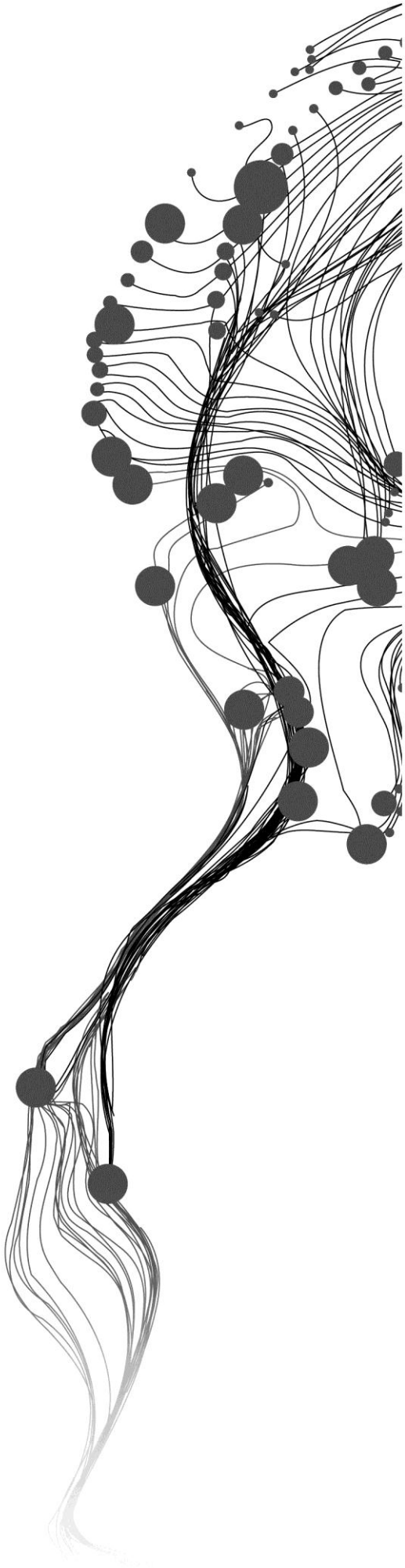
COMPARING HYPERSPECTRAL BASED PREDICTIVE MODELS FOR MAPPING SPATIAL VARIATION OF FOLIAR BIOCHEMISTRY IN NORWAY SPRUCE FOREST

DINESH BABU I.V.
February, 2011

SUPERVISORS:
Dr. M. Schlerf
Dr. A. Vrieling



COMPARING HYPERSPECTRAL BASED PREDICTIVE MODELS FOR MAPPING SPATIAL VARIATION OF FOLIAR BIOCHEMISTRY IN NORWAY SPRUCE FOREST

DINESH BABU I. V.

Enschede, The Netherlands, February, 2011

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Dr. M. Schlerf
Dr. A. Vrieling

THESIS ASSESSMENT BOARD:
Prof. Dr. W. Verhoef (Chair)
Prof. Dr. T. Udelhoven (External Examiner),
Remote Sensing & Geoinformatics Department,
Faculty of Geography and Geosciences, Trier University, Germany

ABSTRACT

Hyperspectral remote sensing of forest foliar biochemicals allows monitoring forest health status. Understanding the spatial variation of foliar biochemicals on the larger areas of forest ecosystem is still very limited. The objective of this research was to compare the empirical hyperspectral approaches stepwise multiple linear regression (SMLR), partial least square regression (PLSR) and boosted regression tree (BRT) for estimating foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) pigments at different spatial scales (laboratory and airborne HyMap spectra); and to understand the spatial variation of foliar biochemicals in Norway spruce forest. This research was carried in Gerolstein test site, West Germany situated in Eifel Mountains. At study site, 13 stands plots were identified to collect leaf samples (n=78). Collected leaf samples were used to represent foliar nutrients information at the leaf scale and in canopies. The SMLR, PLSR and BRT regression models were established between spectral bands from laboratory spectra and airborne HyMap spectra and field measured concentrations of chlorophyll a+b and carotenoids. The most accurate model was identified based on cross validation results. The reliability of predictive model parameters developed at the canopy scale was assessed by bootstrapping. The derived maps of foliar biochemicals was analysed for possible relation with the units of soil substrate using Analysis of Variance (ANOVA). The PLSR model predicted the chlorophyll a+b and carotenoids concentration most accurate than SMLR and BRT at the leaf scale. The sample size was found to be a limitation at the canopy scale to compare models. The predictions of carotenoids at the canopy scale were poor. Nevertheless, an SMLR model was developed to derive a map of leaf chlorophyll a+b concentrations (LCC) in Norway spruce forest. The model parameters were reliable from bootstrapping analysis. The derived map of LCC in Norway spruce reflected distinct spatial patterns for example, low chlorophyll concentrations were found under the low soil nutrient fertility. The differences between the mean of LCC in the soil substrate units (ANOVA result-P value = 0.004) was significant at 95% confidence level. The mean of LCC under the low soil fertility (P value of .000, 0.016, and 0.029) was significantly lower than the other soil substrate units from post-hoc test. The hyperspectral remote sensing of foliar biochemicals explains the underlying soil nutrient availability in forest ecosystem.

Keywords: Chlorophyll, carotenoids, hyperspectral remote sensing or imaging spectroscopy, stepwise multiple linear regression, partial least squares regression, boosted regression tree.

ACKNOWLEDGEMENTS

I would like to thank ERASMUS MUNDUS scholarship programme and Faculty of Geo-information Science and Earth observation (ITC), University of Twente for allowing me to pursue Master of Science programme by providing academic and financial supports.

I express my sincere thanks and gratitude to my first supervisor Dr. Martin Schlerf for his continuous support, critical comments and suggestions from research proposal to the final completion of thesis. He was always very kind to answer my questions. I learned many things from him. I would like to thank specially to my second supervisor, Dr. Anton Vrieling for his eye-opening critical comments and suggestions in analysis and thesis. I really enjoyed the field visit with both of them. I would greatly acknowledge Dr. Anton Vrieling for making nice field trip to study site.

I am greatly obliged to My Course Director Dr. Michael Weir for providing continuous support in studies and critical comments during the presentation. He was so kind to hear my words as a class representative.

My special thanks to Dr. Tiejun Wang and Dr. Thomas Groen for discussing regarding statistical part in my research.

I am greatly indebted to Mr. Shyam Kumar Shah for discussing in studies from the start of module to the final thesis. I would greatly acknowledge Mr. Satya Kumar Negi for giving me a moral support, guidance and sharing about philosophy of life.

Thank you so much to all of my friends especially Srijana Baral Jamarkattel, Rachna shah, Nandin-Erdene Tsendbazar, Aida Taghavi Bayat, Theresa Adjeley Adjaye, Damayanti Sarodja, Saurav Kumar Shrestha, and Loise Nancy Nangira Wandera. I thank all my NRM classmates.

TABLE OF CONTENTS

Abstract i
Acknowledgements ii
List of figures v
List of tables vi
1. INTRODUCTION 1
1.1. General objective 3
1.2. Specific objectives 3
1.3. Research questions 3
1.4. Research hypotheses 3
2. MATERIALS AND METHODS 5
2.1. Study Area 5
2.2. Data 5
2.2.1. Data collection and biochemical assay 5
2.2.2. Hyperspectral Image data 6
2.3. Software used 6
2.4. Research Methods 6
2.5. Pre-processing of spectra 7
2.5.1. Laboratory spectra 7
2.5.2. Airborne HyMap spectra (Continuum removal) 8
2.5.3. Extraction of canopy spectra 8
2.6. Model generation 9
2.6.1. Stepwise Multiple Linear Regression (SMLR) 9
2.6.2. Partial Least Squares Regression (PLSR) 9
2.6.3. Boosted Regression Tree (BRT) 10
2.7. Model validation and comparisons 11
2.7.1. Comparing the predictive ability of models based on the selected wavebands 12
2.7.2. Assessing the reliability of predictive model parameters at canopy scale 12
2.8. Mapping the foliar biochemical maps 13
2.9. Investigation of spatial variation in foliar biochemical maps 13
3. RESULTS AND DISCUSSION 17
3.1. Descriptive statistics of biochemical concentration 17
3.2. Extracted spectra 17
3.3. Stepwise Multiple Linear Regression models 18
3.3.1. Prediction results of SMLR from laboratory spectra 18
3.3.2. Prediction results of SMLR from HyMap image spectra 20
3.4. Partial Least Squares Regression models 21

3.4.1. Prediction results of PLSR from laboratory spectra 21
3.4.2. Prediction results of PLSR from HyMap image spectra 22
3.5. Boosted Regression Tree models 24
3.5.1. Prediction results of BRT from laboratory spectra 24
3.6. Models comparisons 26
3.6.1. Comparing the predictive ability of models at leaf scale 26
3.6.2. Comparing the performance of models at canopy scale 29
3.7. Assessing the reliability of predictive model parameters at canopy scale 30
3.8. Interpretation of the predictive model 32
3.9. Map of foliar chlorophyll a+b concentration 32
3.10. Investigation of spatial variation in foliar biochemicals 33
3.10.1. Sample size for ANOVA 34
3.10.2. ANOVA results 35
3.11. Exploring the relationships of light availability with the leaf chlorophyll concentrations 37
3.12. Summary of research answers 39
3.13. Limitations of this study 40
4. CONCLUSION 41
List of references 43

LIST OF FIGURES

Figure 2-1 Gerolstein test site, Germany; Source: Schlerf, 2006. 5
Figure 2-2 The general workflow of the research 7
Figure 2-3 Continuum removal for chlorophyll absorption of vegetation from (Clark, 1999) 8
Figure 2-4 The spectral characteristics of vegetation; Source : Jensen, (2007) 12
Figure 3-1 Continuum removed extracted spectra from HyMap image of Norway spruce plots; the name of the 13 sample plots in legend 18
Figure 3-2 Measured against estimated leaf chlorophyll a+b (*Cab*) concentration and measured against estimated carotenoids (*Cars*) using laboratory spectra from training of SMLR model using 472-761nm as input data 19
Figure 3-3 Measured against estimated leaf chlorophyll a+b (*Cab)* concentration and measured against estimated carotenoids (*Cars*) using HyMap image spectra from training of SMLR model 20
Figure 3-4 Measured against predicted leaf chlorophyll a+b (*Cab*) and measured against predicted carotenoids (*Cars*) using laboratory spectra from cross validation of PLSR model using 472-761nm as input data 22
Figure 3-5 Predicted against measured leaf chlorophyll a+b (*Cab*) and predicted against measured carotenoids (*Cars*) using HyMap image spectra from cross validation of PLSR model 23
Figure 3-6 Fitted functions of wavebands by BRT model for the prediction of *Cab* (n=78) in R Package; the value in percentage indicates wavebands contribution to predict *Cab* 25
Figure 3-7 Fitted functions of wavebands by BRT model for the prediction of *Cab* (n=78) in R Package; the value in percentage indicates wavebands contribution to predict *Cab.* 26
Figure 3-8 Comparison plots of R^2CV, RMSECV and nRMSECV for the prediction of chlorophyll a+b concentration from SMLR, PLSR and BRT using laboratory spectra 27
Figure 3-9 Comparison plots of R^2CV, RMSECV and nRMSECV for the prediction of carotenoids concentration from SMLR, PLSR and BRT using laboratory spectra 28
Figure 3-10 Histogram and normal quantile plots for the bootstrap replications (2000 bootstrap samples) of the intercept term; the broken vertical line in histogram shows the location of the regression coefficient for the model fit to the original sample. 31
Figure 3-11 Histograms and normal quantile plots for the bootstrap (2000 samples) (a) waveband 692 nm and (b) waveband 738 nm coefficients; the broken vertical line in each histogram shows the location of the regression coefficient for the model fit to the original sample. 31
Figure 3-12 The spectral characteristics of healthy and stressed vegetation; Source : NOAA, (2011) 32
Figure 3-13 Map of leaf chlorophyll concentration a+b for Norway spruce at Gerolstein study site, West Germany 33
Figure 3-14 Spatial location of Norway spruce stands over underground geology at Gerolstein test site, West Germany 34
Figure 3-15 The mean plots of chlorophyll a+b concentration in different geology class 36
Figure 3-16 Map of leaf chlorophyll concentrations over incoming solar radiation 38

LIST OF TABLES

Table 2-1 Soil substrates occurring at the study site from Schlerf, (2006) 13
Table 2-2 Choice of sample size for population (N) Source: Rea & Parker, (2005) 14
Table 3-1 Summary of the biochemical in situ measurements 17
Table 3-2 SMLR model from laboratory spectra (n=78) resampled to HyMap wavebands to predict chlorophyll a+b and carotenoids at leaf scale 18
Table 3-3 The interpretation of wavebands selected by SMLR model from laboratory spectra resampled to HyMap wavebands 19
Table 3-4 SMLR models from HyMap spectra (n=13) to predict chlorophyll a+b and carotenoids at canopy scale 20
Table 3-5 PLSR model from laboratory spectra (n=78) resampled to HyMap wavebands for the prediction of chlorophyll a+b and carotenoids at leaf scale 21
Table 3-6 The wavebands selected by the PLSR model from laboratory spectra resampled to HyMap waveband for the estimation of foliar biochemical 21
Table 3-7 PLSR model from HyMap (n=13) for the prediction of chlorophyll a+b and carotenoids at canopy scale 22
Table 3-8 BRT model from laboratory spectra resampled to HyMap wavebands (n =78) for the prediction of *Cab* and *Cars* at leaf scale 24
Table 3-9 Predictive models of chlorophyll a+b concentration from laboratory spectra resampled to HyMap wavebands (n=78) using input data of 472-761nm 27
Table 3-10 Predictive models of carotenoids from laboratory spectra resampled to HyMap wavebands (n=78) using input data of 472-761nm 27
Table 3-11 Predictive models of chlorophyll a+b at stand level from HyMap wavebands (n=13) 29
Table 3-12 Predictive models of carotenoids at stand level from HyMap wavebands (n=13) 29
Table 3-13 The Summary statistics of regression coefficients 30
Table 3-14 The Summary statistics of the regression coefficients for bootstrap sample 30
Table 3-15 Summary of the mean differences and ANOVA results of chlorophyll concentration a+b in different geology units 35
Table 3-16 Post- Hoc tests of multiple comparisons of leaf chlorophyll a+b concentration in different geology units 35
Table 3-17 Pair wise T-test of mean differences between chlorophyll a+b concentration in different geology units 36
Table 3-18 Correlation coefficient (r) of leaf chlorophyll concentrations with solar radiation, elevation and slope 39
Table 3-19 The significant wavebands to predict chlorophyll a+b and carotenoids at leaf scale. 40

1. INTRODUCTION

Foliar biochemistry is an indicator of various ecosystem processes, e.g. photosynthesis, transpiration, respiration, productivity and nutritional status (Kokaly et al., 2009; Martin et al., 2008). Among the many foliar biochemicals, chlorophyll a+b (*Cab*) pigments are necessary for photosynthesis and carotenoids (*Cars*) are essential for plant survival and nutritional functions (Davies, 2004). Carotenoids (*Cars*) are composed of carotenes and xanthophylls (Bartley & Scolnik, 1995). Chlorophyll a+b is an indicator of stress and phenological stage of vegetation (Blackburn, 1998b). Carotenoids concentration provides much useful insights on vegetation physiological status (Demmig-Adams et al., 1996). In stress and senescence condition of vegetation, the concentrations of chlorophyll decline and the quantity of carotenoids increase and produce the yellow and reds in foliage (Andrew et al., 2002). They are important inputs for forest ecosystem modelling (Lucas & Curran, 1999) and allow monitoring forest health status (Sampson et al., 2003). Conventional assessment of forest health is through visual inspection of trees which is impractical for larger areas. Such ground observations cannot provide detailed data for forest health mapping over larger areas. Remote sensing is a potential tool to produce these quantitative maps on regional scales (Johnson et al., 1994). Quantitative spatial maps of foliar biochemicals can provide an important input for forest management.

Spatial variation in foliar biochemical concentrations is caused by tree species composition, stand age and disturbance history (Ollinger et al., 2002), soil moisture availability (Niemann et al., 2002), nutrient availability (Solberg et al., 1998), elevation (Richardson & Berlyn, 2002), incident solar radiation or light availability (Baltzer & Thomas, 2005; Lee et al., 2000), climate (Magill et al., 2004; Omrad et al., 1999), management practices (Burger, 2009; Lopez-Serrano et al., 2005), insects defoliation (Lawerence & Labus, 2003) and pathogenic fungi (Craine et al., 2009). One of the important factors explaining the spatial variation in foliar biochemical concentration is soil nutrient availability (Miller & Watmough, 2009). Soil is the basic foundation of forest ecosystem (Boyle & Powers, 2000) and it provides nutrients to trees that reflects in the foliar biochemicals (Bauer et al., 1996). Soil properties like pH, moisture, percentage base saturation and cation exchange capacity influences the availability of nutrients to trees (Davis et al., 2007). Trees on productive soils are likely to produce high concentrations of foliar nitrogen (Sariyildiz & Anderson, 2005). In forestry research, previous studies made an effort to understand the relationship of foliar biochemicals with the soil chemistry in small sites (Guang-sheng et al., 2004; Hobbie & Gough, 2002; Zas & Serrada, 2003). The understanding of the spatial patterns of foliar biochemicals in relation with the underlying soil nutrients on the larger areas of forest landscape is quite limited (Ollinger et al., 2002).

