurement of the correlation between the Y variables in this study was carried out by principle component analysis (PCA). In this case, PCA was not used to reduce noise (as discussed above), but instead used for variable analysis. For this purpose, PCA still transforms data into a new reference frame based on maximum variance of the measured data and determines the weights to be applied to each original value in order to transform it to the new dimensional space. These weights are called “loadings” or “eigenvalues”. Now, instead of removing variables with small loadings, analysis was done to determine correlation between the variables. Variables which are correlated will have similar loadings (Gil-Perez, et al., 2010).

Figure 8 shows the loadings for each of the six nutrients in this study into the first three components of the PCA. The only nutrients which could be correlated are Mg and Na, but then only in the Berau study area. Because of the lack of any strong correlation between the nutrients in this study, PLSR was applied individually to each nutrient.

However, as mentioned in section 1.1.2 of this study, many studies have noted the strong correlation between nutrients and genera. For this reason, information on the genera for each field sample was included in the regression.

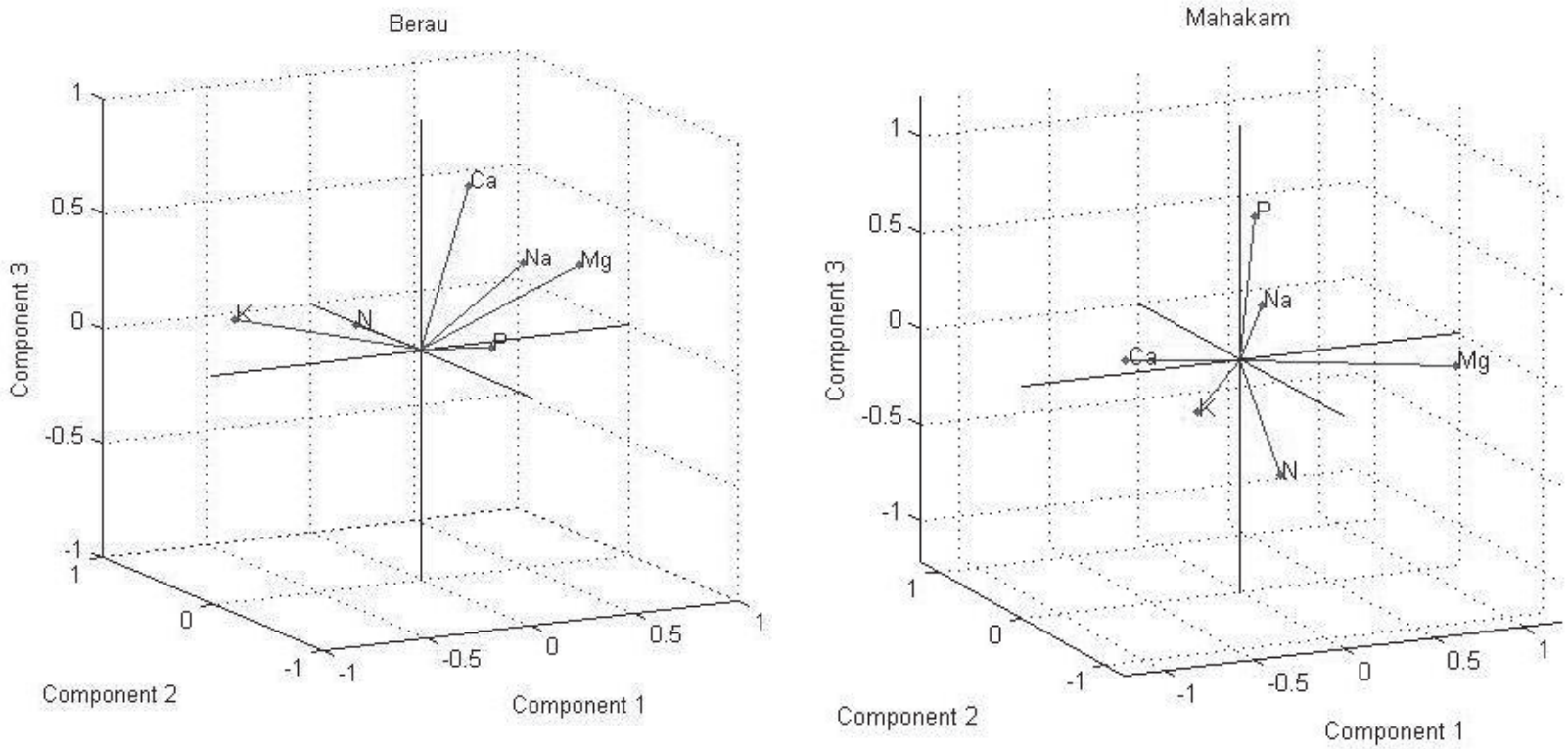

Figure 8 – The PCA loadings for each nutrient in the first three PCA components. Nutrients with similar PCA loadings would be considered correlated. However, the loadings of all nutrients are dissimilar, and therefore PLSR was applied to each nutrient separately.

2.5.2 Model calibration: number of PLS components

Choosing the correct number of components in the PLSR model is important to avoid the problem of overfitting, where the model fits the measured data well but fails to accurately predict anything outside the measured data (Wold, et al., 2001). The ideal number of PLS components can be found by using cross validation (M. Clark & Cramer, 1993; Wakeling & Morris, 1993). In this study, the data was split into 10 groups; nine groups were used to generate the model with the remaining group used as validation. The groups were the rotated so that each group was used for validation and there was a predicted and measured value for each sample point. This process was done first assuming the number of components was one, and then continued iteratively up to a sufficiently high number of components to cover the range of possibilities (this study used 40). The predictive residual sum of squares (PRESS) was calculated for each number of components using the following expression:

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where y_i is the i^{th} observed value and $\hat{y}_i$ is the i^{th} predicted value. The ideal number of components (ncomp) was then calculated via the following expression:

$$ideal\ ncomp = \min\left(\frac{PRESS}{N - A - 1}\right)$$

Where N is the number of observations and A is the number of PLS components used in that iteration. After repeating this process 100 times, PLSR was run again using the calculated ideal number of components to generate concentration predictions. R^2 and nRMSE were calculated for each iteration and averaged.

2.6 Generating nutrient maps

Nutrient maps were generated by applying the model to each pixel in the study area images. To avoid SLC being applied to shrimp ponds, water or other non-mangrove areas, a forest mask was generated. Unsupervised classification was run on the images, and each category was manually determined to be either vegetation or non-vegetation (shrimp ponds, water, etc). All non-vegetation categories were then masked out. Classification was done by the previous research group

(Axelsson, 2011) and the mask was applied to the images by this study using ENVI.

After the forest mask was completed, the SLC physical model was applied to the entire image, creating a simulated spectrum for each pixel. The complete simulated image was then subtracted from the measured image to create a third image of residuals.

Generating nutrient maps proved to be computationally very expensive. A single map using all of the spectral bands would have taken on the range of one to two months to compute from start to finish. For this reason, nutrient maps were only made for the best performing models and only using a limited set of bands.

Images were reduced to a selection of the most important bands (which were determined by plotting the PLS regression coefficients for each band). The importance of an input variable to the respondent variable is proportional to the magnitude of the PLS coefficient (Wold, et al., 2001). Bands with the highest coefficient values have the greatest influence on the nutrient concentration prediction and the most important variables combined can therefore be used as an approximation of the entire spectrum. Coefficient values above a user-defined threshold can be considered important (Gomez et al., 2008).

PLSR repeated 10-fold cross validation was then run again using only the selected wavelengths to get new PLS coefficient values. The model whose R^2 was closest to the mean R^2 of the dataset was then selected to be applied to the image. The PLS coefficients were then applied to the residuals image to generate an estimated nutrient concentration for each pixel. These processes were carried out using Matlab scripts which were generated for this study.

3 Results

Four sets of residuals (figure 9) were generated from the extracted spectra; for both of the study areas, one dataset was derived from spectra simulated with chlorophyll constant and another with chlorophyll varied. The difference between the four datasets is not substantial, with each dataset following a similar pattern. However, overall the variance of the residuals from Berau was consistently larger than the residuals from Mahakam. In all sets, the visible, near infrared (NIR) and shortwave infrared (SWIR) segments of the spectrum are distinct with their own attributes.

The residuals are all smallest in the visible spectrum (between 0.4 and 0.7 µm). However, in the visible there is slightly greater variation in the residuals from chlorophyll constant than chlorophyll varied. The greatest variance is found in the NIR, and while this section of the spectrum follows a noticeable trend, it also harbors the most noise. Variance is relatively small in the SWIR. In Berau, the physical model systematically over-predicts the reflectance, leading to negative residuals.

All four sets contain large bumps around 1.4 and 1.8 µm, which correspond to data gaps due to atmospheric water absorption features at those wavelengths (Cocks et al., 1998). The HyMap camera does not have sensors covering these sections.

3.1 Model results

3.1.1 Nitrogen concentration

Table 3 – Average of 100 nitrogen prediction results.

Input dataset	R^2	RMSE	nRMSE
Mahakam CHL varied	0.462	0.155	18.2
Mahakam CHL const	0.575	0.136	16.0
Mahakam raw	0.446	0.156	18.4
Mahakam MNF	0.482	0.151	17.8
Berau CHL varied	0.257	0.266	16.2
Berau CHL const	0.295	0.253	15.4
Berau raw	0.368	0.242	14.8
Berau MNF	0.115	0.284	17.3