The accurate quantification of foliar biochemical content using remote sensing is still a challenge (Blackburn, 2007). Multispectral remote sensing systems such as Landsat TM have been used for the estimation of foliar biochemicals using spectral vegetation indices (Duchemin, 1999; Richardson et al., 1983). These sensors collect information in 3 to 4 or more in broad spectral bands which can spectrally distinguish vegetation or land cover types, but cannot provide the adequate information to quantify foliar biochemicals (Thenkabail et al., 2004). In vegetation spectra, chlorophyll and carotenoids absorption takes place in specific narrow bands centred at 640nm, 670nm (chlorophyll) and 470nm (carotenoids), which is used to quantify relevant foliar biochemical concentrations (Van der Meer & De Jong, 2001). The major limitation of multispectral data is that the specific absorption feature to quantify relevant foliar biochemical content gets lost in broad spectral bands (Blackburn, 1998b).

Several researchers have shown that hyperspectral remote sensing or imaging spectroscopy data, which are characterized by several continuous (15 to several hundred) narrow spectral bands, can provide crucial information to quantify foliar biochemicals (Asner & Martin, 2009; Blackburn, 2007; Darvishzadeh et al., 2008; Mutanga et al., 2009; Townsend et al., 2003). Currently, statistical methods and physical based models are the two approaches used in hyperspectral remote sensing to estimate the foliar biochemicals concentrations (Im & Jensen, 2008; Majeke et al., 2008). Both methods have advantages and disadvantages. Physical models are known for their accurate estimation of biochemical compositions because they are based on physical principle (Schlerf & Atzberger, 2006). However, most physical models are complicated to implement the model inversion (Atzberger, 2004). They are successful at the leaf level, rather than canopy level, due to the complexity of vegetation canopies (Schlerf & Atzberger, 2006). Physical models have not been used widely as compared to statistical methods (Majeke et al., 2008), because they require a large number of input variables for model parameterization, which are often hard to estimate (Fang et al., 2003). On the other hand, statistical methods are site specific and not robust. Despite these disadvantages, several researchers attempted to use statistical methods (Asner & Martin, 2009; Demetriades-Shah et al., 1990; Gastelluetchegorry et al., 1995; Ge et al., 2008; LaCapra et al., 1996; Wessman, 1992; Zhang et al., 2008). However, the statistical methods are superior in some aspects, such as easy to establish and results are quite acceptable in the same site (Run-he et al., 2003). The focus of this research is on statistical methods.

Statistical methods are divided into univariate and multivariate methods. These methods estimate biochemical concentration by relating field measured biochemical concentrations with the reflectance in wavebands. Univariate methods relate only one independent variable e.g. spectral vegetation indices (reflectance in red and near-infrared bands), whereas multivariate methods relate many independent variables, for example many spectral bands to estimate foliar biochemical concentrations. Studies demonstrated that multivariate statistical methods provide better results in estimating foliar biochemicals rather than univariate regression (Darvishzadeh et al., 2008; Hansen & Schjoerring, 2003). Multivariate statistical methods develop calibration equation relating the field measurements of foliar biochemicals with the spectra using, for example stepwise multiple linear regression (SMLR) or partial least square regression (PLSR). Most of previous studies have used SMLR to estimate foliar biochemicals (Curran et al., 2001; Jacquemoud et al., 1995; Johnson et al., 1994; Kokaly & Clark, 1999; LaCapra et al., 1996). But, SMLR is open to criticism because this method suffers from multicollinearity, overfitting and the selected wavebands fail to correspond with known absorption features (Curran, 1989; Grossman et al., 1996). In contrast, the partial least squares regression (PLSR) method combines the most useful information from hundreds of bands into the factors, similar to the principal components analysis, that avoids the potential overfitting problem (Huang et al., 2004; Wold et al., 2001). Some studies also used partial least squares regression (Cho et al., 2007; Goodenough et al., 2009; Martin et al., 2008).

The limitation of these regression methods is that they assume a linear relationship between the field measurements and the spectra (Mutanga & Skidmore, 2004). However, there are techniques such as neural networks (Keiner & Yan, 1998), support vector machines (Durbha et al., 2007) and boosted regression tree (BRT) (Leathwick et al., 2006) to model nonlinear behaviour. Study in ecology have found that BRT is more interpretable than neural networks and support vector machines (Death, 2007). Studies have used BRT in remote sensing domain for many applications (Chan & Paelinckx, 2008; Ismail & Mutanga, 2010; Lawerence et al., 2006; Lawrence, 2004) and to predict biophysical and biochemical concentrations of crops (Im et al., 2009). Boosted regression tree models do not need to assume a linear relationship between response and explanatory variable. The major advantage of BRT is that it can fit complex nonlinear relationships, and routinely handles interactions effects between predictors, that find best combination of predictors overcoming multicollinearity (Elith et al., 2008). Despite these advantages,

no studies were found that used BRT to estimate forest foliar biochemicals using airborne hyperspectral remote sensing data.

1.1. General objective

This research aimed to test the predictive ability of BRT in comparison with SMLR and PLSR for estimating foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) at different spatial scales from laboratory spectra (non-spatial) and airborne hyperspectral spectra; and to understand the relationship of foliar biochemicals with the underlying soil substrate type in Norway spruce forest.

1.2. Specific objectives

1. To investigate the potential of boosted regression tree (BRT) to estimate foliar chlorophyll a+b and carotenoids and compare its performance with stepwise multiple linear regression (SMLR) and partial least squares regression (PLSR) from laboratory spectra and airborne hyperspectral spectra.
2. To derive maps of foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*).
3. To investigate the spatial variation of chlorophyll a+b (*Cab*) and carotenoids (*Cars*) over different underlying soil substrate types.

1.3. Research questions

1. What are spectral bands significantly predicted the chlorophyll a+b (*Cab*) and carotenoids (*Cars*) in SMLR, PLSR and BRT regression models from laboratory spectra and airborne hyperspectral spectra?
2. Which regression method estimates foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) most accurately from laboratory spectra and airborne hyperspectral spectra?
3. Are foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) concentrations significantly different over the underlying soil substrate types?

1.4. Research hypotheses

1.

H_0: BRT does not estimate foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) more accurately (higher Akaike information criterion (AIC) value) than PLSR and SMLR from laboratory spectra and airborne hyperspectral spectra.

H_1: BRT estimates foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) more accurately (lower Akaike information criterion (AIC) value) than PLSR and SMLR from laboratory spectra and airborne hyperspectral spectra.

2.

H_0: Foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) do not vary significantly over the underlying soil substrate types (95% confidence level).

H_1: Foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) vary significantly over the underlying soil substrate types (95% confidence level).

Assumptions:

- The underlying geology classes represent different soil substrate types.
- The effect of factors other than soil nutrients on foliar biochemicals like climate, diseases or pest attack, elevation, and forest management strategy etc. are all remain constants.
- Base saturation is an indicator of soil nutrient availability.

2. MATERIALS AND METHODS

2.1. Study Area

This research was carried out in Gerolstein test site (50°15'N, 6°40'E) which is located in the central part of the Eifel mountains, Germany (Figure 2-1). This study area was selected for this research because there is a natural variation in soil nutrients and it is expected to result in large variation of foliar chemicals (Schlerf, 2006). Norway spruce is the dominating species in this study area which are important and commonly distributed species in central Europe. These species occur over different soil types in this site from very poor to rich in soil nutrient availability which makes ideal for this research to study foliar biochemicals variation (Schlerf, 2006). The availability of data about this site also made practicable for this research.

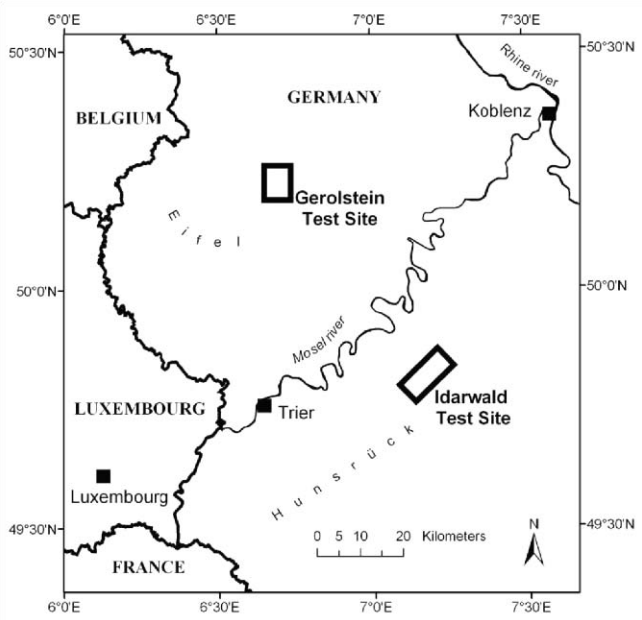


Figure 2-1 Gerolstein test site, Germany; Source: Schlerf, 2006.

The mean annual precipitation is about 800-900mm and the mean annual air temperature is 7°C. This site floor is covered by variety of bedrock types like mainly Devonian quartzite sandstone-schist formations, middle Devonian limestone, Triassic sandstones and quaternary basalts so that different soil substrate types have been developed (Schlerf et al., 2010). These soil substrate types vary with base saturation and moisture.

2.2. Data

This study used data from previous research [Schlerf,(2006)] that investigated hyperspectral remote sensing for estimating forest biochemical and biophysical attributes and results were published in Schlerf et.al.,(2010). The following explanation about methods of data collection and image data follow above references.

2.2.1. Data collection and biochemical assay

At study site, 13 stands of Norway spruce were identified by researcher and sampled in August 2002. Foliar samples were collected following guidelines of the German Environmental Specimen Bank from

top part of crown in three randomly selected trees in each stands (Schlerf et al., 2010). All the collected foliage samples were separated according to their position on the branch by age class (First Year and third year). In total, 78 samples were collected (13 stands x 3trees x 2 age classes). Foliage samples were bagged, labeled and kept in cold boxes in field. The samples were stored in refrigerators at +5^{o}C after transport to nearby laboratory. Sub-samples of 20 g were selected from these samples for reflectance measurements. Needle optical properties were measured in laboratory using ASD Field-Spec-2 spectrometer. The target was illuminated by a 1000 W halogen light source at a Zenith angle of 45^{o} and a distance of 60 cm. The head of instrument was at a nadir 10cm above targets. Relative spectral radiances were obtained for samples between 350 and 2500nm at 1nm spectral resolution. After obtaining reflectance measurements of these samples, the same needles were used for chemical analysis to measure the concentration of chlorophyll a+b and carotenoids.

2.2.2. Hyperspectral Image data

The airborne hyperspectral Mapper (HyMap) sensor with 126 spectral channels was flown over the study site on 14 July, 2003. The HyMap data of 2003 and the field measurements of 2002 collected during same season (summer season) and in same phenological stages were used assuming that the health status of forest didn't change significantly within one year (Schlerf et al., 2010). The HyMap data were recorded at 12.40 hr. at an altitude of average 3400 m above ground level. The spatial resolution of data is about 7m with a full scene covering about 4 km X 13 km. The data were collected during clear sky. The data were radiometrically pre-processed following across-track illumination correction for removal of view angle effects. The image data were corrected for atmospheric effects (for details refer (Schlerf, 2006) and geometrically also corrected using local coordinate system (Gauss Kruger, Zone 2, Ellipsoid : Bessel 1841, Datum: Potsdam).

2.3. Software used

- ENVI
- Arc GIS
- SPSS
- R
- ParLeS
- MATLAB

2.4. Research Methods

The general work flow of this research is outlined in Figure 2-2 below. The works carried in this research follows:

The workflow was divided into following steps:

1. Preprocessing of spectra
2. Model generation at different spatial scales (Lab spectra and Airborne HyMap spectra)
3. Model validation and comparisons
4. Mapping of leaf foliar biochemicals
5. Investigation of spatial variation in foliar biochemical maps

2.5. Pre-processing of spectra

2.5.1. Laboratory spectra

The laboratory reflectance spectra (350nm-2500nm, 350 wavebands) was resampled to HyMap wavebands (128 wavebands) to establish models which predict chlorophyll concentration a+b (*Cab*) and carotenoids (*Cars*).

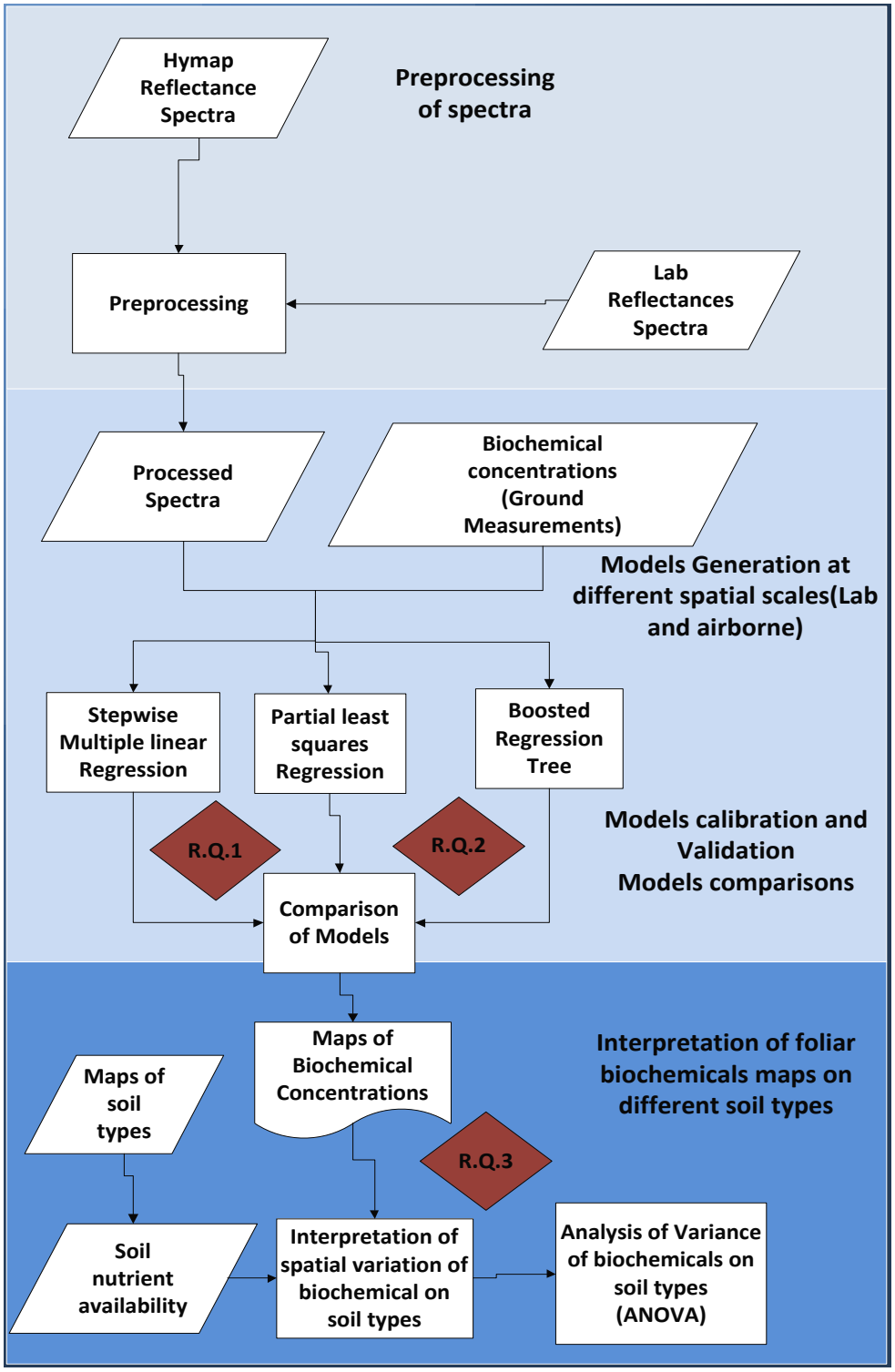


Figure 2-2 The general workflow of the research

2.5.2. Airborne HyMap spectra (Continuum removal)

The HyMap image data were processed using continuum removal (spectral transformation method) before extracting the spectra for analysis using ENVI image processing package. The continuum is defined as a "convex hull of straight line segments over the top of a spectrum to connect local spectral maxima" (Huang et al., 2004)(refer Figure 2-3). The continuum removed (CR) spectra at a certain wavelength was obtained by dividing the original reflectance values for each waveband in the absorption feature by the corresponding reflectance values of the continuum line (convex hull) (Huang et al., 2004; Mutanga & Skidmore, 2004) (refer Figure 2-3). In the continuum removed spectra, the first and last spectral values are equal to one, since they both lie on the continuum line (Mutanga & Skidmore, 2003). The output reflectance values of the CR spectra are scaled between zero and one, which enhance the absorption troughs in vegetation spectra (Schmidt & Skidmore, 2001). This allows for better estimation through the identification of specific absorption features to quantify relevant foliar biochemicals. Spectral transformation is the important process converts original spectra into transformed spectra which were widely used by many researchers for statistical modelling with hyperspectral data (Huang et al., 2004; Kokaly et al., 2009; Run-he et al., 2003). Previous studies have demonstrated that continuum removal performs superior to other spectral transformation methods in the estimation of foliar biochemicals (Kokaly & Clark, 1999; Mutanga & Skidmore, 2004).