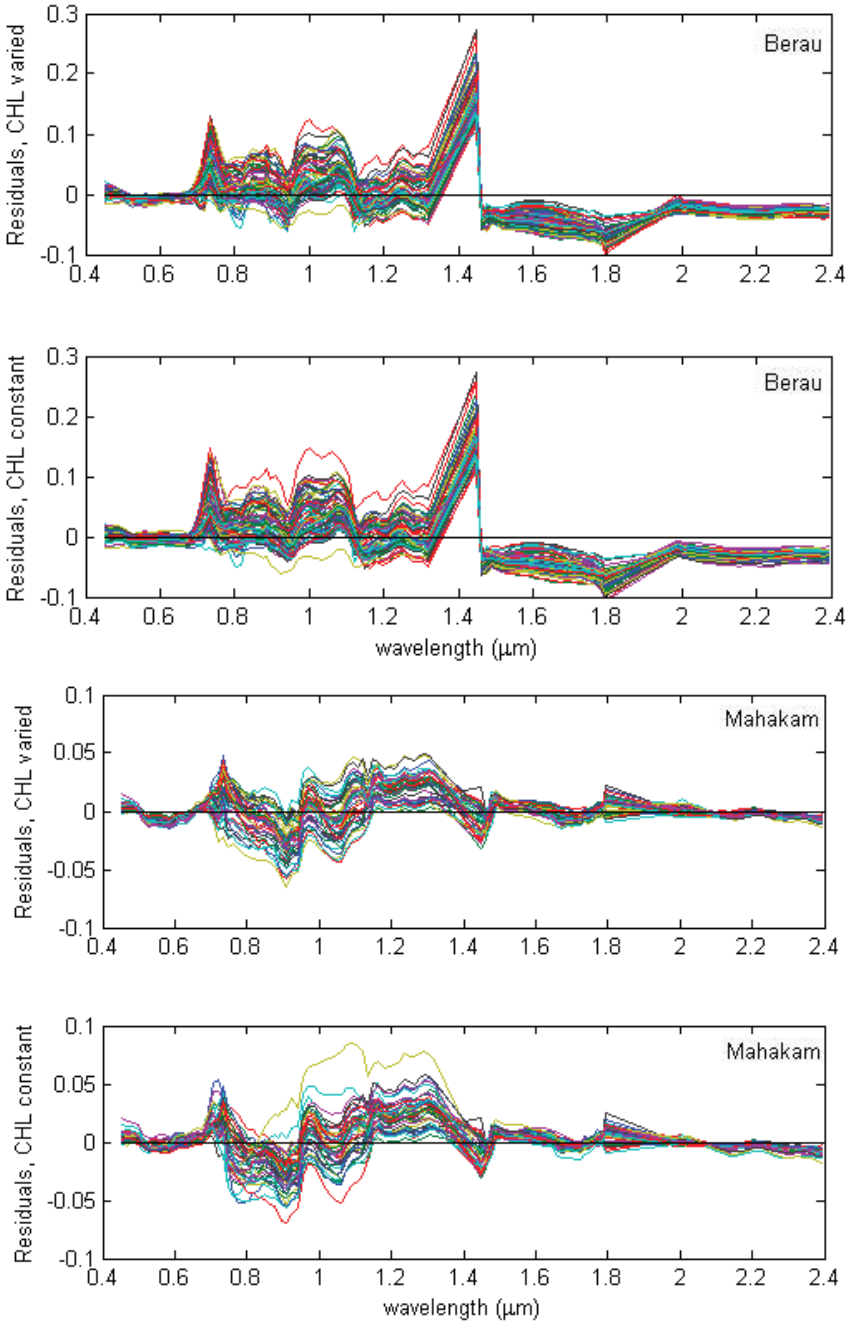

Figure 9 - Residuals (from measured minus simulated spectra) for all of the Berau (top two) and Mahakam (bottom two) sample points.

While all of the datasets in this study performed poorly, N was the only nutrient with somewhat fair results from the model (see full results in table 3). The best performing dataset for Mahakam was the chlorophyll constant residuals (R^2=0.57, nRMSE=16%). Raw spectra performed best in Berau (R^2=0.37, nRMSE=15%). All of the R^2 values are presented as histograms in figure 10. Several of the sets of R^2 results showed non-normal distributions (see also appendix A, figures 19 through 23 histograms of other nutrients), so a Kruskal-Wallis test was used to test the significance between means (Kruskal & Wallis, 1952). The Kruskal-Wallis test is the non-parametric equivalent of an ANOVA test.

Significance testing of the R^2 values of Mahakam N shows that the chlorophyll constant dataset significantly outperforms the other three models ($p<0.01$). At this significance level, the only datasets which were not statistically different in Mahakam were chlorophyll varied and raw spectra (see figure 11). All of the datasets for Berau were statistically significantly different. In both study areas, chlorophyll constant significantly outperforms chlorophyll varied.

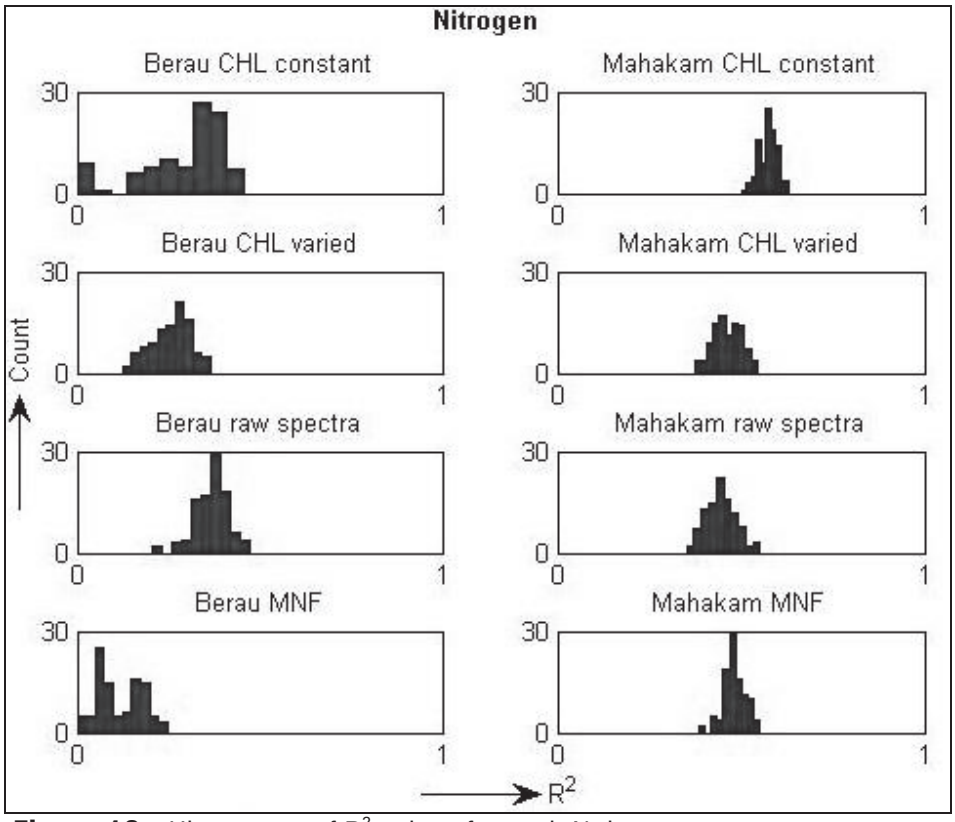

Figure 10 - Histograms of R^2 values for each N dataset.

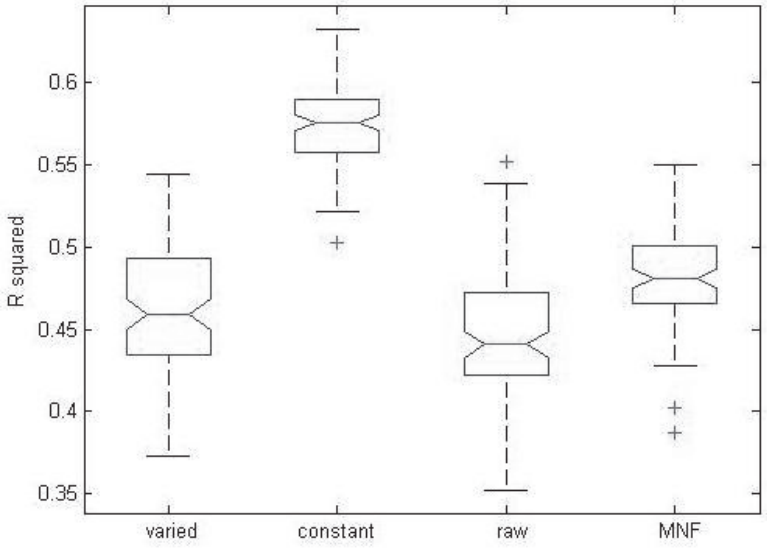

Figure 11 – Results from the Kruskal-Wallis test for nitrogen model R^2 values in Mahakam. The box plot shows that the chlorophyll constant dataset has a significantly higher mean R^2 than the other datasets. There is no statistical difference between chlorophyll varied and raw spectra

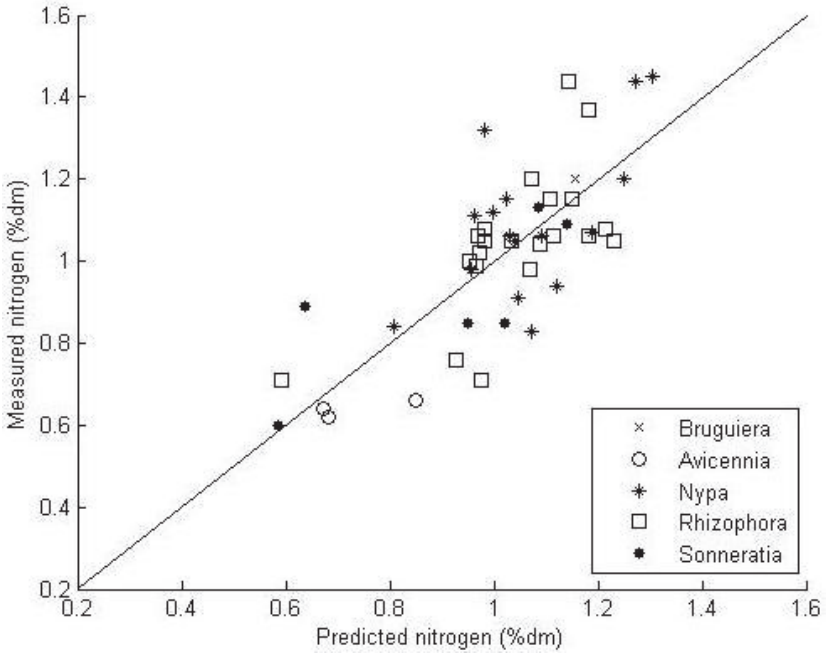

Figure 12 - Nitrogen predictions for Mahakam, chlorophyll constant dataset. R2=0.574 and nRMSE=15.9%. This scatter plot is based off of the results of a single PLSR model, not the average of all 100 iterations. This model was selected because it had the closest R^2 to the mean R^2 of the dataset.

Scatter plots of predictive accuracy were generated for the models. Because averaging prediction results risked overfitting the model, the iteration with an R^2 closest to the mean R^2 of the dataset was used (see example figure 12). Scatter plots showed that predictions were not clumped by genera.

3.1.2 Non-nitrogen nutrient concentrations

While there is some variation in quality, PLSR failed to produce accurate models for all of the remaining nutrients. A complete set of the histograms for each dataset within each model can be found in appendix A. Overall, using the residuals improved results in Mg and Na but not K, while it had little effect, if any, in Ca and P (table 4).

The models for P were the least accurate, and cannot be considered better than simply using the average of the measured values as a prediction. Ca proved only slightly better. In Mahakam here was no statistical difference between any of the models for Ca, while in Berau the difference was significant statistically, but not practically.

Using residuals to model K was significantly worse than using raw or MNF transformed spectra. However, there was no significant difference between raw and MNF for either Berau or Mahakam.

Table 4 - Results showing the best input dataset (based on lowest nRMSE scores) for each nutrient in each study area.

Nutrient	Study Area	Spectra type	R^2	nRMSE (%)
N	Mahakam	Residuals (CHL const.)	0.575*	16.0
	Berau	Raw Spectrum	0.368*	14.8
P	Mahakam	Residuals	0.024	21.8
	Berau	Raw Spectrum	0.083	19.7
K	Mahakam	Raw Spectrum	0.282*	20.7
	Berau	MNF Spectrum	0.181*	17.8
Ca	Mahakam	MNF Spectrum	0.108	31.6
	Berau	Residuals	0.128*	19.9
Mg	Mahakam	Residuals (CHL varied)	0.264*	21.6
	Berau	Residuals (CHL const.)	0.277	22.4
Na	Mahakam	Residuals	0.154*	25.4
	Berau	Residuals	0.382*	20.6

* denotes models where the R^2 from using residuals was either significantly larger or significantly smaller than the alternative models ($p<0.01$).

Mg was, like N, tested for a correlation with chlorophyll. While using residuals produced the most accurate models for both sites, the chlorophyll varied dataset was best in Mahakam while in Berau there was no significant difference between the two residual datasets. In Mahakam, models for chlorophyll varied were significantly better than all other datasets. Again, the model with the R^2 closest to the mean R^2 of the dataset was plotted with the genera included (figure 13). It also shows predictions, with all of the genera distributed widely and no obvious clumping.

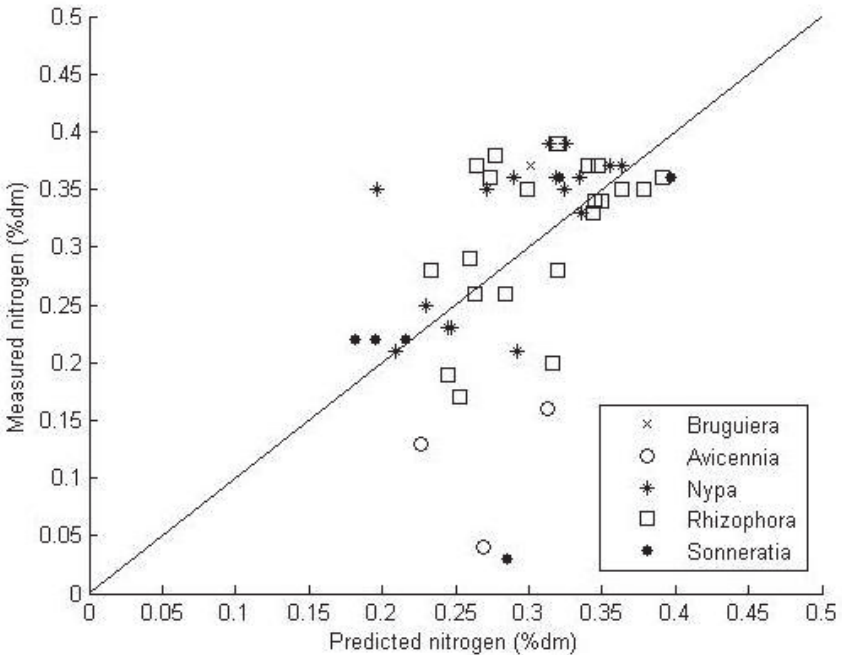

Figure 13 – Magnesium predictions for Mahakam, chlorophyll varied dataset. R^2=0.260 and nRMSE=21.6%.

3.2 Nutrient maps

Model performance was best for N, so N concentration maps were made using the chlorophyll constant models. This study used thresholds of one and five for Mahakam and Berau, respectively to select significant wavelengths (see figure 14). 15 wavelengths were selected for Mahakam, while 22 were selected in Berau. For Mahakam results, there was almost no change to the predictive power of the model (with 15 bands R^2 changed from 0.575 to 0.582 and nRMSE changed from 0.136 to 0.140) and the results were not statistically different ($p>0.01$). With Berau, the R^2 improved significantly (R^2 from 0.295 to 0.380 ($p<0.01$) while the nRMSE stayed almost the same, changing only from .253 to .256).

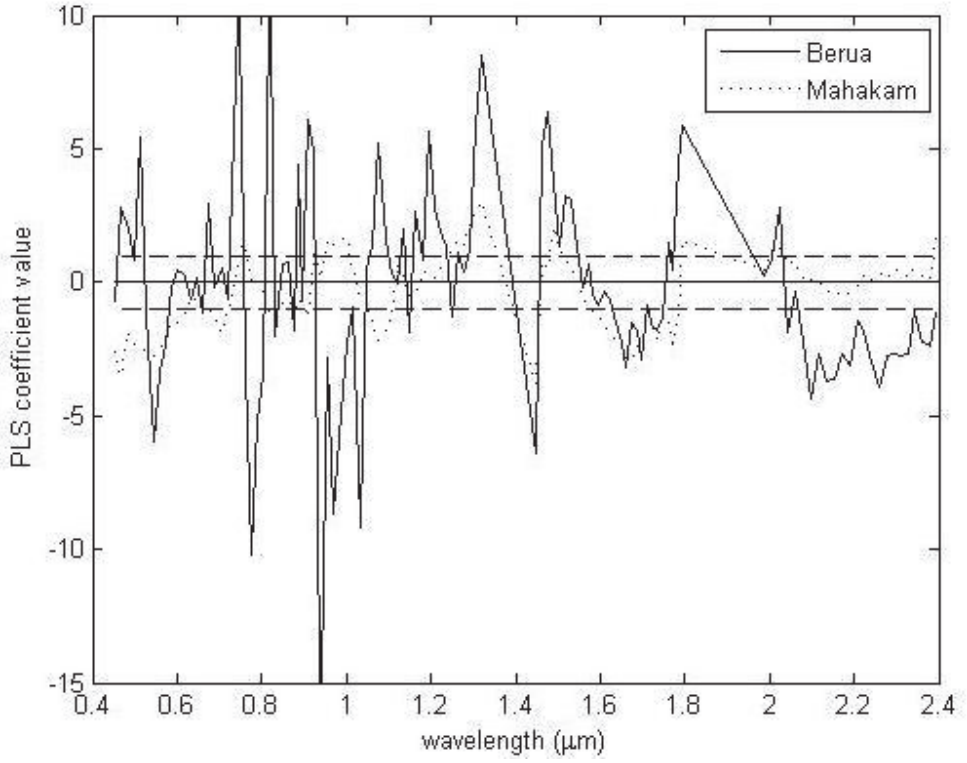

Figure 14 - Plot of the PLS coefficients for both study areas for models with chlorophyll kept constant. Peaks greater than the absolute value of one were used to select wavelengths for final regression model.

Table 5 - Wavelengths (µm) with strong correlations to N concentration, selected from each study area

Berau	0.469, 0.544, 0.676, 0.748, 0.819, 0.909, 0.941, 1.032, 1.077, 1.136, 1.193, 1.447, 1.476, 1.597, 1.662, 1.699, 1.796, 2.025, 2.100, 2.189, 2.260, 2.393
Mahakam	0.469, 0.544, 0.705, 0.762, 0.819, 0.875, 0.909, 1.002, 1.077, 1.319, 1.490, 1.662, 1.773, 2.025, 2.393

Nearly all of the wavelengths in the visible and SWIR spectrum listed in table 5 can be linked to known absorption features involved with N (Card et al., 1988; Cho & Skidmore, 2006; Curran, 1989; Serrano et al., 2002). In the NIR, there were several wavelengths with high coefficient values that do not correspond to any known N-related features. There were 0.762, 0.819, 0.875 and 1.077 in Mahakam and only 0.819 and 1.077 in Berau (all units in µm).

The final nutrient maps can be seen in figures 15 and 16. Black patches in both maps are sections removed due to cloud cover and grayscale rectangular objects are shrimp ponds. Ponds are scattered throughout Mahakam, but are most prominent the southeast corner. Areas of low N concentrations are most noticeable along the coastline (along the west). Areas with N concentrations higher than average are located almost exclusively in large forest stands.

The map for Berau has a clearly unnatural “patchwork quilt” look. Horizontal (east-west) striping from flightpaths are obvious throughout the image. Furthermore, there are also some artifacts in the image creating north-south boundaries within each flightpath. Due to the very prominent patchwork features in the image, it is impossible to link concentration data to any physical features on the map.