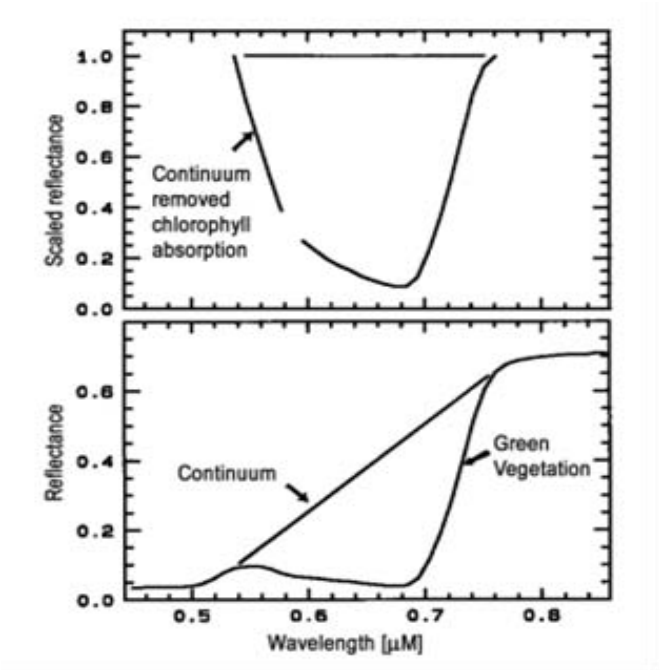


Figure 2-3 Continuum removal for chlorophyll absorption of vegetation from (Clark, 1999)

2.5.3. Extraction of canopy spectra

The field sample points were overlaid on CR HyMap image to extract spectra for analysis. Region growing approach (ROI) was applied using ENVI (Version 4.5) to define the representative pixels around the sample plot stands. At each sample plot in the image, four pixels were defined as a "region of interest" (ROI) in ENVI Package. Growing of ROIs was allowed in all directions with the standard deviation of 2 that resulted in the image pixels of 9-35 around sample plot stands. For further analysis, only the average spectra from the selected pixels were used. These extracted spectra were used for statistical modelling to estimate chlorophyll a+b and carotenoids at the canopy scale.

2.6. Model generation

The SMLR, PLSR and BRT regression models were established between spectral bands from laboratory spectra (non-spatial) and airborne HyMap spectra and concentrations of chlorophyll a+b and carotenoids at leaf and canopy scale respectively. During the process of model development, the models were analysed with both the full spectrum and the reduced spectrum of absorption features (471-761nm) as input data. Here after, models developed with the full spectrum and the reduced spectrum of absorption features (471-761nm) were called as the full spectrum and reduced spectrum models respectively.

2.6.1. Stepwise Multiple Linear Regression (SMLR)

Stepwise multiple linear regression (SMLR) used to identify significant wavebands to predict foliar *Cab* and *Cars* for regression model. It works on assumption that certain numbers of wavelengths have a significant contribution to predict response variable. SLMR runs on the predictor variables to identify statistically significant explanatory predictor wavelengths based on the highest F-value or lowest *p*-value for prediction. First it starts with no explanatory variables (wavelengths) in the regression equation and at each step it adds the statistically significant predictors (lowest *p*-value), until no further entry or removal is possible (Darvishzadeh et al., 2008). At the same time, the SMLR computes the removal statistic to remove the non-significant predictor wavelengths (highest *p*-value) for prediction. In this study, the *P* values for entry and removal were set at ≤ 0.05 and ≥ 0.1 respectively. Forward SMLR was used to predict *Cab* and *Cars* at the leaf and canopy scales using laboratory and HyMap spectra respectively. SMLR result in the predictive equation to predict the studied response variable.

To control model over-fitting, the following measures were implemented from Schlerf et.al., (2010), 1) Combination of predictor wavelengths were selected based on the minimum RMSE (Root mean square error); 2) The number of bands included in the model was controlled by *p* -values for entry and removal; 3) The *p* –values for entry and removal were decreased until no relation observed between spectral data and biochemical. The SMLR analysis was performed using the MATLAB tool. More description about mathematical concepts of SMLR can be referred to this paper Ganesh, (2010).

The main assumptions and criticism in this SMLR follows Curran, (1989): 1) the relationship between predictor wavelengths and chemical concentration are near linear ; 2) the relationship between spectra and chemical concentration is not confounded by other factors such as leaf angle and area distribution, solar incidence angle and sensor view angle, and plant factors; 3) the chemical concentrations have been accurately measured. The criticism of SMLR: 1) the model may over-fit when the number of samples is smaller than the number of predictor wavelengths. 2) It may or may not be possible to explain the selected wavebands by SMLR based on chemical bonds which may occur due to inter-correlation of chemicals. 3) It is also not possible to explain why a certain waveband, which is known to be related to certain biochemical, has not been selected by SMLR.

2.6.2. Partial Least Squares Regression (PLSR)

PLSR is a combination of multiple regression and principal component regression. This technique reduces the large number of spectral variables to a few non-correlated latent variable or factors (Cho et al., 2007). This latent variable is derived from covariance between the few non correlated reflectance spectral variables and the biochemical concentrations. The latent variables or factors represent the relevant information present in the reflectance spectra to predict the foliar biochemical concentrations (Darvishzadeh et al., 2008). It overcomes the problem of overfitting and multicollinearity which is found with SMLR (Hansen & Schjoerring, 2003). It aims to develop a simple linear model. It works on hypothesis that the relationships between response variable and predictor wavelengths are linear as SMLR. It does not result in the simple predictive equation for estimation as SMLR. One problem with PLSR is

that identifying important predictor wavelengths is not straightforward and must use correlograms of reflectance and chemical concentration to identify predictor bands (Huang et al., 2004). The PLSR analysis was performed using ParLeS package which was developed by Viscarra Rossel. The PLSR analysis was performed on the entire reflectance spectra and separately with the reduced spectrum of relevant absorption features (471-761nm) to predict foliar biochemicals. To prevent model over-fitting, the optimal number of factors for estimation was determined based on cross-validation method. A common way to select the number of factors that minimizes the RMSE and AIC (Akaike Information Criterion) (Nguyen & Lee, 2006). More description about PLSR algorithm can refer to these papers (Bastien et al., 2005; Geladi & Dabakk, 2009; Wold et al., 2001).

2.6.3. Boosted Regression Tree (BRT)

The BRT is a machine learning technique which aims to develop a single model by fitting combination of many models for explaining the target variable. It uses algorithms of classification and regression tree (predict response variable by recursive binary splits of predictors) and boosting (combines many simple models to provide improved predictive performance)(Elith et al., 2008). It differs fundamentally from conventional regression techniques. The performance of conventional regression models is of deterministic (i.e. fixed input with fixed output), but BRT is of stochastic (it includes random or probabilistic). This method was invented by Jerome H. Friedman and more description about BRT model can be referred to (Friedman, 2001, 2002). It is also called as stochastic gradient boosting or Tree Boost models or multiple additive regression trees or machine learning regression trees. It works on nonlinear relationships between target variable and predictor variable without assumption on linear relationships. It overcomes the limitation found in SMLR and PLSR models. There is no literature about the use of BRT model in statistical modelling with hyperspectral data for the prediction of forest foliar nutrients.

Below is a brief outline of BRT model as reported by researchers in ecology (Elith et al., 2008; Schonlau, 2005; Sherrod, 2010):

1) The model first finds the average value of target variable or response variable (first guess for predicting response variable) and identify regions of predictors which having most important responses.
2) On each step, the focus of the tree model is to fit on residuals (tree fits to difference between measured response variable and average response variables). Each tree predicts residuals.
3) The residuals (error values) from the first tree are then used for the second tree and growing of tree process is continued through a chain of successive trees using recursive binary splits.
4) The BRT modelling process is stage-wise (not stepwise) which means that existing trees are left unchanged as the model becomes enlarged.
5) The BRT model automatically handles interaction effects between predictors (Multicollinearity).
6) The final model is a linear combination of many trees (hundreds) and made as single final regression model or additive regression model. Each tree is a model. The final predicted value is estimated by adding the weighted contribution of each tree.
7) The variable selection in BRT is obtained by ignoring unimportant predictors when fitting trees. This measures relative influence or contribution of each predictor variable scaled to 100 and quantify the importance of predictors (highest numbers indicating stronger influence on target variable).
8) The unimportant predictors can be removed which have minimal effect on response called as recursive feature elimination. Such simplification of BRT model is used for small datasets.
9) To develop the BRT models, it requires for the specification of two main parameters. The learning rate which defines the contribution of each tree to the growing model and the size (number of splits) of individual trees called as tree complexity. This determines the number of

trees in the model. The BRT model was developed using 0.01 (learning rate) and 5 (tree complexity).

10) The performance of BRT model is stochastic (i.e., it includes random or probabilistic), which improves predictive performance, that decreases the variance of the final model by using random subset of data for building model (Friedman, 2002).

This method is implemented in gbm package version 1.5-7 (Ridgeway, 2006) of R software which was used to develop BRT models. To avoid model over-fitting, the following measures were implemented: 1) the optimal number of trees were identified using cross-validation method which has minimum error and trees were pruned accordingly. 2) Random selection of rows without replacement from full training data was used at each iteration 50% of data is drawn random using bag fraction (functions in R of gbm package) as 0.5. This reduces over-fitting (Friedman, 2002). 3) the unimportant predictors were dropped for the simplification of BRT model (Elith et al., 2008).

2.7. Model validation and comparisons

Two types of validation were used to assess the performance of the studied models: (i) Leave one out cross validation method (LOOCV) for SMLR and PLSR models and (ii) n.fold cross validation for BRT models. As the number of samples were limited, the cross-validation procedure was used which divide the data set into training and test data set. The cross validation method provides unbiased estimations of the prediction error (Cawley & Talbot, 2004; Efron & Gong, 1983). This means that the predicted samples were not same as the samples used for model building. The leave one out cross validation method uses one data for validation and the remaining data for calibration and iteratively runs equal to the sample size to develop models so that each data is used for validation. For example, data with sample size of 78, it develop 78 individual models, each time with data from 77 and the remaining one for validation. This method was used by similar type of studies to evaluate the SMLR and PLSR models (Darvishzadeh et al., 2008; Huang et al., 2004). The BRT model was evaluated using n. fold cross validation procedure. Where, n denotes the number of groups of data made for cross validation (here n = 10). This method divides the data into groups of data such that in each group, equal size was made for training and validation from 50% random subsets. For example, data with sample size of 78, it develop 10 individual models, each time with data from 50% random subsets (without replacement) for training and the remaining for validation. This is an efficient algorithm to assess the stability of BRT model (Blockeel & Struyf, 2002; Friedman, 2002).

The performance of the studied models were assessed using cross-validated statistics: R^2CV (coefficient of determination between measured and predicted values from model), root mean squared error (RMSECV) (the difference between measured and corresponding predicted values from model are each squared and then averaged over number of the sample, and the final is the square root of the average) and normalized mean square error (nRMSECV) (RMSECV divided by mean of measured variable (response variable) expressed as percentage). R^2CV is an indicator of precision of model and RMSECV is the measure of predictive accuracy of model (Tedeschi, 2006).

To identify the best model, among various models with the different number of predictor variables, Akaike information criterion (AIC) was used which includes both RMSECV and number of parameters. The AIC value penalize model for overfitting and identifies the model which produce good accuracy with fewer terms.

The formula to calculate AIC = N log(RMSE) + 2m from Li et al., (2002)

Where N is sample size, m is the number of parameters, RMSE is the Root mean square error of the model, N log(RMSE) represents the model accuracy and 2m denotes model parsimony. This value was computed for all models. The model which has the lower value of AIC was considered as the best parsimonious model (i.e., fewer terms)(Wang & Liu, 2006).

2.7.1. Comparing the predictive ability of models based on the selected wavebands

The wavebands selected by the models were analysed for consistency in relation to the known absorption features in vegetation spectra. The major chlorophyll absorption bands occur in the 0.43, 0.46, 0.64, and 0.66 μm (Figure 2-4)(Curran et al., 1997). The carotenoids absorption takes place in the 0.44 and 0.47μm (Jensen, 2007). The major absorption wavelengths are in the 0.65 -0.67 μm (red spectrum)(Curran et al., 1995). The visible and red-edge regions are the major spectral regions contribute for the absorption of pigments (Figure 2-4). The water absorption bands are 0.97, 1.19, 1.45, 1.94 and 2.70 μm (Figure 2-4).

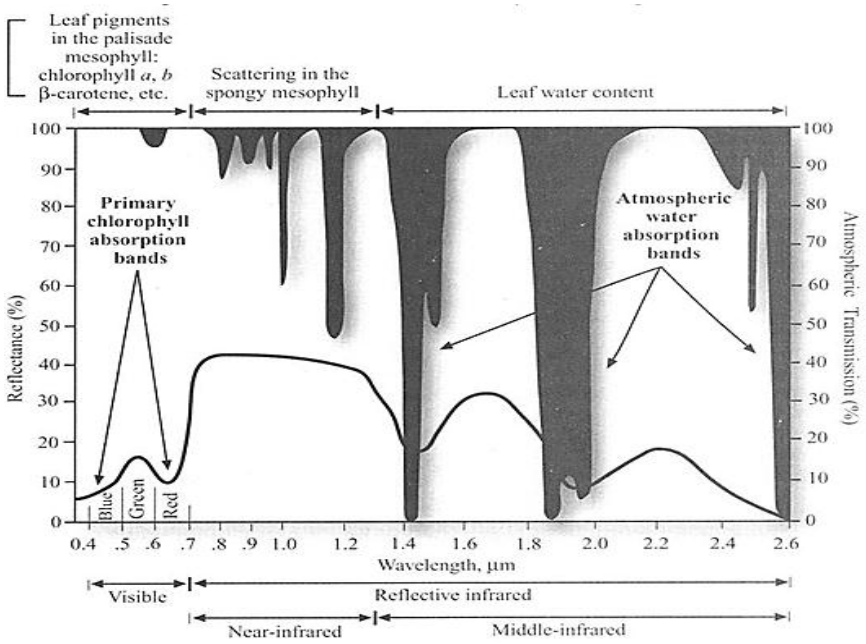


Figure 2-4 The spectral characteristics of vegetation; Source : Jensen, (2007)

2.7.2. Assessing the reliability of predictive model parameters at canopy scale

The sample size is always limitation for these type of studies to represent the foliar biochemical content in the canopies because of destructive sampling, labour-intensive and expensive laboratory techniques (Blackburn, 2007). Here, the sample size was only 13 at the canopy scale. The model coefficients from the small sample size may not be reliable (Fox, 1997). The reliability of the regression parameters can be assessed using bootstrapping (Moulton & Zeger, 1991). Bootstrapping is a nonparametric approach which uses to resample the sample data repeatedly for estimating the relevant characteristics of the population; the key analogy is "the population is to sample as the sample is to bootstrap samples"(Fox, 1997). The meaning of the term bootstrapping is "an illusion to the expression pulling oneself up by ones bootstraps" and considering repeated samples drawn as population (Efron, 1979). This technique allows understanding the variability of regression parameters and measures uncertainty in the estimated standard errors of the coefficients that are more reliable by constructing "pseudo-data" from the sample at hand (Freedman & Peters, 1984). The model parameters accuracy were assessed by 2000 bootstrap samples by resampling with replacement from the sample data in the same manner (Y variable with respective X variables called as fixed X resampling). Each bootstrap sample selects values with replacement from among values of the original sample result in the 2000 set of data sets. Then, these bootstrap samples were used for estimating the regression coefficients. The reliability of the regression coefficients from the original sample was assessed by bias. The bias means the systematic deviation from the true value (Tedeschi, 2006). The bias was computed from the differences between the original coefficients and average coefficients of the bootstrapped samples (true parameter value). The bias of the parameters should be zero for an accurate and unbiased estimator.

2.8. Mapping the foliar biochemical maps

The model which was most accurate developed from HyMap image spectra to predict foliar *Cab* and *Cars* concentration used to produce the maps of foliar biochemicals for spruce pixels alone. Norway spruce stands were digitized from HyMap based on visual interpretation.

2.9. Investigation of spatial variation in foliar biochemical maps

This research expected that the derived foliar nutrients map result in spatial variation to explain the soil nutrient availability. In this site, the spatial variability in soil nutrient availability was mainly due to the different underlying geology that developed the diverse soil substrates varying in soil base saturation and moisture (Schlerf, 2006). There are major four types of underground geology units present in this site namely Buntsandstein, lower Devonian, mid and upper Devonian and quaternary which have developed soil substrate with base saturation of very low, medium, very high and high respectively (Table 2-1). The map was derived from the 13 sample plots located in the eight soil substrate (Table 2-1). The concept of soil substrate is used in Germany for classifying soil types by the Federal Research Centre for Forestry and Forest Products (Schlerf, 2006). The term "soil substrate" describes "the combined grain-size distribution of the soil material, the type of bed rock and the stratification"(Schlerf, 2006).