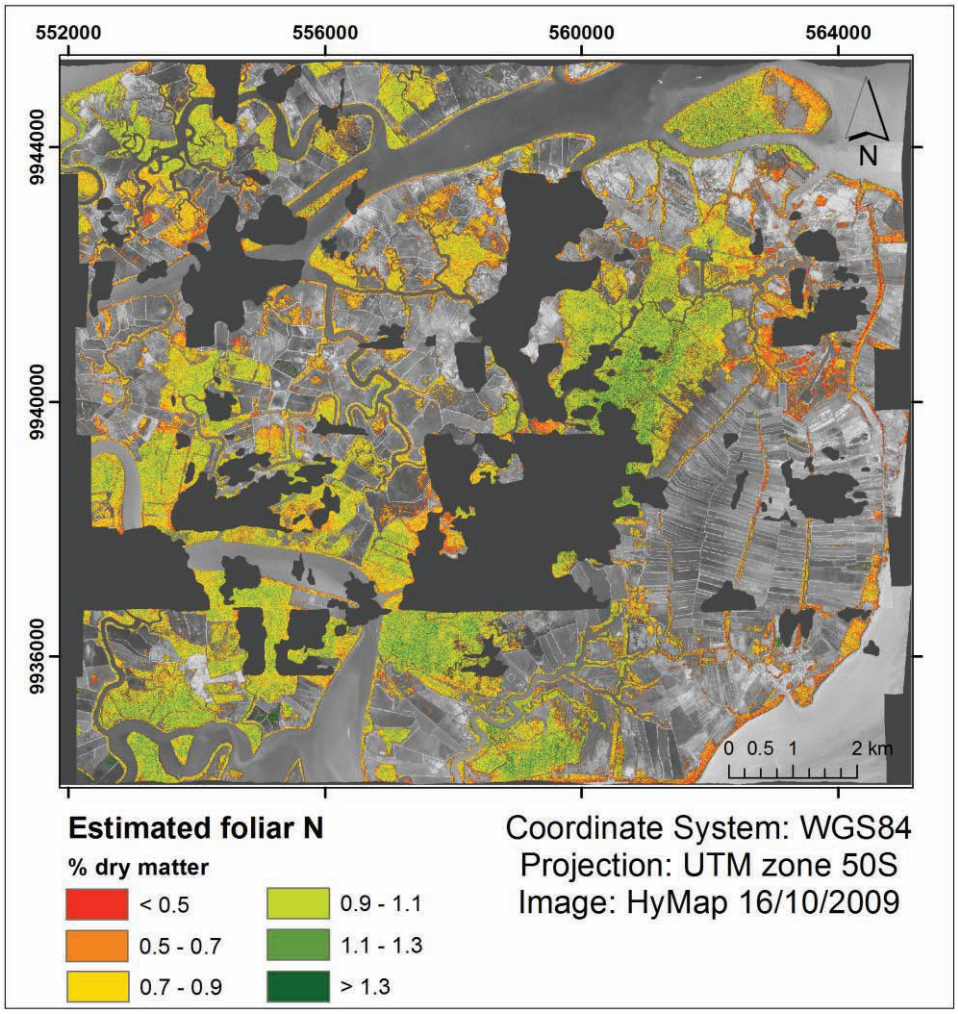

Figure 15 - Map of nitrogen distribution in Mahakam

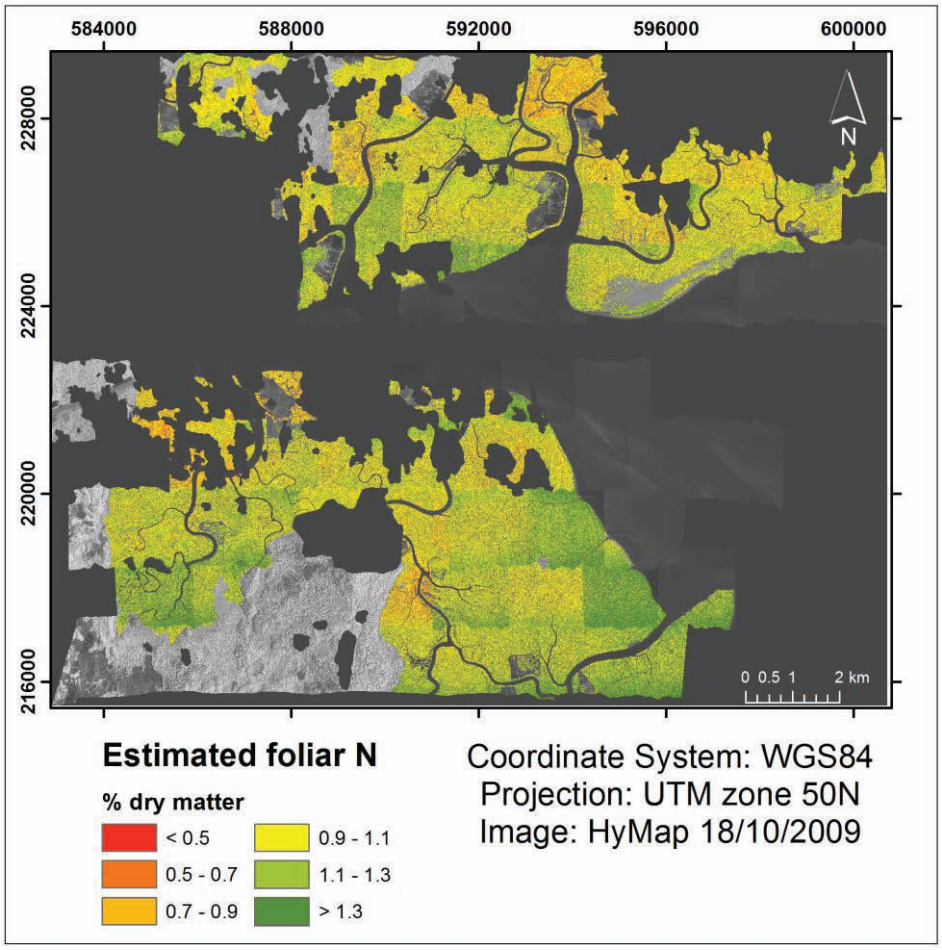

Figure 16 – Map of nitrogen distribution for Berau.

4 Discussion

The patternistic nature of all four sets of residuals indicates that they indeed contain more than just noise. If the physical model was perfect, and simulated spectra were perfect reproductions of reflectance spectra, then the residuals would consist of only random noise. However, residuals show systematic differences between simulated and measured spectra, both within a single dataset and between datasets.

This demonstrates that the SLC model is consistently mis-predicting reflectance and is therefore not an exhaustive physical model. There are other unknown parameters which could be used to create a more complete model. However, the residuals do not necessarily indicate that foliar nutrients are those missing parameters, nor does it indicate that a correlation can be found between nutrients and residuals.

The difference in the residuals between the chlorophyll constant dataset and the chlorophyll varied dataset also shows that chlorophyll's main impact is on the visible spectrum. While this comes as no surprise, it does demonstrate that excluding chlorophyll variation from the physical model successfully kept any information about N and Mg that is correlated with chlorophyll in the residuals.

The difference in the residuals between the visible, NIR and SWIR also demonstrates that the SLC is poorest at predicting reflectance in the NIR. Interestingly, the majority of wavelengths in this study with high PLS coefficients that are not linked to known N-related absorption features are also found in the NIR. This could indicate that the impact of N (or N related biochemicals) is not fully understood in the NIR.

4.1 Criticisms of data quality

Before properly analyzing the potential correlation between foliar nutrients and reflectance residuals, it is important to acknowledge that there were several intractable issues with the data in this study. The first and most obvious problem lies in the collection of data in Berau. Although it has been noted that the Berau data was collected in the same season as the Mahakam data and image data, there was still a one year difference. Seasonal trends have been found in N concentrations of *Rhizophora spp.*, this does not mean that annual trends do not exist (Y. M. Lin et al., 2010). The time lag between data collection casts serious doubts on any conclusion drawn from the

Berau data, and makes it impossible to make a clear comparison between Berau and Mahakam concentrations.

There are also problems with the Berau image data. As was discussed in section 2.3 of this paper, problems were discovered in the initial radiometric correction of the image data by HyVista and the reflectance in the visible was demonstrated to be unnaturally high. Radiometric corrections were implemented but they have clearly introduced an artifact into the data that raises serious question to any conclusions drawn from the data.

A third issue arises when looking at the R^2 results from Berau. When the number of bands was reduced from 113 to 22, the R^2 increased from 0.295 to 0.380. When the accuracy of a model changes from band reduction, it means that something impacting the model was removed. Usually, this means either noise or multicollinear bands. For PLSR, improving results via band reduction is contrary to expectation, as PLSR should not be impacted by multicollinearity. Removing bands with low PLS coefficients should only remove redundant information or information that has little to no impact on the spectra. The model results from Mahakam, for example, did not change with band reduction. Furthermore, PLSR should actually accommodate noisy bands by relegating them to insignificant eigenvectors (Wold, et al., 2001), so reducing these bands should have little or no impact on model accuracy.

The improvement in the Berau model means that some of the bands in the data were actually having a significant negative impact on the model accuracy. Whatever the cause, this clearly indicates a problem with the data. Possibly, removing the bands was removing some of the artifacts caused by the radiometry correction mentioned above.

All of this leads to the conclusion that the Berau data is too flawed to derive any useful conclusions. At the very least results from Berau must be viewed with extreme skepticism. However, these issues do not apply to the Mahakam data.

Another problem with the data, not directly related to Berau, is the quality of the Ca measurements. Analysis of the variance of nutrient measurements shows questionable results for Ca (see figure 17). The variance of the measurements in Berau is extremely high, and does not correspond to the Ca measurements in Mahakam. All of the other nutrients have similar means between the two study areas and generally overlap in the interquartile ranges. Unfortunately, there is not detailed documentation about the chemical analysis of

the ground sample leaves from the field work of this study. Atomic absorption spectrophotometry is a commonly used method in many fields and there are no indications that calcium should pose any unique problems. The variance does not mean necessarily that the Berau measurements are incorrect and the Mahakam measurements are correct. It does mean that the Ca results are suspect and should be used with caution. A previous study of mangrove forests found Ca concentrations between 0.9% and 2% dry matter, which corresponds the mean of both study sites. Measurements of other forests have found Ca concentrations up to 7% of dry matter (DeHayes et al., 1999; Gil-Perez, et al., 2010).

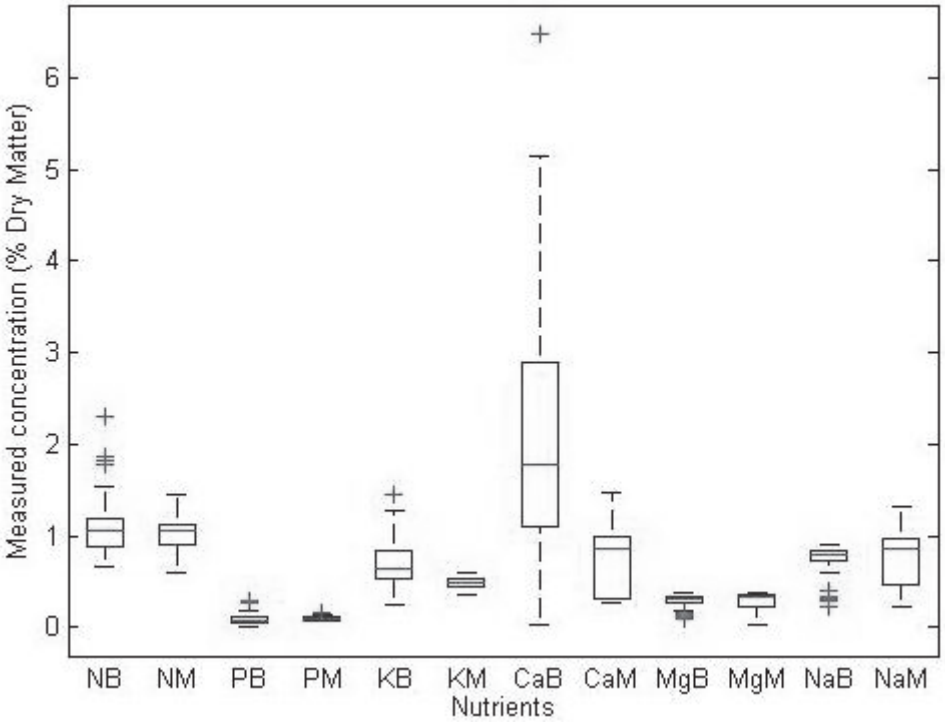

Figure 17 – Distribution of measured nutrient concentrations. NB signifies "Nitrogen, Berau", NM signifies "Nitrogen, Mahakam", etc. Nutrient concentrations between the two study sites are highly correlated, except Ca, which indicates there was potentially a problem with the measurement of Ca.