Table 2-1 Soil substrates occurring at the study site from Schlerf, (2006)

Abbreviation	Sample plots (Number)	Soil substrate	Underlying geology	Base saturation
VLdu	GS 16, GS 17, GS 18	Loam formed from quartzite, sandstone and shale	Lower Devonian	Medium
wDLT		Wet loam cover above clay loam	Lower Devonian	Medium
KVLdm	GS 9	Loam formed from limestone and dolomite	Mid and upper Devonian	Very high
Ssm	GS 19, GS 21, GS 24, GS 25, GS 10b	Sands	Mid -Buntsandstein	Very low
DLsm		Loam cover above sandstone	Mid -Buntsandstein	Very low
BVL	GS 2, GS 3, GS 4, GS 8	Loam formed from basalt	Quaternary	high
Kol+		Colluvium	Quaternary	high
DLB		Loam cover above basalt	Quaternary	high

This forest is not intensively managed and the soil nutrients are stable (Personal communication with forest authorities during field visit). The derived foliar biochemical map was analysed in relation to the underground soil properties (from geology classes). The non-availability of the spatial information regarding the soil substrate was found to be a limitation. The geology classes or units were assumed representative of the soil substrate types. For example, a mid-buntsandstein geology unit represents the soil substrate (sands) which has the very low base saturation (Table 2-1). Similarly, a quaternary geology unit represents the soil substrate which has the high base saturation (Table 2-1). Base saturation was assumed as an indicator of soil nutrient availability. The term soil base saturation means the amount of exchangeable cations (Ca, Mg, K, and Na) other than aluminum and hydrogen in soil expressed as percentage. A higher fraction of clay content and organic matter of soil results in higher soil base saturation. Base saturation is an indicator of soil nutrient availability (i.e. high base saturation has high fertility) which reflects in the foliar nutrient status of vegetation (Sariyildiz & Anderson, 2005). The parent

materials or geology is strongly responsible for the development of soil (Clemens et al., 2010; Neff et al., 2006). The mineralogy of parent soils is the important factor determines the functioning of forest ecosystems (Sariyildiz & Anderson, 2005).

The analysis of spatial variation was performed using Analysis of Variance (ANOVA), post-hoc test and pair wise t-test. The ANOVA is a statistical test used to test hypothesis that whether the mean between groups are equal or not. The post-hoc test and pair wise t-test provides an additional information on which means differs significantly from each other. It is not possible to consider the whole derived map (pixels) to compare the mean of foliar biochemical concentration in each geology class based on t-statistic. The below is the reason follows Moore & McCabe, (2006):

1) The sample size strongly influences the P-value of ANOVA test. The difference between the mean of two groups, which is insignificant at a specified confidence level in a small sample, can become significant in a larger sample.
2) The t statistic to compare the mean of two groups is reasonably robust when the sample sizes are equal.

Because of these reasons, there is a lot of uncertainty to consider all pixels in the derived map for ANOVA analysis. So, the samples were taken for ANOVA analysis. This infers how the mean of foliar biochemical concentration for Norway spruce (target population) differs significantly over the underlying soil substrate. Similarly, McNeil et al., (2008) applied random sampling design on the derived map of canopy nitrogen from Hyperion (hyperspectral data) for analysing the relationships between the canopy nitrogen and the soil carbon: nitrogen ratio.

For the purpose of sampling, stratified random sampling design was applied on the derived map of foliar biochemical concentration. Each pixel was considered as sample unit. Each geology unit was considered as stratum. The derived map of foliar biochemical content was overlaid on the geology class shapefile in Arc GIS. Then, the number of pixels in each stratum (geology unit) was calculated in Arc GIS in order to determine the sample size based on the known population number. Table 2-2 shows the list of sample size for population (N) with allowable errors.

Table 2-2 Choice of sample size for population (N) Source: Rea & Parker, (2005)

	95% level of confidence			99% level of confidence		
N	±3%	±5%	±10%	±3%	±5%	±10%
500	250	218	81	250	250	124
1000	500	278	88	500	399	143
1500	624	306	91	750	460	150
2,000	696	323	92	959	498	154
3,000	788	341	94	1,142	544	158
5,000	880	357	95	1,347	586	161
10,000	965	370	96	1,556	622	164
20,000	1,014	377	96	1,687	642	165
50,000	1,045	382	96	1,777	655	166
100,000	1,058	383	96	1,809	659	166

The above table 2-2 was used as references to decide the number of samples for ANOVA analysis. The formula underlying in the above table is from Rea & Parker,(2005):

$$n = \frac{Z^2_{\alpha/2}\sigma^2}{E^2} \cdot \sqrt{\frac{N}{N-1}}$$

Where n= number of samples to determined based on Z value and allowable error. Then this was corrected for finite population N.

N= number of population size

σ is the variance of the population (0.5 is used for sample size determination)

$Z^2_{\alpha/2}$ is the Z value at α=0.05 (two-tailed) (e.g. 1.96 for 95 % confidence level)

E^2 is the maximum allowable error.

Random points (equal to the number of samples determined) were generated in Arc GIS package at 1m apart from each point in each stratum. The generated random points were used to extract samples from the foliar biochemical concentration map for ANOVA analysis. The ANOVA, post-hoc test and pairwise t-test analysis was performed using SPSS package.

3. RESULTS AND DISCUSSION

3.1. Descriptive statistics of biochemical concentration

Table 3-1 Summary of the biochemical in situ measurements

	Biochemical	n	Min	Max	Mean	Median	Std.Dev	CV (%)
Year1	*Cab*	39	1.71	4.82	2.75	2.67	0.66	24
	Cars	39	0.32	0.63	0.46	0.45	0.06	13
Year3	*Cab*	39	2.13	5.89	3.60	3.45	0.88	24
	Cars	39	0.45	0.86	0.64	0.65	0.08	12
Year1+Year3	*Cab*	78	1.71	5.89	3.17	3.02	0.88	27
	Cars	78	0.32	0.86	0.55	0.54	0.12	21
Average of two age classes and 3 trees per forest stand	*Cab*	13	2.71	3.62	3.21	3.23	0.30	9
	Cars	13	0.48	0.61	0.55	0.56	0.04	7

n = Sample size; Min = Minimum; Max = Maximum; Std.Dev = Standard Deviation; CV = Coefficient of variation; *Cab* = Chlorophyll a+b concentration [mg g^{-1} dry matter]; *Cars* = Carotenoids concentration [mg g^{-1} dry matter]; Total 78 samples = 13 stands x 3 trees x 2 age classes.

In order to link canopy reflectance to biochemical measurements, the biochemical data (First year and third year needles) on the leaf level were averaged by the mean of both age classes to stand level(Table 3-1). It was justified since concentration in both needle age classes was positively correlated. The correlation coefficient, r between *Cab* in Year1 and Year 3 was 0.70. Models were developed at the leaf scale from laboratory spectra, by pooling samples of both age classes to predict *Cab* and *Cars* age class independently.

3.2. Extracted spectra

The spectra was extracted from CR removed HyMap image of entire spectrum (435-2400nm) and also with CR removed selected absorption wavebands from 545–768 nm for analysis. So that, model was developed both with the full spectrum and the reduced spectrum of selected absorption (545-768nm) wavebands. This spectral transformation was not applied with the laboratory spectra. It was justified from previous study results that spectral transformations didn't have much pronounced positive effect in prediction with laboratory spectra for statistical modelling (Schlerf et al., 2010). Figure 3-1 shows extracted spectra from CR HyMap image which was used for regression analysis.

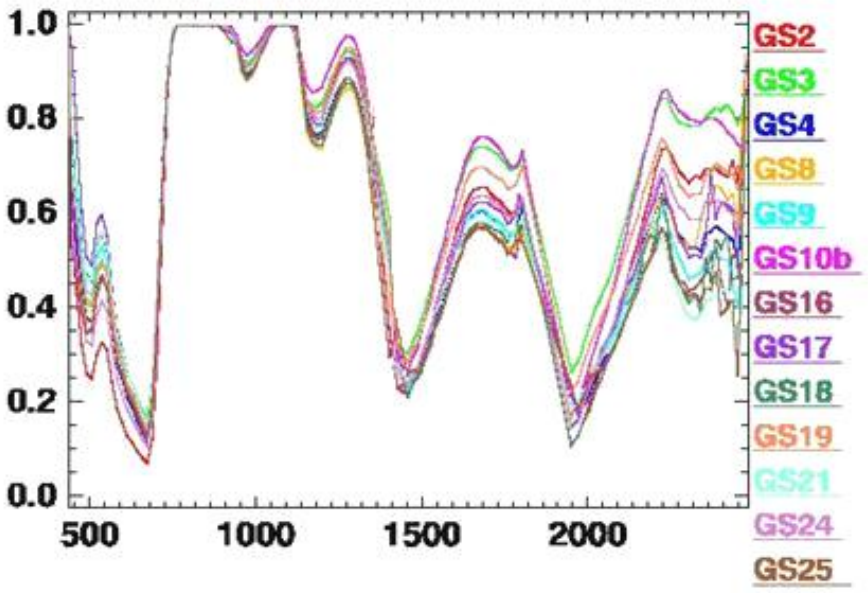


Figure 3-1 Continuum removed extracted spectra from HyMap image of Norway spruce plots; the name of the 13 sample plots in legend

X axis = wavelength (nm); y-axis = Continuum removed reflectance (0 to 1)

3.3. Stepwise Multiple Linear Regression models

3.3.1. Prediction results of SMLR from laboratory spectra

The SMLR models were developed using the entire spectrum and reduced spectrum of 471-761 nm to predict chlorophyll a+b and carotenoids at the leaf scale. Refer to the Table 3-2 for the results.

Table 3-2 SMLR model from laboratory spectra (n=78) resampled to HyMap wavebands to predict chlorophyll a+b and carotenoids at leaf scale

Biochemical	Training		Cross-validation					
	R^2	RMSE	R^2CV	RMSE -CV	nRMSE -CV	Input data (Total No. of predictors in brackets)	AIC	No. of wavebands selected
Cab	0.75	0.31	0.72	0.33	10.41%	472-761nm(20)	-31.56	3
	0.80	0.28	0.61	0.40	12.61%	Full spectrum (128)	-25.04	3
Cars	0.73	0.05	0.71	0.06	10.90%	472-761nm(20)	-91.30	2
	0.75	0.05	0.72	0.06	10.90%	Full spectrum (128)	-91.30	2

Units of RMSE for *Cab* and *Cars* [mg g^{-1} dry matter]; *Cab* is the chlorophyll a+b content, *Cars* is the carotenoids content.

Table 3-3 The interpretation of wavebands selected by SMLR model from laboratory spectra resampled to HyMap wavebands

Biochemical	Input data (nm)	Selected wavebands in (nm) (interpretation in brackets related to known absorption features denoted as @)
Cab	472-761	472(*Cb*@460),549(green peak),761(Red edge)
	Full spectrum	403(blue),549(green peak),1346(W@1400)
Cars	472-761	472(*Cb*@460),549(green peak)
	Full spectrum	549(green peak),1401(W@1400)

Cb is the chlorophyll b; W is water; Red edge (rapid change of vegetation reflectance in near infrared regions is the end of chlorophyll absorption); wavebands at 472nm was related (located in ±12 nm of an absorption feature of a biochemical *Cb* with which biochemical of interest was correlated); wavebands at 1401 nm was located around absorption feature of water; waveband at 549nm in the green peak of reflectance. These interpretation from absorption feature used in Kokaly & Clark, (1999) study.

With laboratory spectra, the best SMLR predictive model to predict the chlorophyll a+b concentrations was found with the input data of 472-761nm (lower AIC value and lower nRMSECV) (Table 3-2). There is no difference between the full spectrum and the reduced spectrum (471-761nm) SMLR models to predict carotenoids (same nRMSECV of 10.9% and same AIC value) (refer Table 3-2). The analysis of the selected wavebands by the full spectrum SMLR model revealed that the selected wavebands at 1400 nm related to water absorption bands (Table 3-3). The wavebands chosen by the reduced spectrum SMLR model to determine the leaf chlorophyll a+b and carotenoids concentrations were located in the spectral regions that are known to be important to detect these pigments (Table 3-3). However, they are not located in the main absorption wavelengths. The scatter plots show linear relationships between measured and estimated values of *Cab* and measured and estimated values of *Cars* from SMLR model (Figure 3-2).

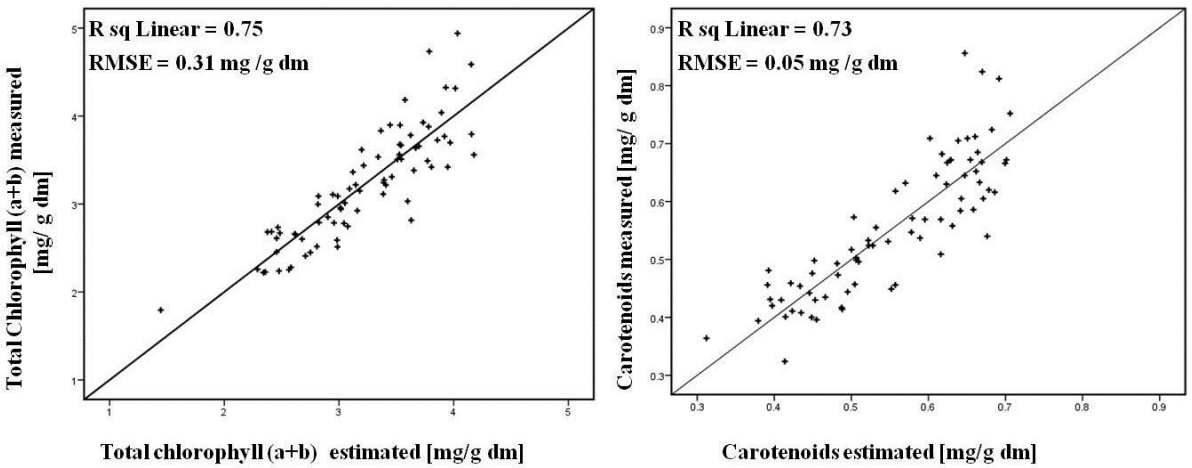


Figure 3-2 Measured against estimated leaf chlorophyll a+b (*Cab*) concentration and measured against estimated carotenoids (*Cars*) using laboratory spectra from training of SMLR model using 472-761nm as input data

The derived results of the model using SMLR was in agreement with comparable previous laboratory spectra studies. For example, Curran et al., (2001) estimated the chlorophyll content in pine needles for a pigment in the range of 0.19 to 2.05 mg g^{-1} with a R^2 of 0.74 using SMLR. Yoder & Pettigrew-Crosby, (1995) estimated the chlorophyll content in maple leaves from reflectance spectra with a R^2 of 0.72 using SMLR. Curran et al., (2001) estimated pine needles chlorophyll a, chlorophyll b and chlorophyll a+b with

a RMSE of 0.03, 0.03, and 0.011 mg g^{-1} respectively from reflectance spectrometry using SMLR. O'Neill et al., (2002) estimated the chlorophyll concentration for Sitka spruce with a R^2 of 0.73 for calibration.

3.3.2. Prediction results of SMLR from HyMap image spectra

Table 3-4 show the results of SMLR model obtained from HyMap image spectra.

Table 3-4 SMLR models from HyMap spectra (n=13) to predict chlorophyll a+b and carotenoids at canopy scale

Biochemical	Training		Cross-validation				
	R^2	RMSE	R^2CV	RMSECV	nRMSECV	Wavebands selected(nm)	Input Data (Total No. of predictors in brackets)
Cab	0.84	0.11	0.72	0.15	4.6%	738,692	554-768nm(15)
	0.69	0.15	0.60	0.18	5.6%	722,692	Full spectrum(126)
Cars	0.37	0.03	0.20	0.04	7.2%	-	554-768nm(15)
	No significant model						Full spectrum(126)

Units of RMSE for *Cab* and *Cars* [mg g^{-1} dry matter]; *Cab* is the chlorophyll a+b content; *Cars* is the carotenoids content

At canopy scale, input data was limited to the 554-768nm (15 predictors) in order to avoid overfitting with the small number of samples (13). With canopy spectra, the best SMLR predictive model to predict chlorophyll concentrations were found with the input data of 554-768nm with a R^2CV of 0.74 and accurate (nRMSECV of 4.6%), rather than the full spectrum (refer Table 3-4). The wavebands chosen by the full spectrum and the reduced spectrum (554-768nm) SMLR models to predict the leaf chlorophyll concentrations varies because of the different continuum removal values; however they are located very close together in the spectral regions. The scatter plots between measured and estimated values of *Cars* at the canopy scale show the cross validated points in the scatter diagram are away from 1:1 line indicates poor predictions (Figure 3-3). The predictive SMLR model to estimate total chlorophyll concentration with CR spectra was based on the wavebands at 692nm (major absorption wavelengths) and along the red edge (738nm).

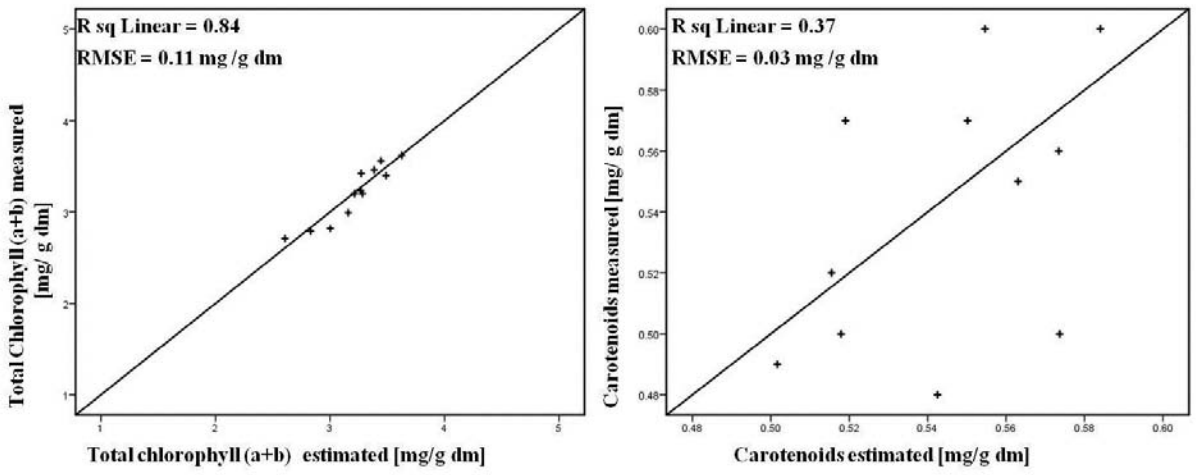


Figure 3-3 Measured against estimated leaf chlorophyll a+b (*Cab)* concentration and measured against estimated carotenoids (*Cars*) using HyMap image spectra from training of SMLR model

The obtained SMLR model accuracy was in agreement with previous comparable airborne hyperspectral studies. For instance, Moorthy et al., (2008) mapped the chlorophyll a+b concentration for conifer needles using coupled leaf and canopy models with a RMSE of 5.3 μg cm^{-2} for a pigment in the range of 25.7 -45.9 μg cm^{-2}. Zarco-Tejada et al., (2004) mapped the chlorophyll content by using coupled leaf and canopy models for conifer needle with a RMSE of 8.1 μg cm^{-2} for a pigment in the range of 26.8-56.8 μg cm^{-2}. Similarly, Darvishzadeh et al., (2008) mapped the canopy chlorophyll content by SMLR with a R^2CV of 0.72 and RMSECV of 0.33 g m^{-2} in grasslands using HyMap data. Mutanga & Skidmore, (2004) mapped the canopy nitrogen concentrations (comparable to chlorophyll) in grass using continuum removed HyMap spectra with a correlation coefficient (r) of 0.78 for calibration and RMSE of 0.05 for validation from SMLR. Huber et al., (2008) predicted canopy nitrogen (comparable to chlorophyll) in a mixed forest species using continuum removed spectra from HyMap with a RMSE of 0.411 of dry weight for validation using SMLR.

3.4. Partial Least Squares Regression models

3.4.1. Prediction results of PLSR from laboratory spectra

The PLSR model was developed using laboratory spectra resampled to HyMap wavebands and their results follows:

Table 3-5 PLSR model from laboratory spectra (n=78) resampled to HyMap wavebands for the prediction of chlorophyll a+b and carotenoids at leaf scale

Biochemical	Training		Cross validation			Input data (Total No. of predictors in brackets)	No.of PLS factors	AIC
	R^2	RMSE	R^2CV	RMSE CV	nRMSE CV			
Cab	0.78	0.30	0.76	0.31	9.8%	472-761nm (20)	4	-31.67
	0.85	0.25	0.78	0.30	9.6%	Full Spectrum(128)	7	-26.78
Cars	0.77	0.05	0.74	0.05	9.1%	472-761nm(20)	4	-93.48
	0.82	0.04	0.76	0.05	9.1%	Full Spectrum(128)	7	-87.48

Units of RMSE for *Cab* and *Cars* [mg g^{-1} dry matter]; *Cab* is the chlorophyll a+b content, *Cars* is the carotenoids content.

Table 3-6 The wavebands selected by the PLSR model from laboratory spectra resampled to HyMap waveband for the estimation of foliar biochemical

Biochemical	Input Data(nm)	Wavebands(nm)	
		Positive coefficient	Negative coefficient
Cab	472-761	472, 487, 503, 656, 671	533, 563, 731
	Full Spectrum	403, 447, 457, 487, 503	533, 702, 1346, 1802,
Cars	472-761	685, 671, 472	533, 549, 563
	Full Spectrum	671, 965	533, 702, 1334

The best PLSR model for the estimation of *Cab* and *Cars* were found with the input data of 472-761 nm based on the lower AIC value (Table 3-5). The important wavelengths to predict the chlorophyll a+b concentrations were located in the 470-480 nm (blue regions), 503-563nm (green peak) and 656-731nm (red edge) and for carotenoids prediction wavebands were located in the 472nm, 533-563nm and 670-680 nm (Table 3-6). The analysis of the selected wavebands by the PLSR model revealed that the differences

in reflectance between healthy and stressed vegetation due to changes in pigment level can be detected in the green peak and along the red edge (Gitelson et al., 2005; Haboudane et al., 2002; Mutanga & Skidmore, 2007). This result confirms the previous studies report that the wavebands at 472nm (blue spectral region) is strongly responsible for the absorption of *Cars* pigment (Blackburn, 1998b; Penuelas et al., 1995). Figure 3-4 shows the scatter plots between measured and predicted values of *Cab* and measured and predicted values of *Cars* and the relationship is near linear.

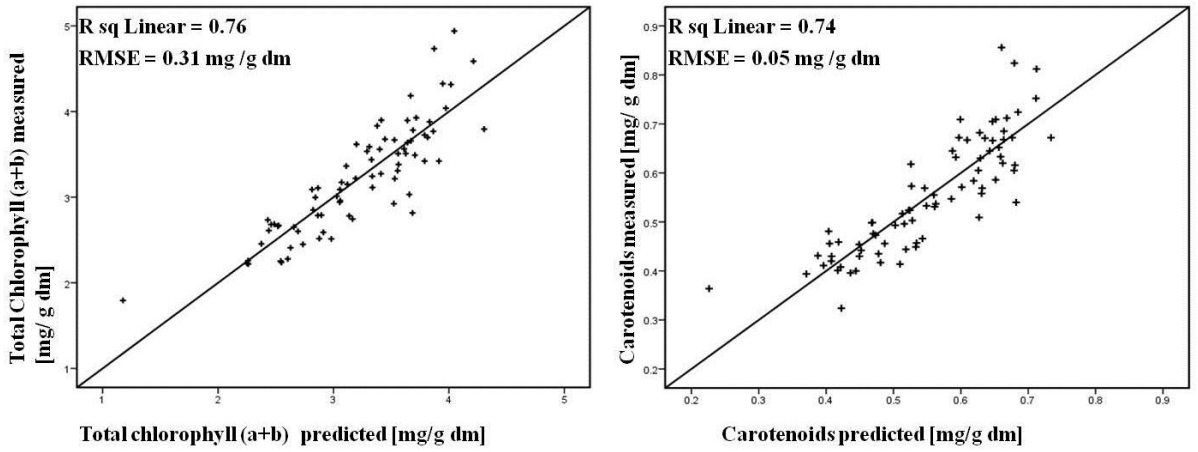


Figure 3-4 Measured against predicted leaf chlorophyll a+b (*Cab*) and measured against predicted carotenoids (*Cars*) using laboratory spectra from cross validation of PLSR model using 472-761nm as input data

The derived results of PLSR model was in agreement with comparable previous laboratory spectra studies using similar model. For example, Xiaobo et al., (2010) predicted chlorophyll content a+b and carotenoids in cucumber leaves for a pigment in the range of 0.6 to 4.0 mg g^{-1} dry matter and 0.46 – 2.16 mg g^{-1} dry matter respectively, using spectrum from 430 to 770 nm, to predict *Cab* with a R^2 of 0.80 (calibration) and RMSE (validation) is 0.5 with 7 PLS factors, to predict *Cars* with a of R^2 of 0.86 (calibration) and RMSE (validation) of 0.2 with 8 PLS factors. Asner & Martin, (2008) estimated leaf carotenoids content in tropical forest species with a R^2 of 0.89 for calibration and RMSE of 0.19 mg g-1 for a pigment range of 0.36 to 2.68 mg g-1.

3.4.2. Prediction results of PLSR from HyMap image spectra

Table 3-7 shows the results of the PLSR model obtained from HyMap image spectra for the prediction of *Cab* and *Cars*.

Table 3-7 PLSR model from HyMap (n=13) for the prediction of chlorophyll a+b and carotenoids at canopy scale

	Training		Cross-validation				
Biochemical	R^2	RMSE	R^2CV	RMSECV	nRMSECV	No.of PLS factors	Input data (nm)
Cab	0.72	0.15	0.60	0.18	5.6%	2	554-768
Cars	0.39	0.03	0.20	0.04	7.2%	2	554-768
Cab	No significant model						Full spectrum
Cars	No significant model						Full spectrum

Units of RMSE for *Cab* and *Cars* [mg g^{-1} dry matter]; *Cab* is the chlorophyll a+b content, *Cars* is the carotenoids content.

The wavelengths selected by the PLSR model to predict the chlorophyll a+b concentrations were at the major absorption wavelengths (661-677 nm) and along the red edge (722-738nm), which is related to the spectral regions contributes for absorption. The result for the prediction of *Cars* was poor (Table 3-7). The PLSR model didn't result in any significant model with the input data of full spectrum (Table 3-7). Figure 3-5 shows the scatter plots between measured and predicted values of *Cab* and measured and predicted values of *Cars* at the canopy scale.

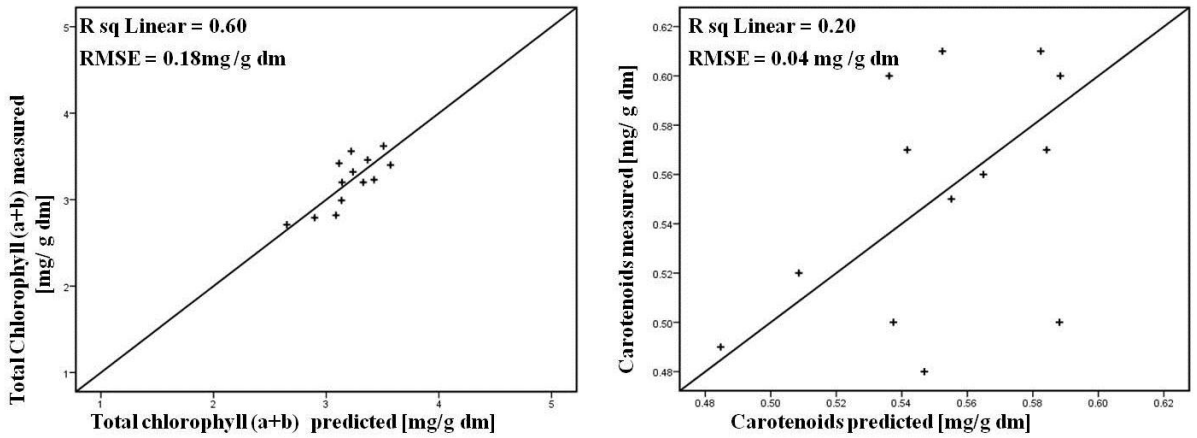


Figure 3-5 Predicted against measured leaf chlorophyll a+b (*Cab*) and predicted against measured carotenoids (*Cars*) using HyMap image spectra from cross validation of PLSR model

The derived result of the PLSR model to predict chlorophyll a+b concentrations confirm the findings of comparable airborne hyperspectral studies using similar model. For example, Darvishzadeh et al., (2008) estimated the canopy chlorophyll content in heterogeneous grassland ecosystem using first derivative reflectance spectra of HyMap data from PLSR model for a pigment in the range of 17.1 to 49.66 gm^{-2} with a R^2CV and RMSECV of 0.66 and 0.37 respectively. Martin et al., (2008) estimated the canopy nitrogen (comparable with chlorophyll) in a mixed forest species from AVIRIS data using PLSR with a R^2 of 0.69 – 0.89 for validation in different sites. Goodenough et al., (2009) mapped the canopy chlorophyll b for Douglas fir species from AVIRIS data using PLSR with a R^2 of 0.85 for calibration. Huang et al., (2004) estimated the nitrogen concentrations in eucalyptus foliage from HyMap data using PLSR with a R^2 0.75 for calibration. However, lower accuracy was obtained for predicting the canopy chlorophyll content in eucalyptus foliage using the reflectance indices of wavebands 850nm, 710nm and 680nm from Hyperion data with a correlation coefficient of 0.42 (Coops et al., 2003a).

3.5. Boosted Regression Tree models

3.5.1. Prediction results of BRT from laboratory spectra

Table 3-8 shows the results of BRT model prediction from laboratory spectra resampled to HyMap wavebands.

Table 3-8 BRT model from laboratory spectra resampled to HyMap wavebands (n =78) for the prediction of *Cab* and *Cars* at leaf scale

Bio-chemical	Training		Cross-validation			Input data (nm)	No.of predictors selected	AIC	Important wavebands selected(nm)
	R^2	RMSE	R^2 CV	RMSE CV	nRMSE CV				
Cars	0.80	0.05	0.68	0.07	12.7%	472-761	20	-81.03	716, 549, 563
Cab	0.78	0.30	0.54	0.47	14.9%	472-761	14	-65.04	716, 549, 563
Cars	0.92	0.03	0.65	0.07	12.7%	Full spectrum	112	107.27	716, 549, 563
Cab	0.82	0.38	0.56	0.43	16.4%	Full spectrum	103	123.84	716, 549, 563
Cars	0.72	0.06	0.68	0.07	12.7%	716,549,563	-	-84.08	716, 549, 563
Cab	0.70	0.41	0.58	0.46	14.5%	716,549,563	-	-20.49	716, 549, 563

Units of RMSE for *Cab* and *Cars* [mg g^{-1} dry matter]; *Cab* is the chlorophyll a+b content, *Cars* is the carotenoids content.

The most important wavebands to predict chlorophyll a+b and carotenoids concentrations were located in the green peak of the reflectance and in the narrow bands of red edge. But the unimportant predictors were also used by the BRT model. The wavebands at the 716, 549 and 563nm have more contribution above 10% for the prediction of *Cab*. The rest of the wavebands have less than 10% contribution for the prediction of chlorophyll a+b (Figure 3-6 and Figure 3-7). This was found similar for the prediction of *Cars* also (not shown here). The model with the input data of full spectrum and the reduced spectrum of absorption features (471-761nm) showed the same wavebands as high contribution for prediction (Table 3-8). This highlights that information for the prediction of *Cars* and *Cab* was available only in the 716, 549 and 563nm spectral regions. For the simplification of BRT model, the most important predictors (716, 549 and 563nm) were included in the model. It was found that the model performance was similar in validation with the input data of wavebands at 716, 549 and 563nm and 472-761nm. The models calibrated well but lack in the validation errors increased with the left out samples which resulted in nRMSECV higher than calibration. This problem can be solved by using more generalized larger samples for better prediction.

Similarly, Jungho Im et al., (2009) estimated nitrogen using BRT for short duration woody crops (60 samples) using the 63 reflectance variables of hyperspectral data from 400 to 980nm, with an accuracy of R^2 (calibration) of 0.79 and cross validation RMSE of 0.48 and found that BRT model performs well in calibration but lacks in validation errors.

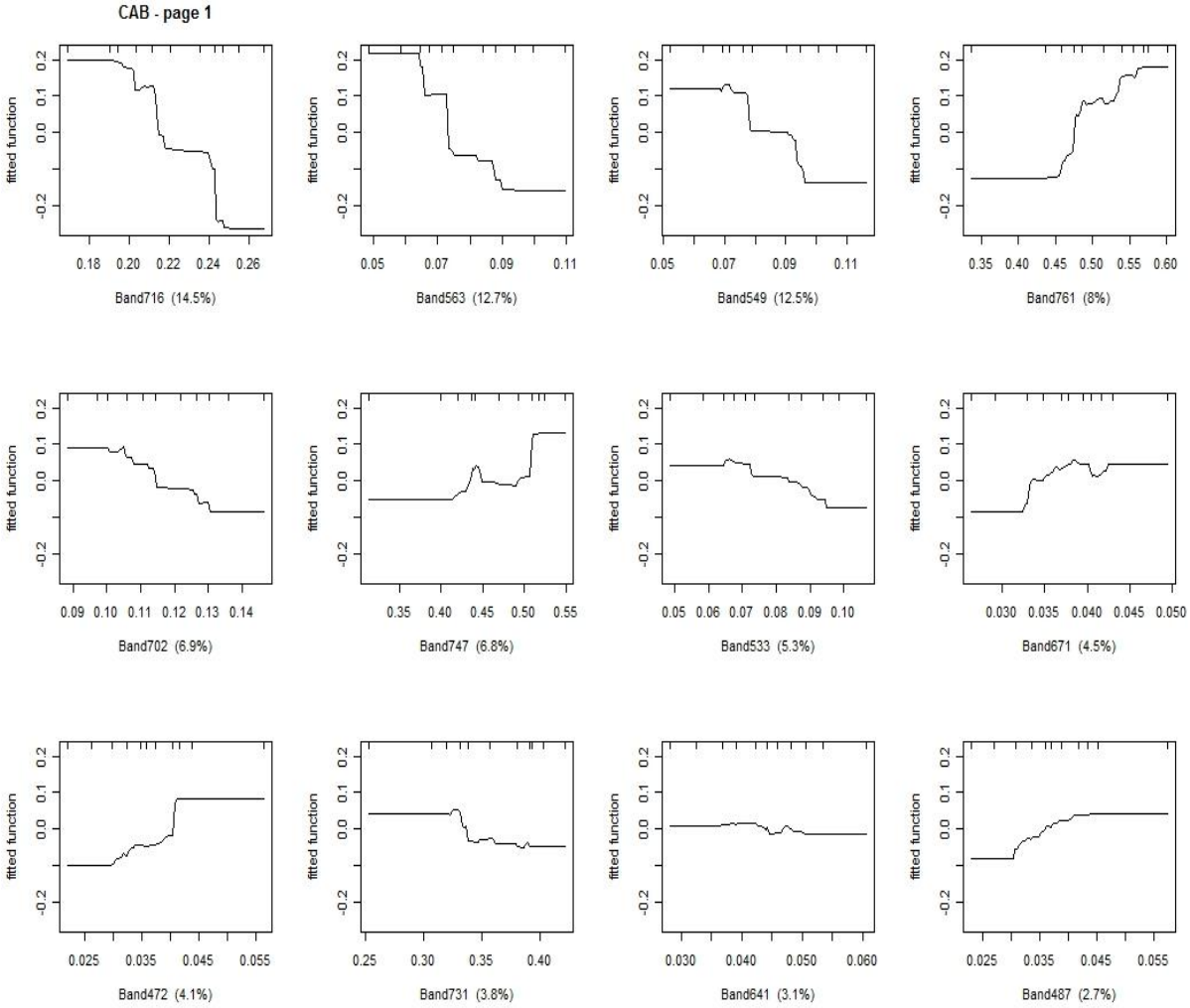


Figure 3-6 Fitted functions of wavebands by BRT model for the prediction of *Cab* (n=78) in R Package; the value in percentage indicates wavebands contribution to predict *Cab*