The final problem present in the data regards the small number of field observations. Certainly, there are often arguments for a greater number of observations in almost any study. Furthermore, there are several examples of studies finding strong correlations between nutrients and reflectance spectra using a similar number of observations (Ferwerda & Skidmore, 2007; Mutanga, et al., 2004).

However, the results of this study could have benefitted dramatically with more samples for reasons regarding PLSR.

Figure 18 demonstrates the relationship between the number of PLS components used in the PLS regression and the percentage of variance that is explained by the model. In this particular example, the ideal number of components for the model (i.e. the number which resulted in the lowest PRESS/(N-A-1) was three. Using three components, the PLSR model was only able to explain about 40% of the variance in the model. This means that 60% of the information in the data was lost, because of the number of components parameter. As mentioned in the section 2.2 of this study, field samples were selected with the intention of maximizing the variance of nutrient concentrations in the ecosystem. Therefore, gathering more samples should not dramatically increase the amount of variance that needs to be explained by the model. However, increasing the sample size would increase the number of components that could be used in the model. It is clear from the figure that increasing the number of components in the model by even a few components would mean the model could incorporate a substantially greater percentage of the variance in the data.

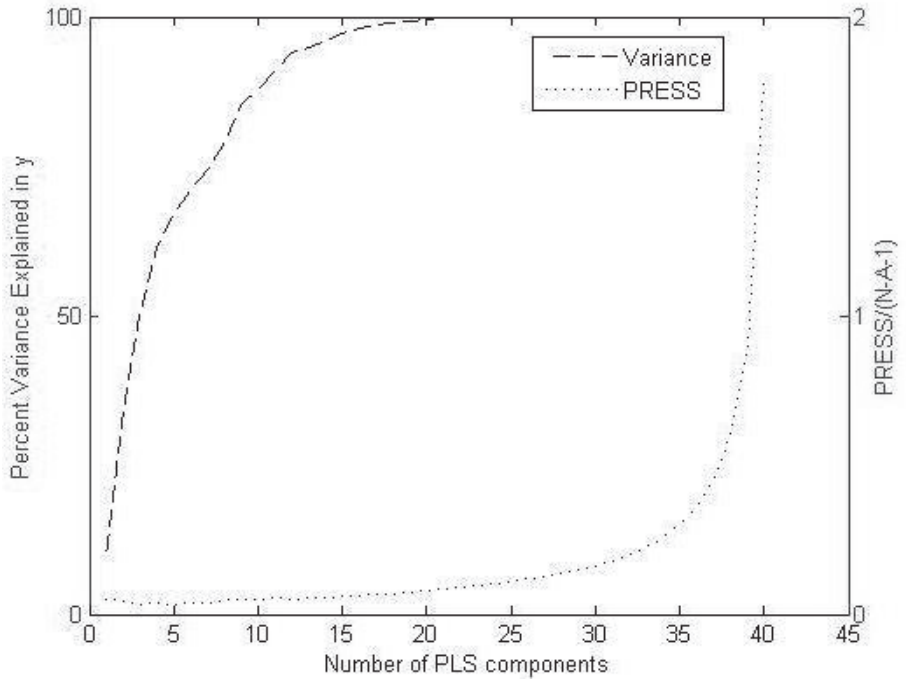

Figure 18 - Effect of PLS components on the percentage of variance explained by PLSR models. This example is from the K model of Mahakam raw data. It demonstrates that a low percentage of the information available in the data was actually used in the regression.

4.2 Relationship between residuals and nutrients in Mahakam

There is no definitive pattern among the nutrients regarding which dataset resulted in the best predictive model. The models for P and Ca not only failed to differentiate between the datasets, but actually failed to generate useful predictive models at all. The most likely reason for this is the extremely small concentrations of P in mangrove leaves. P concentrations were several times smaller than the other nutrients and an order of magnitude smaller than N concentrations. At these concentrations they may simply not impact the reflectance spectra in a detectable way.

The remaining nutrients give a good indication that applying regression to spectral residuals can improve model predictive power. Models for N, Mg and Na all significantly improved with the use of spectral residuals. The model for K, however, was significantly worse using spectral residuals. It is not clear why K models performed worse with the residual dataset. The best guess is that K is strongly correlated with one of the parameters in the physical model and K concentration information is lost in simulated spectrum subtraction. K is not a component of any organic matter, but plays a role in cell processes like regulating stomata (evapotranspiration) and cell turgor (Eakes et al., July 1991). Possibly K concentration is linked to leaf moisture content in the physical model.

When investigating the usefulness of keeping chlorophyll constant, only Mg and N were considered, as they are the only two nutrients in the study that are contained within the chlorophyll molecule. This study demonstrated that keeping chlorophyll constant in the physical model significantly improved the regression model for N, but not for Mg. This indicates that the spectral correlation between N and chlorophyll is stronger than Mg with chlorophyll. This supports a study performed on wheat plants which investigated the impact of N and Mg deficiencies on foliar reflectance (Ayala-Silva & Beyl, 2005). While a correlation was found between Mg deficiency and chlorophyll absorption (when deficient, reflectance in the visible increased by 25-36%) the correlation was significantly greater for N (reflectance increased by 75-97%). Perhaps smaller correlation between chlorophyll and Mg means that the tradeoff between reducing “noise” (physical spectrum removal) and maintaining signal (keeping chlorophyll constant) leans in the favor of reducing noise.

Surprisingly, MNF transformed data rarely resulted in better models than raw spectra. MNF was chosen under the assumption that it

would provide a consistently better model than raw data. It is unclear why the MNF performed so poorly, but it is possible that because the nutrients have such small impact on the original reflectance spectra, some of this information was lost when the MNF transformation discarded most of the noisy bands.

4.3 Nutrient maps

The Mahakam nutrient map does not demonstrate any correlation between shrimp ponds and N concentrations. This is partially because shrimp ponds are scattered throughout the entire image, and there are therefore very few parts of the map that could be considered isolated from the ponds. However, there is also no apparent correlation between the local density of shrimp ponds and N concentrations. The area of highest shrimp pond density (encompassing much of the southwest quadrant of the map) is adjacent to both the areas of highest N concentration and lowest N concentration.

The scatter plots of model predictions also demonstrate that the correlation between nutrients and genera has been successfully removed in the PLSR process. This means that the N maps are mapping N concentration and not just genera distribution.

The distribution of N concentrations does appear to be impacted by two factors; proximity to the coast and the size of the forest stand. The areas of lowest N concentrations are along the western coast facing the sea (or along the main river banks), while the area of highest N concentration is also the largest forest stand. Low values on the coast are possibly due to the wave action of the ocean. While mangroves thrive in tidal inundation, the higher energy of waves are more likely to disturb soil, remove nutrients and hinder mangrove growth (Kathiresan & Qasim, 2005).

It should be noted that it is beyond the scope of this study to determine the healthy range of N concentrations for the Kalimantan mangroves, which, of course, depends on each genus as each has different biochemical needs (Siddiqi, 1995). However, the ranges measured in this study do correspond with, on the low end, ranges measured in other studies (P. Lin & Wang, 2001; Wang et al., 2003). These studies observed many of the same mangrove genera as this study, in China and the USA. Therefore, this map does not support the claim that nutrients from shrimp ponds are having an effect on the surrounding mangroves. However, this map does demonstrate that higher forest fragmentation is correlated with lower N levels. Forest fragmentation can lead to species loss and genera loss in

tropical forests (Laurance et al., 2006) and may be a larger threat to the Mahakam mangroves than nutrient leaching.

5 Conclusions

This study has demonstrated the viability of a novel hybrid model method for predicting nutrient concentrations from remote sensing. PLS regression on spectral residuals from the SLC physical model significantly improved model predictions in three (N, Mg and Na) out of the six nutrients tested for Mahakam. The data from Berau was deemed too flawed to support conclusions.

For N and Mg, physical model residuals were generated via two methods, with the intention of determining if correlations with chlorophyll would impact the modeling of nutrients. This study found that removing chlorophyll from the physical model improved the model results when modeling N using spectral residuals (from R^2=0.462 to R^2=0.575 and from nRMSE=18.3% to nRMSE=16.0%) but diminished model results when modeling Mg using spectral residuals (from R^2=0.263 to R^2=0.156 and from nRMSE=21.6% to nRMSE=23.2%).

This novel method could have broad uses for nutrient modeling. It also could be applied generically in any remote sensing field where physical models have been developed but unknown parameters still exist. The method has also demonstrated the ability to help highlight previously unknown relationships between predictor variables (reflectance) and respondent variables (nutrient concentrations).

N concentration maps were also created using models derived from the residuals with chlorophyll kept as a constant. Because genera were included as respondent variables in the PLS regression, these maps were not impacted by the correlation between genera and nutrient concentrations. The maps were used to investigate a possible spatial correlation between shrimp ponds and N concentrations. No correlation was found, but correlations between N concentrations and both stand size and proximity to coastline were noticed.

6 Recommendations

There are several improvements which could be made to this study to achieve higher model accuracy and more relevant results. The most obvious improvement would be the inclusion of different data. Of course, having a dataset that would not manifest as many problems as Berau would help analysis. However, it might be even more beneficial to also include a biochemical that has a well established record of being successfully modeled. If, for example, this method was used to model chlorophyll concentrations (by keeping chlorophyll constant in the physical model) there would be some guarantee that a statistical model would be able to achieve some predictive success. This would avoid the problem encountered with P and Ca (and to a lesser extent Na, Mg and K) where all of the datasets resulted in poor models and no strong comparison could be made.

There are also a few changes which could be made to the methods which might improve future studies in this area. Wold (2001) states that PLSR can accommodate information in the input parameters which is unrelated to the respondent variable (whether it be random noise or systematic). However, this has been contested (Cheng & Wu, 2006) and a modified PLSR has been developed which should improve on PLSR results under these circumstances.

While subtracting out the simulated spectra should remove most of the information unrelated to nutrient concentrations in the measured spectra, it is still unknown how much of the residual spectra will directly correlate to nutrient concentrations. Certainly some of the six nutrients used in this study will be less correlated to the spectra than others (P for example), and in these cases MPLSR might prove useful.

Another improvement to the method should be the replacement of the MNF transformation with different techniques to improve data quality. It is acknowledged that MNF was a poor choice for a comparison with this method. The method developed in this study is specifically designed to model parameters that have not been included in existing physical models. The types of parameters that are not included in physical models are inherently ones that have a small impact on the output variables (in this case, reflectance). MNF operates by smoothing small features in data, which could easily have led to the loss of valuable information in this study. Instead, it is recommended that future studies with this model use data

enhancement methods that do not eliminate small features, such as continuum removal or water removal.

References

Alongi, D. M. (2008). Mangrove forests: Resilience, protection from tsunamis, and responses to global climate change. *Estuarine, Coastal and Shelf Science, 76*(1), 1-13.

Asner, G. (2008). Hyperspectral Remote Sensing of Canopy Chemistry, Physiology, and Biodiversity in Tropical Rainforests *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests* (pp. 261-296): CRC Press.

Asner, G., & Martin, R. E. (2008). Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing of Environment, 112*(10), 3958-3970.

Axelsson, C. R. (2011). *Predicting mangrove leaf chemical content from hyperspectral remote sensing using advanced regression techniques.* (MSc Thesis) University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.

Ayala-Silva, T., & Beyl, C. A. (2005). Changes in spectral reflectance of wheat leaves in response to specific macronutrient deficiency. *Advances in Space Research, 35*(2), 305-317.

Blackburn, G. A. (2007). Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany, 58*(4), 855-867.

Bouillon, S., Borges, A. V., Castañeda-Moya, E., Diele, K., Dittmar, T., Duke, N. C., et al. (2008). Mangrove production and carbon sinks: A revision of global budget estimates. *Global Biogeochem. Cycles, 22*(2), GB2013.

Braga-Neto, U. M., & Dougherty, E. R. (2004). Is cross-validation valid for small-sample microarray classification? *Bioinformatics, 20*(3), 374-380.

Card, D. H., Peterson, D. L., Matson, P. A., & Aber, J. D. (1988). Prediction of leaf chemistry by the use of visible and near infrared reflectance spectroscopy. *Remote Sensing of Environment, 26*(2), 123-147.

Cheng, B., & Wu, X. (2006). An Modified PLSR Method in Prediction. *Journal of Data Science, 4*(3), 257-274.

Cho, M. A., & Skidmore, A. K. (2006). A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment, 101*(2), 181-193.

Clark, M., & Cramer, R. D. (1993). The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quantitative Structure-Activity Relationships, 12*(2), 137-145.

Clark, R. N., & Roush, T. L. (1984). Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *J. Geophys. Res., 89*(B7), 6329-6340.

Cocks, T., Jenssen, R., Steward, I., Wilson, I., & Shields, T. (1998). The HyMap airborne hyperspectral sensor: the system, calibration and performance. *Proceedings of the 1rd EARSeL Workshop on Imaging Spectrometry*, 37-42.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment, 84*(1), 1-15.

Coppin, P. R., & Bauer, M. E. (1996). Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sensing Reviews, 13*(3-4), 207-234.

Corredor, J., & Morell, J. (1994). Nitrate depuration of secondary sewage effluents in mangrove sediments. *Estuaries and Coasts, 17*(1), 295-300.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment, 30*(3), 271-278.

DeHayes, D. H., Schaberg, P. G., Hawley, G. J., & Strimbeck, G. R. (1999). Acid Rain Impacts on Calcium Nutrition and Forest Health. *Bioscience, 49*(10), 789-800.

Delegido, J., Alonso, L., González, G., & Moreno, J. (2010). Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC). *International Journal of Applied Earth Observation and Geoinformation, 12*(3), 165-174.

Donato, D. C., Kauffman, J. B., Murdiyarso, D., Kurnianto, S., Stidham, M., & Kanninen, M. (2011). Mangroves among the most carbon-rich forests in the tropics. [Article]. *Nature Geoscience, 4*(5), 293-297.

Eakes, D. J., Robert, D. W., & Seiler, J. R. (July 1991). Water Relations of Salvia splendens 'Bonfire' as Influenced by Potassium Nutrition and Moisture Stress Conditioning. *J. Amer. Soc. Hort. Sci., 116*(4), 712-715.

Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks—a review. *Pattern Recognition, 35*(10), 2279-2301.

Ellison, A. M., & Farnsworth, E. J. (1996). Anthropogenic disturbance of Caribbean mangrove ecosystems: Past impacts, present trends, and future predictions. *Biotropica, 28*(4), 549-565.

Evans, J. R. (1989). Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia, 78*(1), 9-19.

Ferwerda, J. G., & Skidmore, A. K. (2007). Can nutrient status of four woody plant species be predicted using field spectrometry? *Isprs Journal of Photogrammetry and Remote Sensing, 62*(6), 406-414.

Field, C., & Mooney, H. A. (1986). The photosynthesis-nitrogen relationship in wild plants. In T. J. Givnish (Ed.), *On the Economy of Plant Form and Function* (pp. 25-55). Cambridge, UK: Cambridge Univ. Press.

Ganapol, B. D., Johnson, L. F., Hlavka, C. A., Peterson, D. L., & Bond, B. (1999). LCM2: A coupled leaf/canopy radiative transfer model. *Remote Sensing of Environment, 70*(2), 153-166.

Ge, S., Carruthers, R., Spencer, D., & Yu, Q. (2008). Canopy assessment of biochemical features by ground-based hyperspectral data for an invasive species, giant reed (*Arundo donax*). *Environmental Monitoring and Assessment, 147*(1), 271-278.

Gil-Perez, B., Zarco-Tejada, P. J., Correa-Guimaraes, A., Relea-Gangas, E., Navas-Gracia, L. M., Hernandez-Navarro, S., et al. (2010). Remote sensing detection of nutrient uptake in vineyards using narrow-band hyperspectral imagery. *Vitis, 49*(4), 167-173.

Giri, C., Zhu, Z., Tieszen, L. L., Singh, A., Gillette, S., & Kelmelis, J. A. (2008). Mangrove forest distributions and dynamics (1975–2005) of the tsunami-affected region of Asia†. *Journal of Biogeography, 35*(3), 519-528.

Gomez, C., Lagacherie, P., & Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma, 148*(2), 141-148.

Green, E. P., Clark, C. D., Mumby, P. J., Edwards, A. J., & Ellis, A. C. (1998). Remote sensing techniques for mangrove mapping. *International Journal of Remote Sensing, 19*(5), 935-956.

Haenlein, M., & Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics, 3*(4), 283-297.

Harris, J., Ponomarev, P., Shang, J., & Rogge, D. M. (2006). Noise reduction and best band selection techniques for improving classification results using hyperspectral data: Application to lithological mapping in Canada's Arctic. *Canadian Journal of Remote Sensing, 32*(5), 341-354.

Hemminga, M. A., Slim, F. J., Kazungu, J., Ganssen, G. M., Nieuwenhuze, J., & Kruyt, N. M. (1994). Carbon outwelling from a mangrove forest with adjacent seagrass beds and coral reefs (Gazi Bay, Kenya) *Marine Ecology Progress Series, 106*, 291-301.

Heumann, B. W. (2011). Satellite remote sensing of mangrove forests: Recent advances and future opportunities. [Article]. *Progress in Physical Geography, 35*(1), 87-108.

Holguin, G., Gonzalez-Zamorano, P., de-Bashan, L. E., Mendoza, R., Amador, E., & Bashan, Y. (2006). Mangrove health in an arid environment encroached by urban development—a case study. *Science of The Total Environment, 363*(1–3), 260-274.

Holguin, G., Vazquez, P., & Bashan, Y. (2001). The role of sediment microorganisms in the productivity, conservation, and rehabilitation of mangrove ecosystems: an overview. *Biology and Fertility of Soils, 33*(4), 265-278.

Jacquemoud, S., Ustin, S. L., Verdebout, J., Schmuck, G., Andreoli, G., & Hosgood, B. (1996). Estimating leaf biochemistry using the PROSPECT leaf optical properties model. *Remote Sensing of Environment, 56*(3), 194-202.

Kambhatla, N., & Leen, T. K. (1997). Dimension Reduction by Local Principal Component Analysis. *Neural Computation, 9*(7), 1493-1516.

Kathiresan, K., & Qasim, S. (2005). *Biodiversity of mangrove ecosystems*. New Delhi (India): Hindustan Publishing.

Kauffman, J., Heider, C., Cole, T., Dwire, K., & Donato, D. (2011). Ecosystem Carbon Stocks of Micronesian Mangrove Forests. *Wetlands, 31*(2), 343-352.

Keithley, R. B., Mark Wightman, R., & Heien, M. L. (2009). Multivariate concentration determination using principal component regression with residual analysis. *Trends in Analytical Chemistry, 28*(9), 1127-1136.

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends in Ecology and Evolution, 18*(6), 299-305.

Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgasim, A. A., & Gao, F. (2000). Inversion methods for physically-based models. *Remote Sensing Reviews, 18*(2-4), 381-439.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2.

Kreb, D., & Budiono. (2005). Cetacean diversity and habitat preferences in tropical waters of East Kalimantan, Indonesia. *Raffles Bulletin of Zoology, 53*(1), 149-155.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*, 583-621.

Laurance, W. F., Nascimento, H. E. M., Laurance, S. G., Andrade, A., Ribeiro, J. E. L. S., Giraldo, J. P., et al. (2006). Rapid decay of tree-community composition in Amazonian forest fragments (Vol. 103, pp. 19010-19014). U. S. A.: Proc. Natl. Acad. Sci.

Liang, S. (2005). *Quantitative remote sensing of land surfaces*. US: Interscience.

Lin, P., & Wang, W.-q. (2001). Changes in the leaf composition, leaf mass and leaf area during leaf senescence in three species of mangroves. *Ecological Engineering, 16*(3), 415-424.

Lin, Y. M., Liu, X. W., Zhang, H., Fan, H. Q., & Lin, G. H. (2010). Nutrient conservation strategies of a mangrove species *Rhizophora stylosa* under nutrient limitation. *Plant and Soil, 326*(1), 469-479.

Majeke, B., van Aardt, J. A. N., & Cho, M. A. (2008). Imaging spectroscopy of foliar biochemistry in forestry environments. *Southern Forests: a Journal of Forest Science, 70*(3), 275-285.

Marschner, H. (1995). *Mineral nutrition of higher plants*. London: Academic press.

Martin, M. E., & Aber, J. D. (1997). High Spectral Resolution Remote Sensing of Forest Canopy Lignin, Nitrogen, and Ecosystem Processes. *Ecological Applications, 7*(2), 431-443.

Martin, M. E., Plourde, L. C., Ollinger, S. V., Smith, M. L., & McNeil, B. E. (2008). A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sensing of Environment, 112*(9), 3511-3519.

Minchinton, T. E. (2001). Canopy and substratum heterogeneity influence recruitment of the mangrove Avicennia marina. *Journal of Ecology, 89*(5), 888-902.

Mutanga, O., & Skidmore, A. K. (2004). Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment, 90*(1), 104-115.

Mutanga, O., Skidmore, A. K., & Prins, H. H. T. (2004). Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment, 89*(3), 393-408.

Nagelkerken, I., Blaber, S. J. M., Bouillon, S., Green, P., Haywood, M., Kirton, L. G., et al. (2008). The habitat function of mangroves for terrestrial and marine fauna: A review. *Aquatic Botany, 89*(2), 155-185.