Y axis = residuals of predicted variable; X axis = spectral reflectance

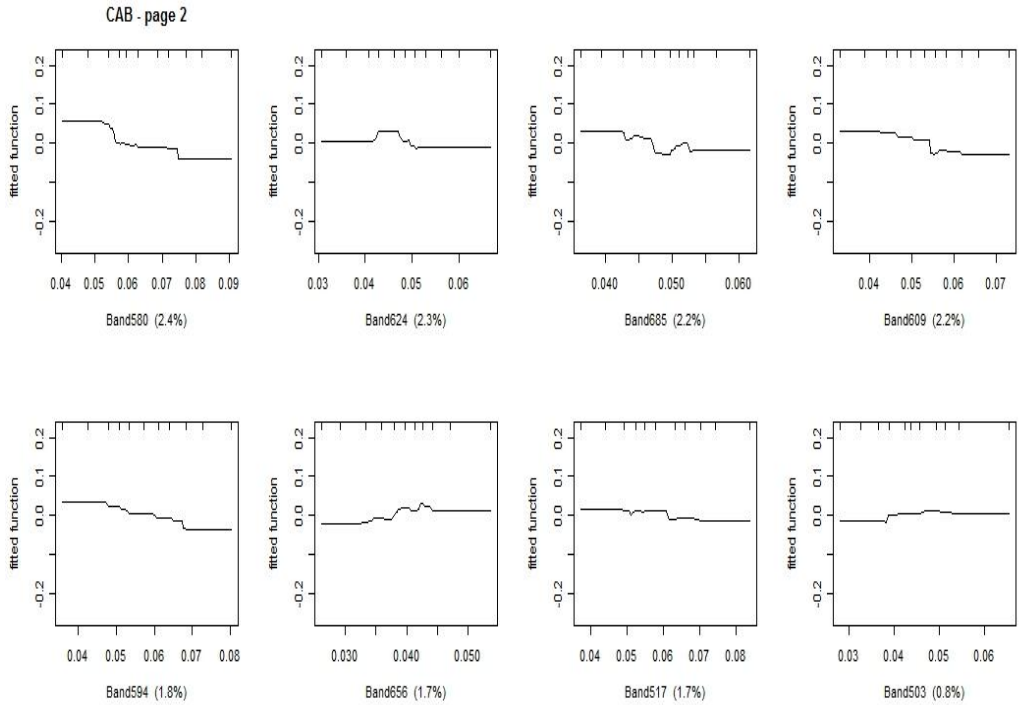


Figure 3-7 Fitted functions of wavebands by BRT model for the prediction of *Cab* (n=78) in R Package; the value in percentage indicates wavebands contribution to predict *Cab*.

Y axis = residuals of predicted variable; X axis = spectral reflectance of bands.

3.6. Models comparisons

The selected models from laboratory spectra and airborne HyMap spectra were compared based on the cross-validated statistic, AIC value and the selected wavebands. In all the models (SMLR, PLSR and BRT), the reduced spectrum models (472-761nm) were found to predict chlorophyll a+b and carotenoids better than using the entire spectrum as input data at the leaf scale.

3.6.1. Comparing the predictive ability of models at leaf scale

Comparison of the selected wavebands by SMLR, PLSR and BRT models to predict chlorophyll a+b and carotenoids (Table 3-3, Table3-6 and Table 3-8) revealed that the PLSR model was alone based on the wavebands located in the major absorption wavelengths (650-670nm). These are the indicative for better performance of the PLSR model than SMLR and BRT models. When analysing the selected wavebands in all models, revealed that the BRT model (refer table 3-8) was consistent in highlighting the same wavebands to predict chlorophyll a+b and carotenoids concentrations with the full spectrum and also with the reduced spectrum (472-761nm) as input data. This was found to be a limitation in the SMLR and PLSR models (refer Table 3-3 and Table 3-6). However, the BRT model failed to select the major absorption wavelengths (650-670nm). The wavebands chosen by the BRT model to determine the chlorophyll a+b concentrations were located in the narrow bands of the red edge (716nm) and wavebands along the edge of the green peak (549 and 563nm). This was also the case with this model for the prediction of carotenoids concentrations. The wavebands chosen by the SMLR and PLSR models to predict the chlorophyll a+b concentrations were in agreement with previous studies that the wavebands corresponding for the prediction of these pigments concentrations were available approximately in 530-630 nm and narrower bands around 700nm in many leaves (Carter & Knapp, 2001; Curran et al., 2001; Gitelson et al., 2003; Imanishi et al., 2010; Yoder & Pettigrew-Crosby, 1995). The result reflects that the

visible and red edge spectral regions are very important for the prediction of these pigments (Gitelson et al., 2005). The result highlights physical principle that the blue spectral regions are strongly responsible for the absorption of chlorophyll a and chlorophyll b (Hansen & Schjoerring, 2003).

The analysis of the selected wavebands by SMLR and PLSR models to determine the carotenoids content highlights that they strongly absorb radiation in the 400nm–500nm spectral regions (Dash et al., 2008; Gitelson et al., 2002). All models selected the wavebands in the spectral regions (blue, green peak and red-edge regions) which are reported by many researchers in literature to predict chlorophyll a+b and carotenoids concentrations. Table 3-9 and Table 3-10 show the cross validation performance of SMLR, PLSR and BRT models.

Table 3-9 Predictive models of chlorophyll a+b concentration from laboratory spectra resampled to HyMap wavebands (n=78) using input data of 472-761nm

Model	Cross-validation			AIC	Selected Wavebands (nm)	No .of Predictors
	R^2CV	RMSECV	nRMSECV			
SMLR	0.72	0.33	10.4%	-31.56	472, 549, 761	3
PLSR	0.76	0.31	9.8%	-31.67	472, 487, 503-563, 656, 671, 731	4(PLS Factors)
BRT	0.58	0.46	14.5%	-20.49	716,549,563	3

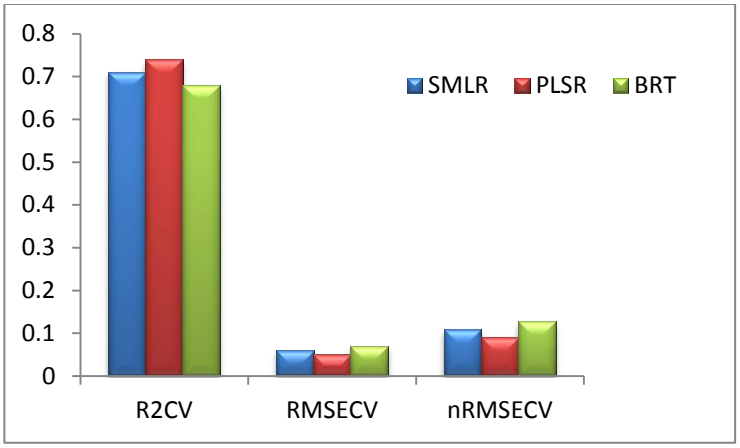


Figure 3-8 Comparison plots of R^2CV, RMSECV and nRMSECV for the prediction of chlorophyll a+b concentration from SMLR, PLSR and BRT using laboratory spectra

Table 3-10 Predictive models of carotenoids from laboratory spectra resampled to HyMap wavebands (n=78) using input data of 472-761nm

Model	Cross-validation			AIC	Selected Wavebands (nm)	No .of predictors selected
	R^2CV	RMSECV	nRMSECV			
SMLR	0.71	0.06	10.9%	-91.30	472, 549	2
PLSR	0.74	0.05	9.1%	-93.48	472,533-549,670-680	4(PLS Factors)
BRT	0.68	0.07	12.7%	-84.08	716,549,563	3

Figure 3-8 and Figure 3-9 show a graphical comparison of cross-validated statistics of SMLR, PLSR, and BRT models to predict the chlorophyll a+b and carotenoids concentrations respectively. With laboratory spectra, the PLSR model resulted in better predictive performance (most accurate) than SMLR and BRT

models which has higher R^2CV, lower RMSECV and lower nRMSECV to predict *Cab* and *Cars* (Table 3-9, Table 3-10). On the ground of AIC value, the PLSR was found to be best parsimony model to predict *Cab* and *Cars* which has very low value (Table 3-9 and Table 3-10). However, the differences between SMLR and PLSR were in terms of nRMSECV less than 1%. The simple SMLR model was found to predict chlorophyll a+b and carotenoids concentrations with reasonable accuracy at the leaf scale.

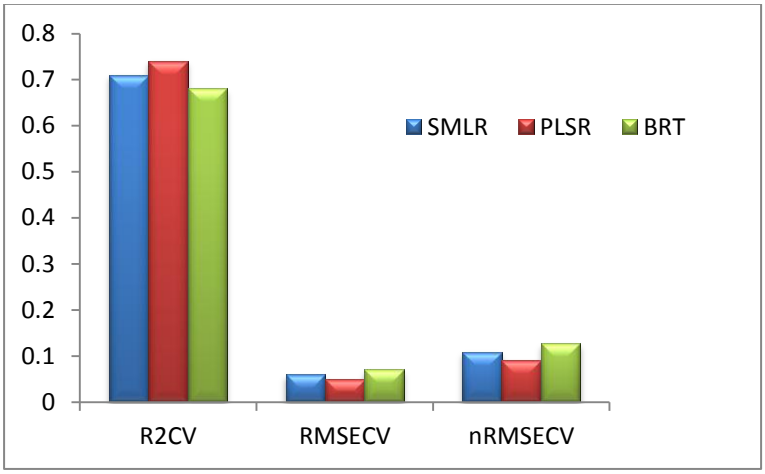


Figure 3-9 Comparison plots of R^2CV, RMSECV and nRMSECV for the prediction of carotenoids concentration from SMLR, PLSR and BRT using laboratory spectra

However, the SMLR model can be criticised that the selected wavebands failed to correspond with the major absorption wavelengths (650-670nm) to predict the chlorophyll a+b concentrations. In contrasts, the PLSR model provided better results in prediction (most accurate than SMLR) with the selected wavebands located in the major absorption wavelengths (650-670nm). The better performance of the PLSR model than SMLR model at the leaf scale confirm the findings of several researchers (Atzberger et al., 2010; Coops et al., 2003b; Darvishzadeh et al., 2008; Nguyen & Lee, 2006; Smith et al., 2003).

The BRT model was found to predict chlorophyll a+b and carotenoids in lower R^2CV, higher RMSECV and high AIC value as compared with SMLR and PLSR models (refer Figure 3-8, Figure 3-9). The reason is that SMLR and PLSR models used only one data for validation (leave-one out cross validation), but the BRT model used the 50% random subsets of data for validation (n. fold cross validation). The model comparisons can be criticised based on the use of different cross validation procedure. The justification behind the use of different cross validation is that, the n. fold cross validation is an efficient procedure to assess the stability of BRT model (Friedman, 2002); and the leave one out cross validation is an efficient method which provides true unbiased generalization errors in linear models (Cawley & Talbot, 2003). Even the simple linear model (SMLR) predicts with the least error than BRT model. The reason is that the relationship between reflectance (recorded in the controlled environment) and pigments are linear. But, at the canopy scale multiple factors (i.e. orientation of leaves, % canopy ground coverage and soil/litter surface reflectance) influences the reflectance which allows nonlinearity (Barry et al., 2009; Blackburn, 2007; Chen et al., 2007; Skidmore et al., 2010). The BRT model has the potential to model nonlinear relationships between reflectance variables and pigments. The flexibility of BRT modelling allows for overfitting which requires careful consideration to simplify the predictor variables by using most important predictors based on their contribution. The BRT model statistically differs from conventional linear models. The BRT model can be criticised on the ground that there is no P-value to determine the significance of predictors, model coefficients and degrees of freedom in the model. However, the BRT model was found to perform well in calibration equalled to PLSR and SMLR models, but it lacks only in the validation error due to the lack of larger generalized sample size. The predictive performance of BRT is stochastic (i.e., random or probabilistic) which requires the larger samples of generalized to make random subsets of data for training and validation (Im et al., 2009). Due to the lack of larger and

generalized sample data, this research failed to reject null hypothesis that BRT does not estimate foliar biochemical concentration most accurate than PLSR and SMLR. Similarly, Im et al., (2009) compared the PLSR and BRT for the prediction of biophysical and biochemical concentrations for crops using hyperspectral bands and found that the PLSR model resulted in good performance, but the BRT with a larger sample size appears to be a robust model due to its flexibility modelling to sample data.

3.6.2. Comparing the performance of models at canopy scale

The sample size of 13 didn't support to develop the BRT model. However, SMLR and PLSR models were developed to predict chlorophyll a+b and carotenoids concentration at canopy scale.

Table 3-11 Predictive models of chlorophyll a+b at stand level from HyMap wavebands (n=13)

Model	Cross validation			AIC	No. of predictors selected	Wavebands
	R^2CV	RMSECV	nRMSECV			
SMLR	0.72	0.15	4.7%	-6.71	2	738, 692
PLSR	0.60	0.18	5.6%	-5.68	2 (PLS Factor)	707, 722, 738 , 661, 677

The selected wavebands by SMLR and PLSR models explain that the wavebands at the major absorption features (650-670nm) are very sensitive to pigment concentrations at the canopy scale (Blackburn, 1998a; Sari et al., 2005). The result highlights that chlorophyll is a strong absorber of radiation and the chlorophyll absorption feature is not masked by leaf water (Schlerf et al., 2010). The model comparison of SMLR and PLSR at the canopy scale could not be statistically reliable due to the limited number of samples. However, the SMLR model resulted in higher R^2CV, lower nRMSECV and lower AIC value as compared to PLSR model (Table 3-11).

Table 3-12 Predictive models of carotenoids at stand level from HyMap wavebands (n=13)

Model	Cross validation		
	R^2CV	RMSECV	nRMSECV
SMLR	0.20	0.04	7.2%
PLSR	0.20	0.04	7.2%

The results of *Cars* (very low R^2CV) at the canopy scale with SMLR and PLSR models were poor (Table 3-12). The models were accurate (nRMSECV 7.2%) but not precise (R^2CV 0.20). The predictions by an accurate but imprecise model are unrealistic (Tedeschi, 2006). The reason could be that the low concentrations of carotenoids in spruce needles with respect to chlorophyll a+b do not have significant contribution to absorption at canopy reflectance. Carotenoids are generally masked by chlorophyll in most of the leaves (Britton et al., 2008). The absorption features of chlorophyll a+b and carotenoids are highly overlapped each other (Blackburn, 2007). It is very difficult in separating and estimating these pigments in their overlapping absorption features. Foliar biochemicals contribution to the canopy reflectance are influenced by the reflectance of soil, illumination conditions and other canopy biophysical attributes (Asner, 1998). Because of these factors, the low content of *Cars* pigment contribution to the canopy reflectance is very minimal. This is not applicable to laboratory spectra, since the reflectance is recorded in the controlled environment. The estimation of chlorophyll and carotenoids concentrations separately works very well at the leaf scale using empirical methods but fails at the canopy scale or stand levels due to multiple scattering in intact leaves (Ustin et al., 2009). This justifies that why carotenoids concentrations were estimated only at the leaf scale from laboratory spectra but not in the canopy scale.

3.7. Assessing the reliability of predictive model parameters at canopy scale

Table 3-13 The Summary statistics of regression coefficients

	Unstandardized coefficients	Standardized coefficeints	Std.error	t-value	P-value	95% confidence Interval
Intercept	33.369	-	4.118	8.102	0.001	24.19, 42.54
Waveband 692nm	8.258	0.865	1.858	4.445	0.001	4.11, 12.39
Waveband 738	-34.822	-1.406	4.821	-7.223	0.001	-45.56, -24.08

The standard error of the model coefficients is the standard deviation of the coefficient. This indicates the measure of the precision with regression coefficient. It is a measure of the typical deviation of a sample estimate of the coefficient from the true parameter value. (Dougherty, 2007). The units of the standard error are the units of corresponding variable. The standard error of the model coefficients from very small sample size may not be reliable (Fox, 1997). The reliability of the regression coefficients was assessed using bootstrapping. Refer Table 3-14 for the results of bootstrap analysis.

The predictive equation:
Cab = -34.822*(CR waveband at 738nm) +8.258*(CR waveband at 692nm) + 33.369..... Equation (1)

Table 3-14 The Summary statistics of the regression coefficients for bootstrap sample

Variables	Original coefficients of sample	Average coefficients of bootstrap sample	Standard error (original sample)	Bootstrap standard error	Bias	95% confidence intervals
Intercept	33.369	33.364	4.118	3.631	-0.005	26.26, 40.49
Band 692	8.258	8.241	1.858	1.641	-0.016	5.057, 11.493
Band 738	-34.822	-34.831	4.821	4.222	0.009	-43.17,-26.50

Bias = the differences between original coefficients of the sample and average coefficients of bootstrapped samples. This indicates the accuracy of an estimator; bootstrap standard error is the standard deviation of the regression coefficient in its bootstrap distribution.