Oxmann, J., Pham, Q., Schwendenmann, L., Stellman, J., & Lara, R. (2010). Mangrove reforestation in Vietnam: the effect of sediment physicochemical properties on nutrient cycling. *Plant and Soil, 326*(1), 225-241.

Parks, P. J., & Bonifaz, M. (1994). Nonsustainable Use of Renewable Resources: Mangrove Deforestation and Mariculture in Ecuador. *Marine Resource Economics, 09*(1).

Patrik, R. (1999). The ecological basis for economic value of seafood production supported by mangrove ecosystems. *Ecological Economics, 29*(2), 235-252.

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*(1), 125-135.

Powell, N., & Osbeck, M. (2010). Approaches for Understanding and Embedding Stakeholder Realities in Mangrove Rehabilitation Processes in Southeast Asia: Lessons Learnt from Mahakam Delta, East Kalimantan. *Sustainable Development, 18*(5), 260-270.

Primavera, J. H. (1998). Mangroves as Nurseries: Shrimp Populations in Mangrove and Non-mangrove Habitats. *Estuarine, Coastal and Shelf Science, 46*(3), 457-464.

Primavera, J. H. (2000). Development and conservation of Philippine mangroves: institutional issues. *Ecological Economics, 35*(1), 91-106.

Ramoelo, A., Skidmore, A. K., Schlerf, M., Mathieu, R., & Heitkonig, I. M. A. (2011). Water-removed spectra increase the retrieval accuracy when estimating savanna grass nitrogen and phosphorus concentrations. *Isprs Journal of Photogrammetry and Remote Sensing, 66*(4), 408-417.

Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications, 29*(1), 65-74.

Reef, R., Feller, I. C., & Lovelock, C. E. (2010). Nutrition of mangroves. *Tree Physiology, 30*(9), 1148-1160.

Rubin, J. A., Gordon, C., & Amatekpor, J. K. (1999). Causes and Consequences of Mangrove Deforestation in the Volta Estuary, Ghana: Some Recommendations for Ecosystem Rehabilitation. *Marine Pollution Bulletin, 37*(8–12), 441-449.

Serrano, L., Peñuelas, J., & Ustin, S. L. (2002). Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment, 81*(2–3), 355-364.

Shao, J. (1993). Linear Model Selection by Cross-Validation. *Journal of the American Statistical Association, 88*(422), 486--494.

Siddiqi, N. A. (1995). Site Suitability for raising *Nypa fruticans* plantations in the Sundarbans mangroves. *Journal of Tropical Forest Science, 7*(3), 405-411.

Skidmore, A. K., Varekamp, C., Wilson, L., Knowles, E., & Delaney, J. (1997). Remote sensing of soils in a eucalypt forest

environment. *International Journal of Remote Sensing, 18*(1), 39-56.

Spalding, M., Kainuma, M., & Collins, L. (2010). *World atlas of mangroves*. London: Earthscan.

Townsend, A. R., Cleveland, C. C., Asner, G. P., & Bustamante, M. M. C. (2007). Controls over foliar N:P ratios in tropical rain forests. *Ecology, 88*(1), 107-118.

Verhoef, W., & Bach, H. (2007). Coupled soil-leaf-canopy and atmosphere radiative transfier modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sensing of Environment, 109*(2), 166-182.

Vovides, A., López-Portillo, J., & Bashan, Y. (2011). N-fixation along a gradient of long-term disturbance in tropical mangroves bordering the gulf of Mexico. *Biology and Fertility of Soils, 47*(5), 567-576.

Wakeling, I. N., & Morris, J. J. (1993). A test of significance for partial least squares regression. *Journal of Chemometrics, 7*(4), 291-304.

Wandera, L. N. N. (2011). *Mapping chlorophyll concentration in a mangrove forest by model inversion approach applied to hyperspectral imagery.* (MSc Thesis) University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.

Wang, W.-Q., Wang, M., & Lin, P. (2003). Seasonal changes in element contents in mangrove element retranslocation during leaf senescene. *Plant and Soil, 252*(2), 187-193.

Wilkie, M. L., & Fortuna, S. (2003) Forest Resources Assessment Working Paper 63. *Status and trends in mangrove area extent worldwide*. Rome: Forestry Department, Food and Agriculture Organization of the United Nations.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109-130.

Zhang, Y., Chen, J. M., Miller, J. R., & Noland, T. L. (2008). Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. *Remote Sensing of Environment, 112*, 3234-3247.

Zwieten, P. A. M. v., Sidik, S. A., Noryadi, I., & Suyatna, I. (2006). Aquatic Food Production in the Coastal Zone: Data-based Perceptions on the Trade-off between Mariculture and Fisheries Production of the Mahakam Delta and Estuary, East Kalimantan, Indonesia. In T. P. Tuong, J. W. Gowing, B. Hardy & C. T. Hoanh (Eds.), *Environment and Livelihoods in Tropical Coastal Zones: Managing agriculture-fishery-aquaculture -*

conflicts (pp. 219-236). Wallingford, UK: CABI Publishing, in association with IRRI.

Appendix A: R^2 histograms

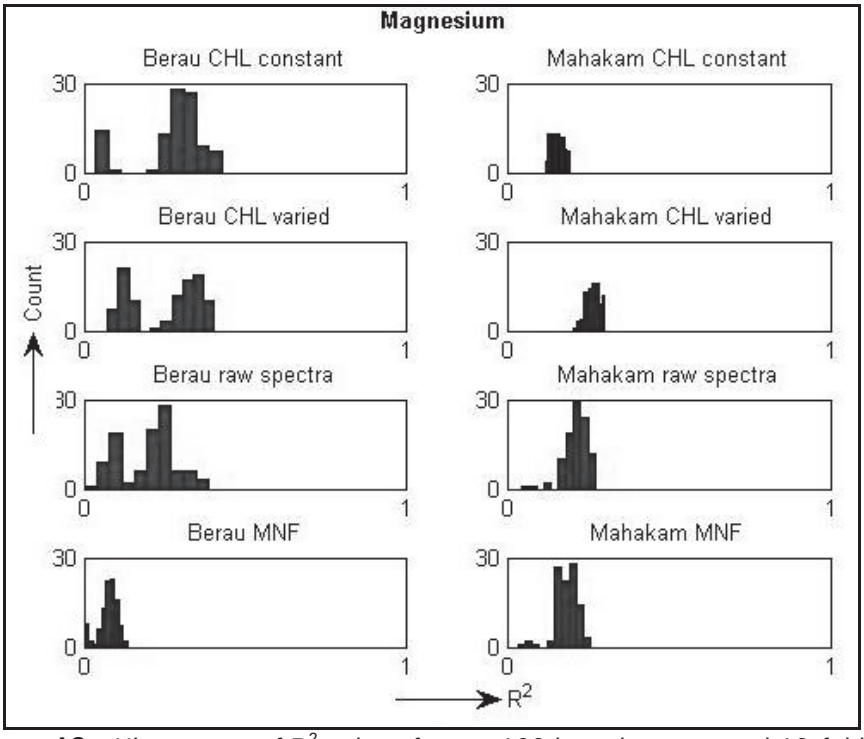

Figure 19 - Histograms of R^2 values from a 100 iteration repeated 10-fold cross validation of Mg content.

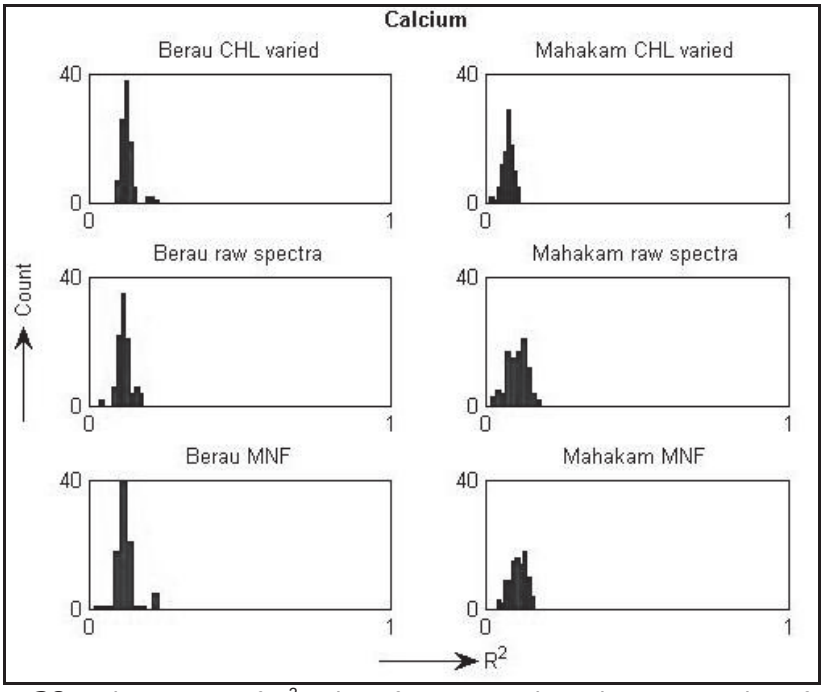

Figure 20 - Histograms of R^2 values from a 100 iteration repeated 10-fold cross validation of Ca content.

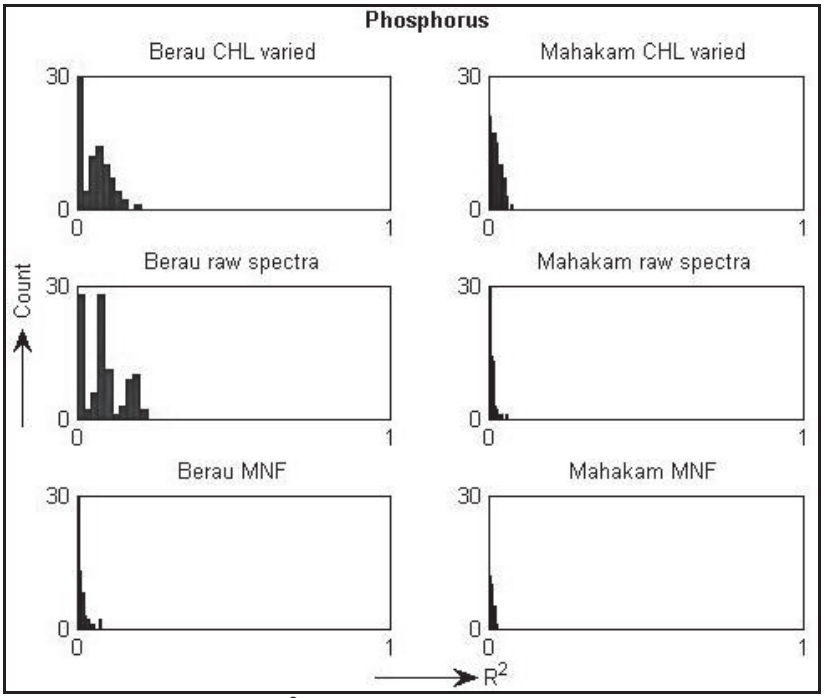

Figure 21 - Histograms of R^2 values from a 100 iteration repeated 10-fold cross validation of P content.

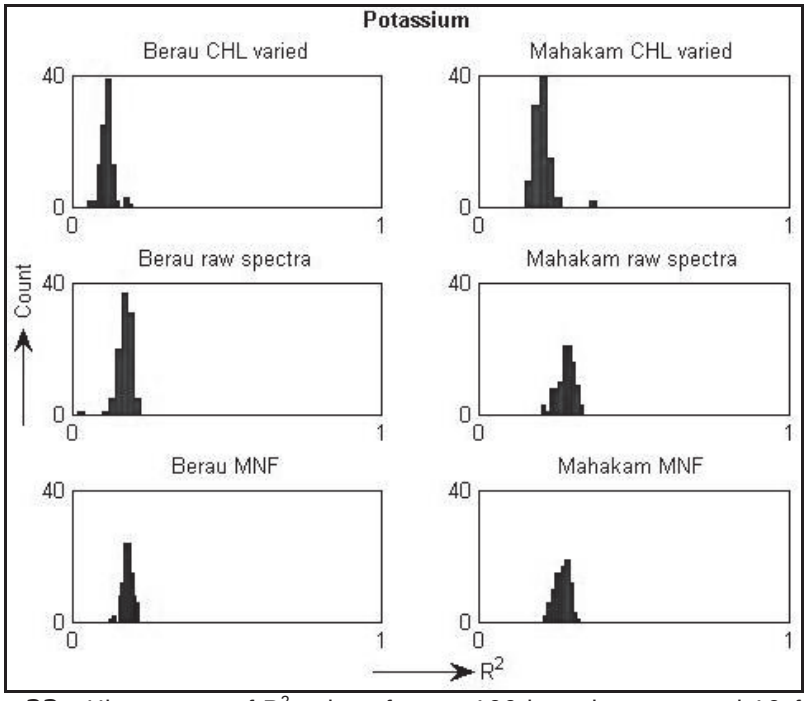

Figure 22 - Histograms of R^2 values from a 100 iteration repeated 10-fold cross validation of K content.

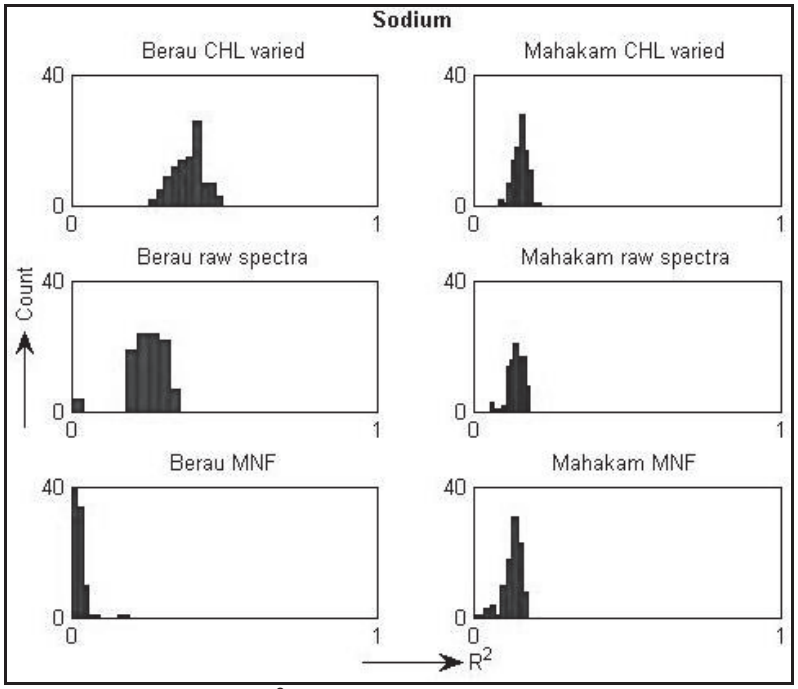

Figure 23 - Histograms of R^2 values from a 100 iteration repeated 10-fold cross validation of Na content.

Appendix B: Maps of ground sample locations

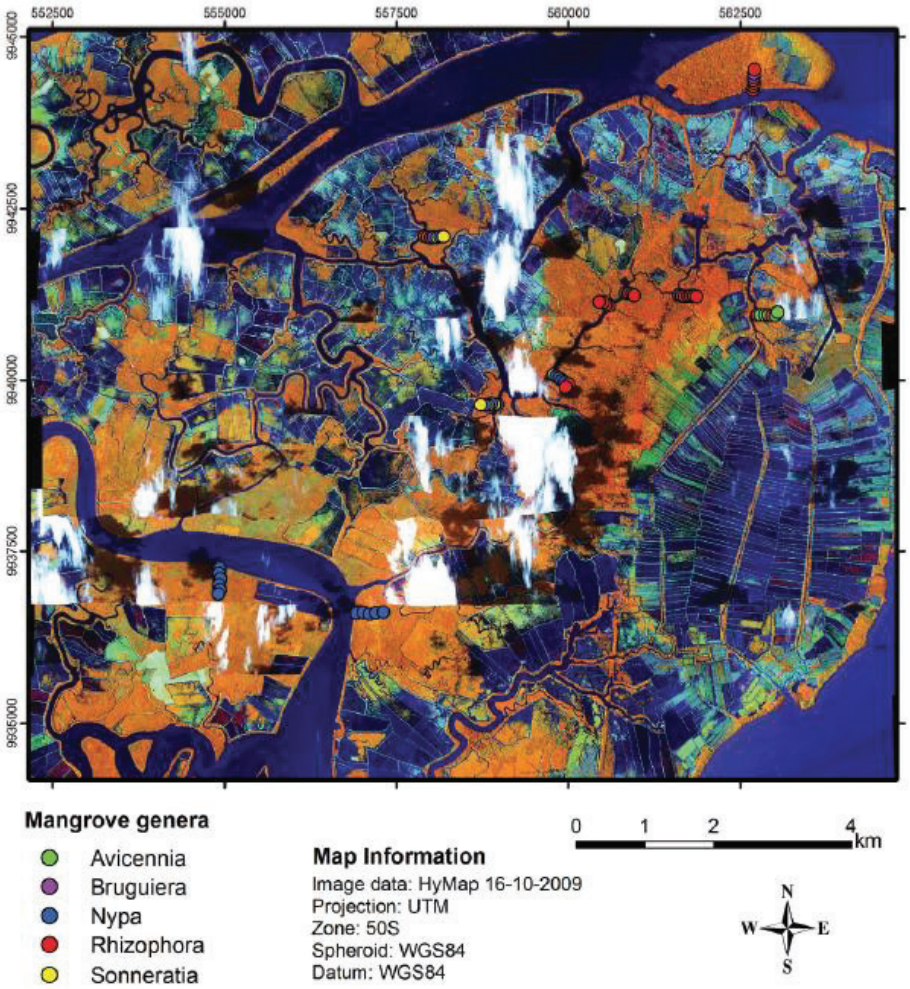

Figure 24 - False color map of the Mahakam study area. Sample locations are designated with circles. Color of circles indicates species. Map was generated by Christoffer Axelsson (Axelsson, 2011).

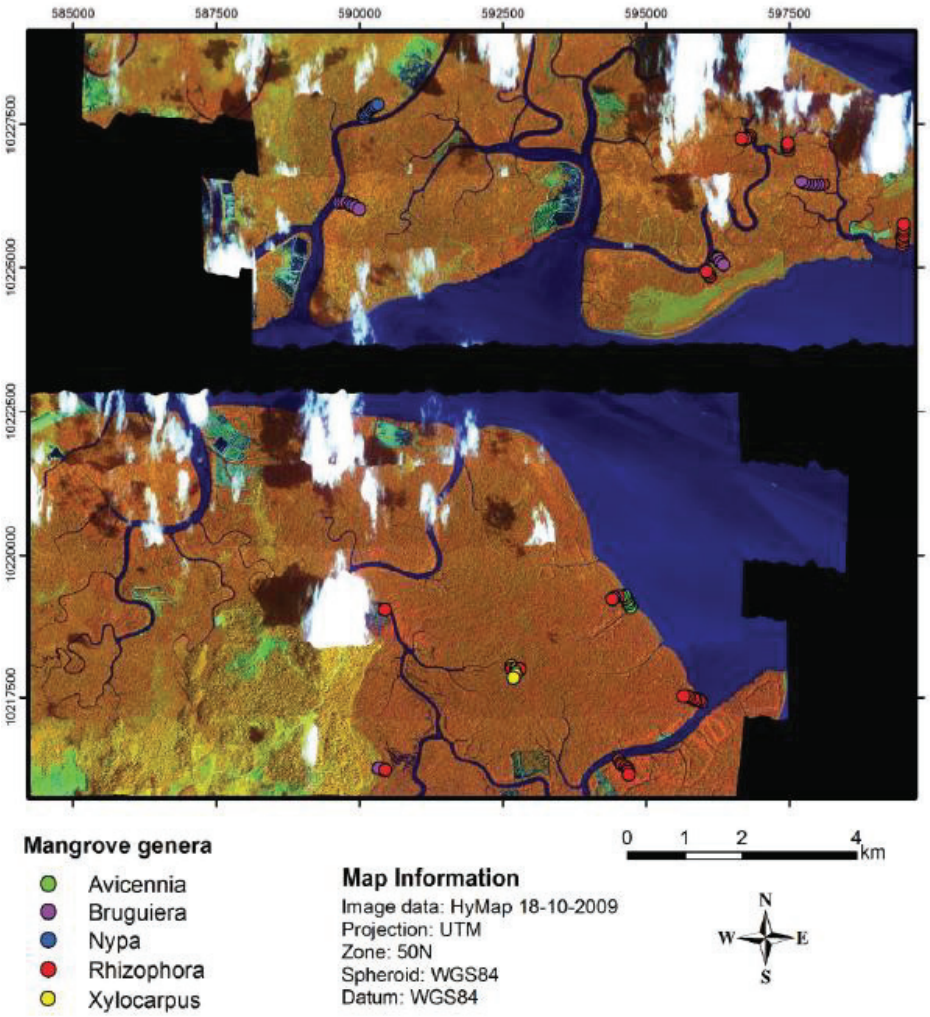

Figure 25 - False color map of the Berau study area. Sample locations are designated with circles. Color of circles indicates species. Map was generated by Christoffer Axelsson (Axelsson, 2011).