The interpretation of bootstrap results were made from Suat & Dervis, (2007). The bootstrap distributions of the intercept, waveband 692nm and waveband 738nm coefficients were reasonably normal distributed but somewhat heavy-tailed (Figure 3-10, 3-11 below). The confidence intervals of the coefficients were computed based on the bootstrap standard error at 95% level. The differences between the 95% confidence intervals (CI) of the original sample and the bootstrap samples were of marginal. This indicates the precision of the coefficients. The coefficients of the original sample are biased if its bootstrap distribution is not centred at the true value of the parameter (Moore & McCabe, 2006). The parameters or estimators of the model (derived from sample) are consistent and efficient when they are very close to the mean of population (Clark & Hosking, 1986). The coefficients of the model from the original sample lie close to the mean of the bootstrap distribution (Figure 3-10, Figure 3-11). Here, comparing the bootstrap average coefficients with the corresponding coefficients of estimated from the original sample were similar and shows that there is a very little bias in bootstrap coefficients (Table 3-14). Thus, the regression parameters of the model are reliable.

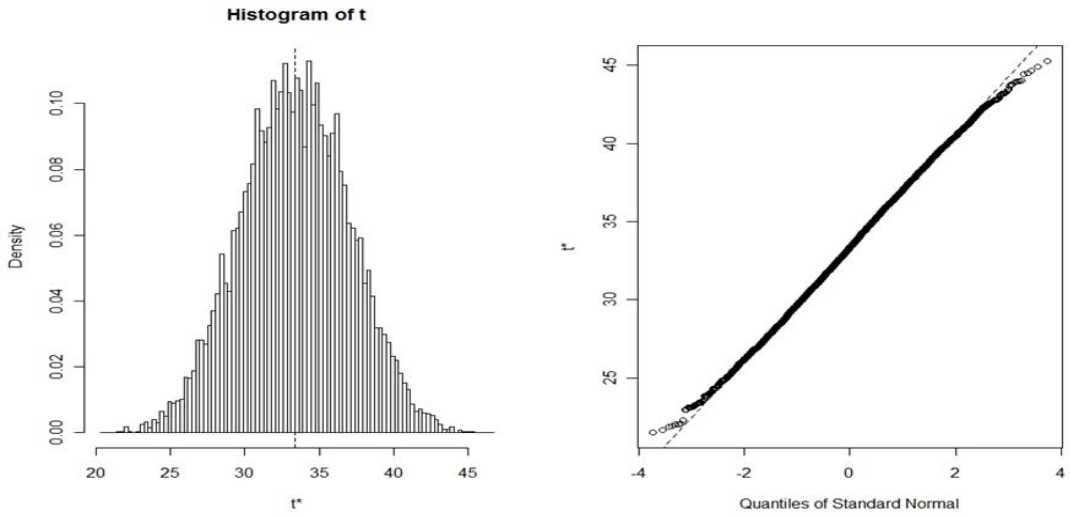


Figure 3-10 Histogram and normal quantile plots for the bootstrap replications (2000 bootstrap samples) of the intercept term; the broken vertical line in histogram shows the location of the regression coefficient for the model fit to the original sample.

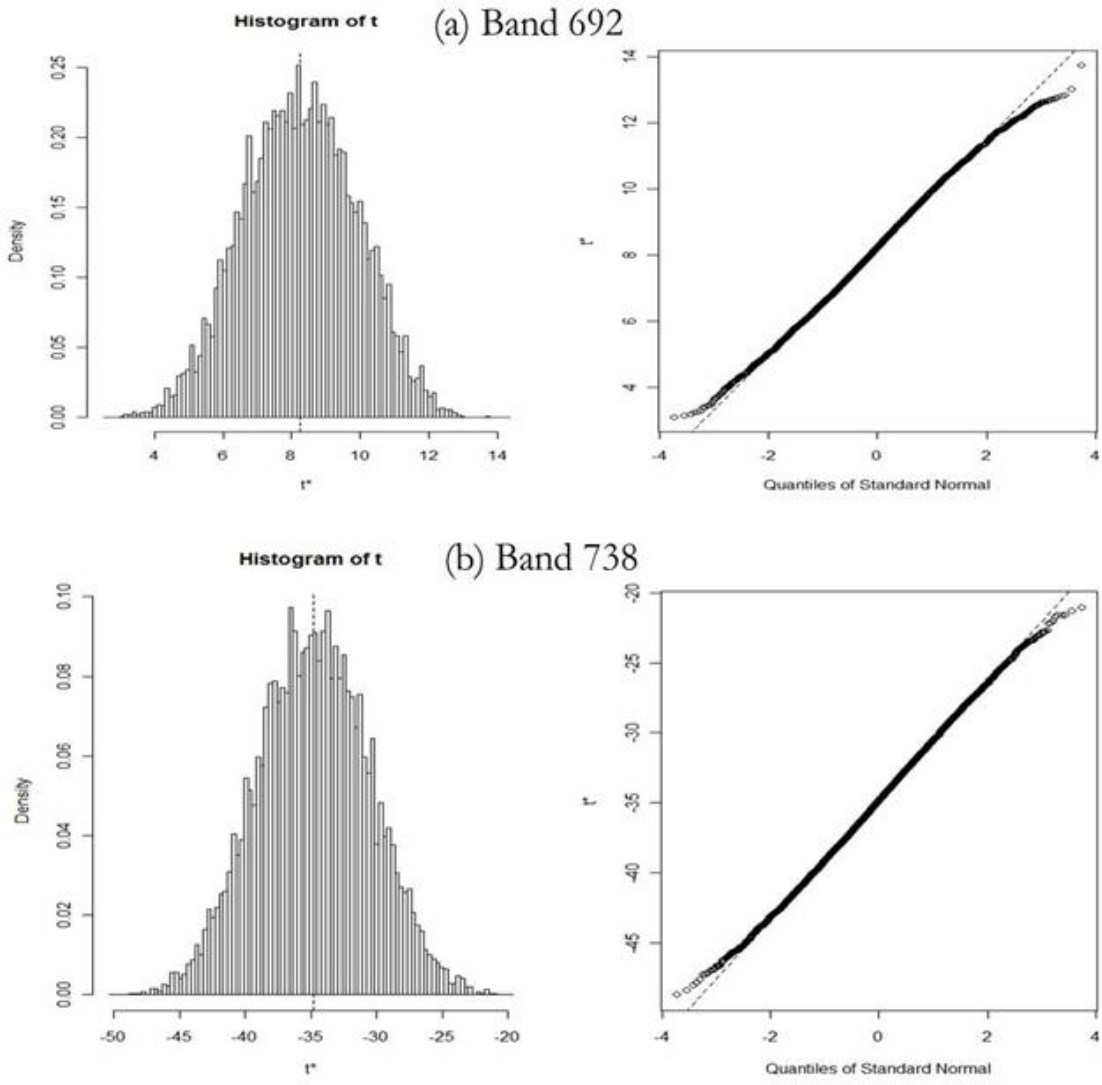


Figure 3-11 Histograms and normal quantile plots for the bootstrap (2000 samples) (a) waveband 692 nm and (b) waveband 738 nm coefficients; the broken vertical line in each histogram shows the location of the regression coefficient for the model fit to the original sample.

3.8. Interpretation of the predictive model

The regression coefficient of the waveband at 738nm indicates a strong negative relationship with chlorophyll a+b concentration. The waveband at 738nm has a larger contribution towards the prediction of chlorophyll than the waveband at 692nm since they have higher magnitude of the standardized coefficient (refer Table 3-13) (Bring, 1994). The negative regression coefficient of the CR waveband at 738nm (located in the red-edge region) agrees with the understanding that decreasing chlorophyll concentration in stressed vegetation leads to a shift of the red edge inflection point towards shorter wavelengths (Figure 3-12). The red-edge region means abrupt reflectance change in the 680-740nm, termed the red-edge position, caused by the combined effects of chlorophyll absorption and leaf scattering (Dawson & Curran, 1998). This result confirms previous study findings that the red edge position shifts according to changes in the chlorophyll content that can detect vegetation health status (Helmi Zullhaidi & Nasrulhapiza, 2009; Mutanga & Skidmore, 2007; Ruiliang et al., 2003). The positive regression coefficient of CR waveband at 692 nm, which is located closely to the major absorption feature (650-670nm) of pigment, agrees with the general understanding that the chlorophyll concentration increase with more absorption. The wavebands selected by the model at the 738 nm and 692 nm reflects that the increase in chlorophyll concentration results in the broadening of the absorption feature centred around 692 nm (red spectrum) due to strong absorption (Dawson & Curran, 1998; Zhang et al., 2007). This causes the red edge position to shift towards longer wavelengths. The low regression coefficient of the CR waveband at 692nm is highly sensitive to high chlorophyll concentrations, whereas the high regression coefficient of the CR waveband at 738nm is highly sensitive to low chlorophyll concentrations.

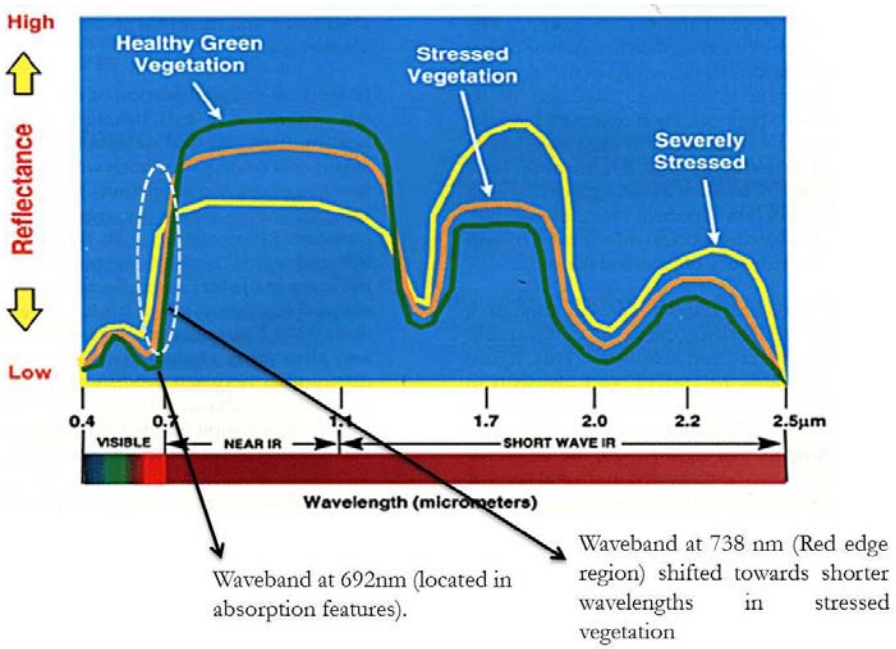


Figure 3-12 The spectral characteristics of healthy and stressed vegetation; Source : NOAA, (2011)

3.9. Map of foliar chlorophyll a+b concentration

The predictive model derived using SMLR from HyMap was most accurate in prediction used to produce the map of leaf chlorophyll a+b (LCC) concentration. Norway spruce stands were digitized based on visual interpretation from HyMap. The predictive model developed using CR spectra and SMLR (Equation 1) was applied on the HyMap image to compute the map of leaf chlorophyll a+b concentration (LCC) for pixels representing spruce alone. The histograms of the LCC (not shown here) from the map revealed that the values are in a range from 2 to 6 mg/g dry matter. The map revealed some distinct spatial

variation; there are some areas in red colour (very high *Cab*) and some areas were in blue (low *Cab*) (Figure 3-13). The detailed interpretation of the spatial patterns of LCC in relation to the underground soil properties were made below (Section 3.10).

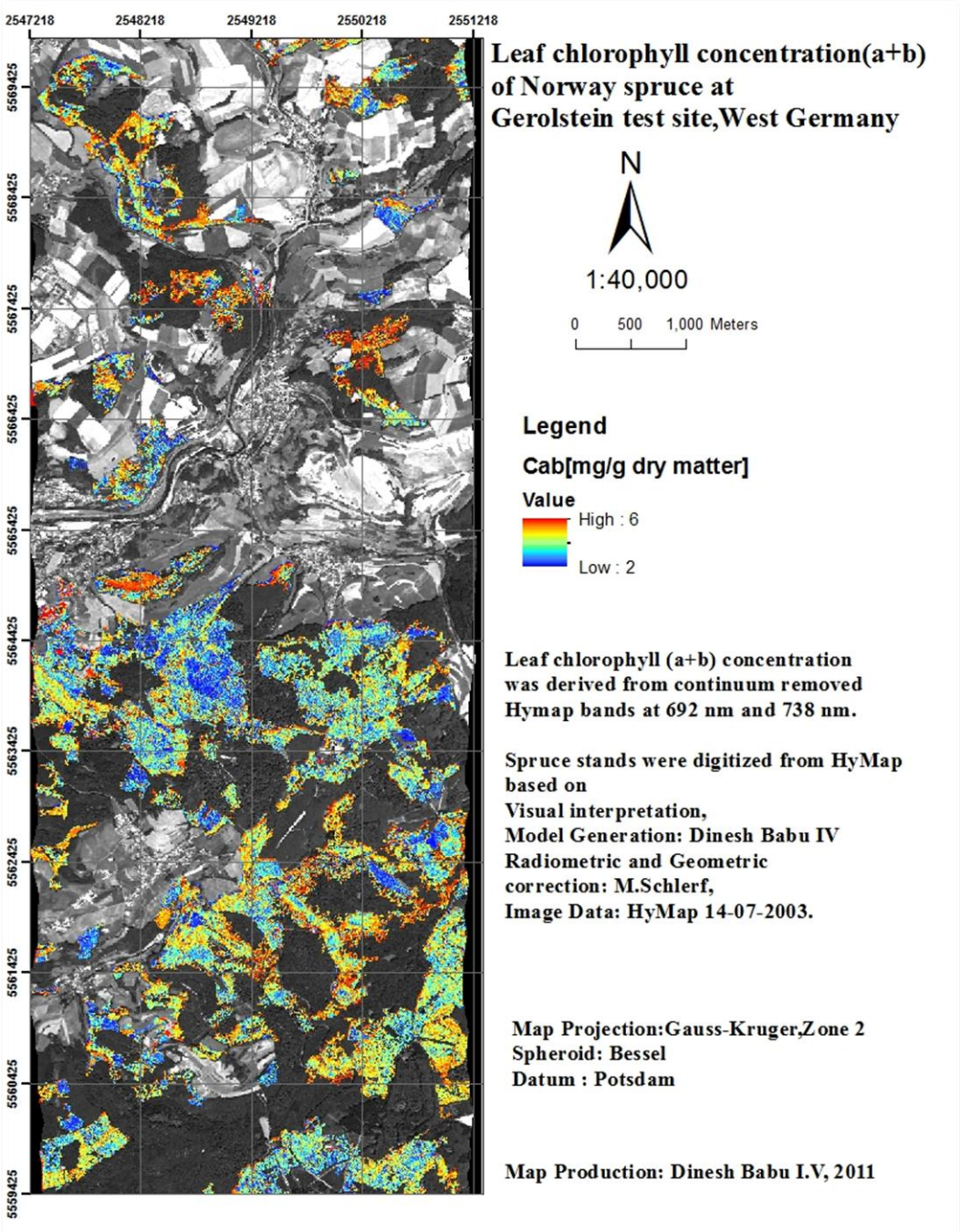


Figure 3-13 Map of leaf chlorophyll concentration a+b for Norway spruce at Gerolstein study site, West Germany

3.10. Investigation of spatial variation in foliar biochemicals

The geology shapefile (represents soil substrates) obtained from the forest management was used to investigate the relationship between the LCC and the underlying soil substrate types. In figure 3-14, the LCC were very low (blue colour) over the buntsandstein geology class and very high (red colour) over the

quaternary geology class. Analysis of Variance (ANOVA) was used to find whether the LCC varied on the different soil substrates.

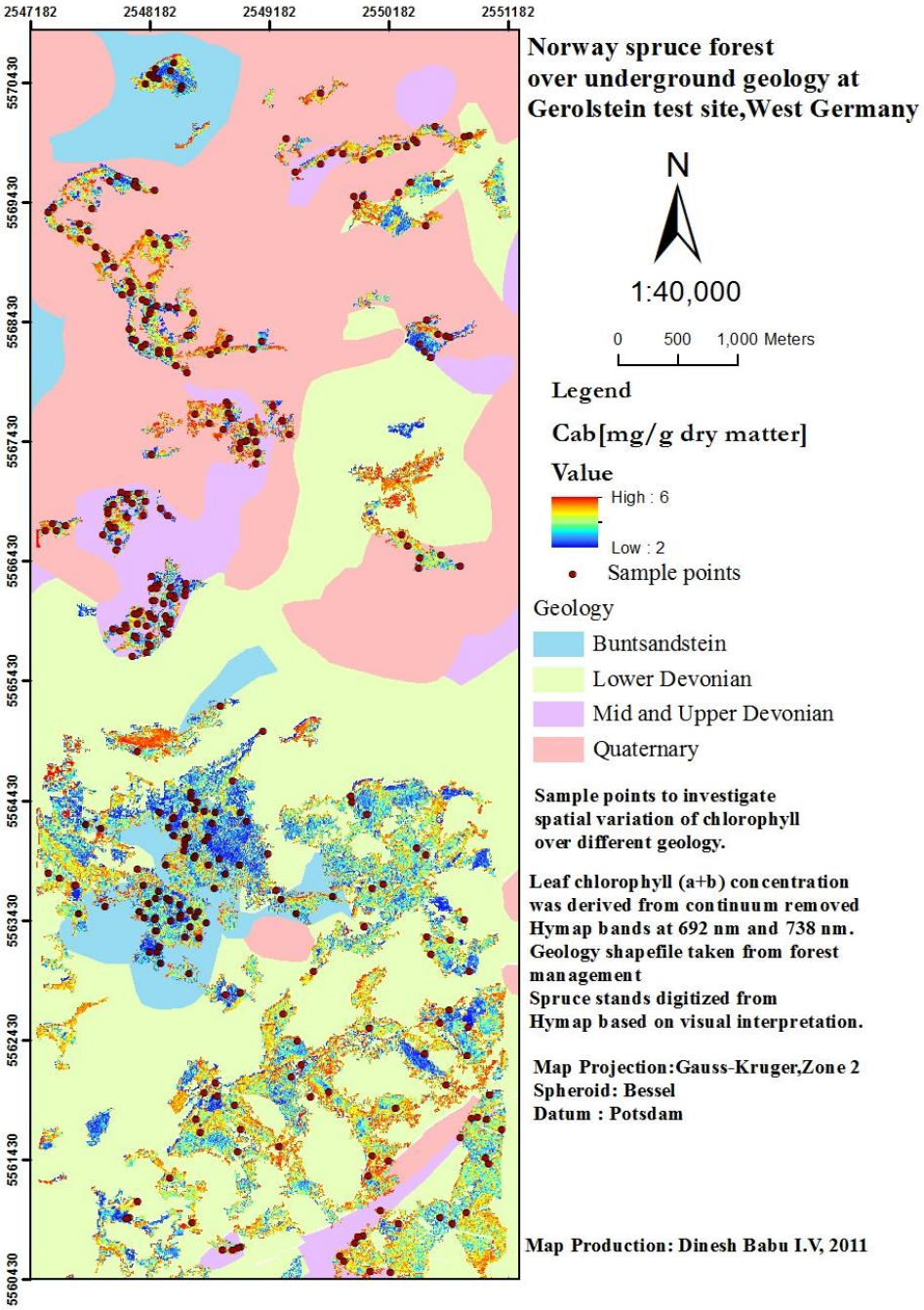


Figure 3-14 Spatial location of Norway spruce stands over underground geology at Gerolstein test site, West Germany

3.10.1. Sample size for ANOVA

The numbers of pixels (map of LCC) in each geology unit were calculated in Arc GIS found to be more than 10000. For a population more than 10000, 96 numbers of samples were considered as representative at 95% level of confidence with ±10% allowable error (refer Table 2-2). The 96 number of samples were taken from each geology unit for ANOVA analysis. The total sample size was 96 X 4 (geology units) = 384. The samples were extracted following method described in section 2.9.

3.10.2. ANOVA results

Table 3-15 Summary of the mean differences and ANOVA results of chlorophyll concentration a+b in different geology units

Groups	Count	Sum	Average	Variance
Quaternary	96	357.357	3.722469	0.573571
Mid-Buntsandstein	96	324.0942	3.375981	0.396279
Lower Devonian	96	344.5398	3.588957	0.416558
Mid and Upper Devonian	96	346.6643	3.611087	0.416155

Source of Variation	SS	df	MS	F	P-value
Between Groups	6.033779	3	2.01126	4.463113	0.004271
Within Groups	171.2434	380	0.450641		
Total	177.2772	383			

The P-value (0.004271) of ANOVA analysis indicates that the differences between the mean of LCC in the different soil substrates (represented by geology class) was significant at 95% confidence level (Table 3-15). Refer table 3-16 and 3-17 for differences between the mean of LCC in different geology class from post–hoc and pairwise t-test respectively.

Table 3-16 Post- Hoc tests of multiple comparisons of leaf chlorophyll a+b concentration in different geology units

(I) Geology	(J) Geology	Mean Difference (I-J)	Significance (Bold indicates significant less than 0.05)	95% Confidence Interval	
				Lower Bound	Upper Bound
Quaternary	Mid-Buntsandstein	0.346	**.000**	0.155	0.537
	Lower Devonian	0.133	.169	-0.057	0.324
	Mid and Upper Devonian	0.111	.251	-0.079	0.301
Mid-Buntsandstein	Quaternary	-0.346	**.000**	-0.537	-0.155
	Lower Devonian	-0.212	**.029**	-0.403	-0.022
	Mid and Upper Devonian	-0.235	**.016**	-0.425	-0.044
Lower Devonian	Quaternary	-0.133	.169	-0.324	0.057
	Mid-Buntsandstein	0.212	**.029**	0.022	0.403
	Mid and Upper Devonian	-0.022	.819	-0.212	0.168
Mid and Upper Devonian	Quaternary	-0.111	.251	-0.301	0.079
	Mid-Buntsandstein	0.235	**.016**	0.044	0.425
	Lower Devonian	0.022	.819	-0.168	0.212

Table 3-17 Pair wise T-test of mean differences between chlorophyll a+b concentration in different geology units

		Paired Differences		
		t	df	Significance(2-tailed)(P-value) (Bold indicates significant less than 0.05)
Pair.1	Quaternary – Lower Devonian	1.246	95	0.216
Pair.2	Quaternary – Mid and Upper Devonian	1.142	95	0.256
Pair.3	Quaternary – Mid-Buntsandstein	3.424	95	**0.001**
Pair.4	Lower Devonian- Mid and Upper Devonian	-0.236	95	0.814
Pair.5	Lower Devonian- Mid-Buntsandstein	2.526	95	**0.013**
Pair.6	Mid and Upper Devonian –Mid-Buntsandstein	2.562	95	**0.012**

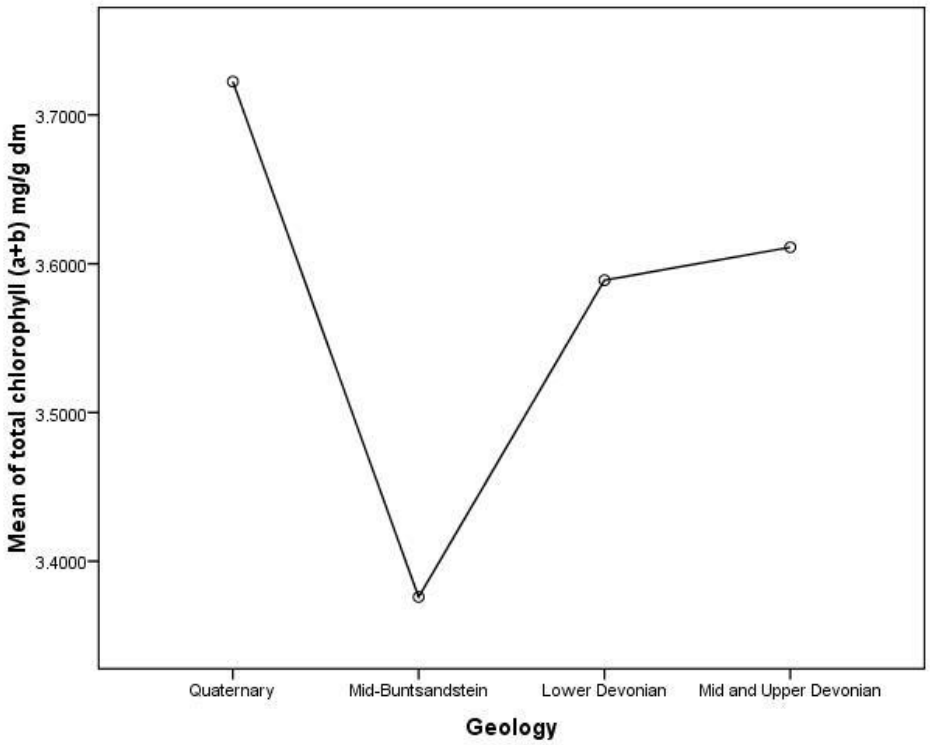


Figure 3-15 The mean plots of chlorophyll a+b concentration in different geology class

Figure 3-15 shows the plot of mean LCC in different geology class. The mean of the LCC over the mid-buntsandstein (represents low soil fertility substrate) was significantly lower than the other soil substrate units at 95% confidence level (Table 3-16 and Table 3-17). But, the mean of LCC in the soil substrates represented by the lower Devonian, quaternary and mid-upper Devonian geology classes were not significantly different from each other. The variation of soil nutrient availability and soil base saturation in the lower Devonian, quaternary and mid-upper Devonian geology units were not reflected in the variation of canopy nutrient concentrations. The reason could be that there exist feedback mechanism between the biochemical constituents of leaves with the soil properties (Sariyildiz & Anderson, 2005). The soil substrates developed from the lower Devonian, quaternary and mid-upper Devonian geology units have the base saturation ranging from high to medium; they have some fractions of clay content which can decompose litter and improve their soil properties, which feedback to tree productivity through nutrient

availability. This feedback mechanism allows trees to lower their dependence on the soil nutrients (Covelo et al., 2008). However, the soils developed from mid-buntsandstein geology are of only sands with the very low base saturation (no clay content), that cannot decompose litter to improve their soil properties. Moreover, trees on soils of low fertility generally produce leaf litters that decompose at slower rates (Hattenschwiler, 2005). Previous studies in forestry have shown that the same tree species growing on different soil types develop the underlying forest floor characteristics that vary in humus and nutrients (Cote et al., 2002; Sariyildiz & Anderson, 2003). This result highlights that only very poor soil nutrient availability conditions result in the depressed foliar biochemical concentrations. This understanding provides a useful guide for the forest management to map and monitor the stands which suffer due to nutrient deficiency by possible fertilization. This result agrees with the expectation of this research that the foliar nutrients of spruce result in spatial variation due to the large variation in the underground soil base saturation.

These results were in agreement with similar type of studies in forestry. For example, Sariyildiz & Anderson, (2005) found that the hemicellulose and lignin composition in the leaves of beech, oak and chestnut significantly vary on the different soil types. Wang & Klinka, (1997) demonstrated that the foliar concentrations of magnesium, potassium, phosphorous and nitrogen were positively correlated with the soil nutrient concentrations in white spruce (*Picea glaunca*). The relationship between the soil nutrients and the foliar nitrogen and phosphorous concentrations in young densely planted mini plots of *Pinus radiate* and *Cupressus lusitanica* were significant (Davis et al., 2007). The relationship between the foliar canopy nitrogen concentrations and the underlying soil nitrogen cycling was significant in Northeastern US forest (Pardo et al., 2007). These studies were carried on the small sites in forest.

To best of our knowledge, literatures understanding the relation from remote estimation of foliar nutrient concentrations with the underlying geology on larger area of forest landscape are very limited. There are few comparable studies in literature which analyzed the spatial variation of biochemicals estimated from remote sensing. For example, Ollinger et al., (2002) examined the relationship between the spatial patterns of species level foliar nitrogen concentrations and the soil nitrogen availability and soil carbon: nitrogen ratios, and found that they all vary as function of stand age and disturbance history in white mountains of New Hampshire in North American landscape. Canopy foliar nitrogen was used as an indicator of ecosystem response to elevated atmospheric nitrogen deposition in Adirondack park, New York (McNeil et al., 2007). Empirical imaging spectroscopy approaches reflected the spatial variation of canopy chlorophyll concentrations along an elevation gradient in two coniferous species (Richardson & Berlyn, 2002), at leaf scale chlorophyll concentrations responses to changes in soil nutrient availability in temperate deciduous trees (Baltzer & Thomas, 2005), and canopy water and nitrogen maps provided insights on relationship between canopy chemistry variation with biological invasion in Hawaiian montane forest landscape (Asner & Vitousek, 2005).

3.11. Exploring the relationships of light availability with the leaf chlorophyll concentrations

Further the analysis was extended beyond the objective of this research, to investigate the variation in foliar biochemical concentrations of Norway spruce which is unexplained by geology classes. Along with soil nutrient availability, light availability also an another important factor which determines the pigment concentrations in forest landscape (Lee et al., 2000). The solar radiation strongly penetrates in canopies under steep angles; leaves in sun have higher chlorophyll concentrations (Andrew et al., 2002). Solar radiation is strongly influenced by terrain parameters such as slope, aspect and elevation (Kumar & Skidmore, 2000). Solar radiation was calculated for four weeks before acquisition date of sensor in the interval of 3 days using model developed within Geographical information system package (Arc GIS 10); the input to the model is a Digital Elevation Model (DEM). This model derives the incoming solar radiation for every pixel in a floating point type from DEM raster surface and has the units of watt hours

per square meter (WH/m²). For more description about this model, readers can refer to this page (ESRI, 2011). Here, DEM of 1meter spatial resolution was used to compute the incoming solar radiation. The histogram of the incoming solar radiation from the derived map reveal that they are ranging from 77980.55 to 465660.40 WH/m² (not shown here) (Figure 3-16). The simple random sampling was applied on the overlay of derived solar radiation and leaf chlorophyll a+b concentrations map. For a large population size, 385 samples were considered representative at 95% confidence level with ±5% allowable error (Rea & Parker, 2005) (refer section 2.9). The samples were extracted from both solar radiation and leaf chlorophyll concentrations map in the same corresponding pixels (Figure 3-16) for correlation analysis.

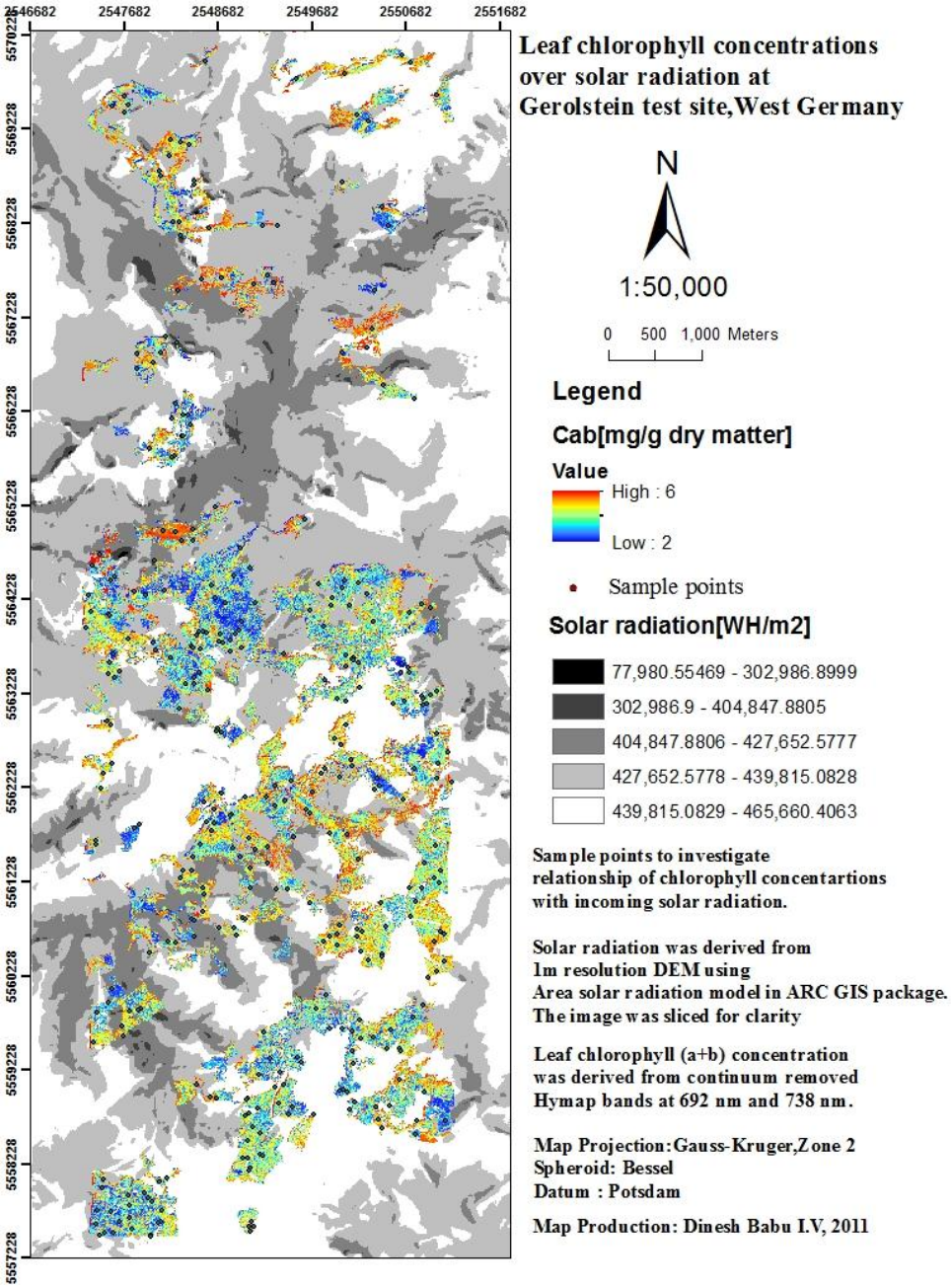


Figure 3-16 Map of leaf chlorophyll concentrations over incoming solar radiation

Similarly, the same random samples were used to investigate relationship between the leaf chlorophyll concentrations with elevation and slope (elevation and slope map not shown here). Refer Table 3-18 for the results of correlation between chlorophyll concentrations with solar radiation, elevation and slope.

Table 3-18 Correlation coefficient (r) of leaf chlorophyll concentrations with solar radiation, elevation and slope

Variables	Solar radiation	Elevation	Slope
Leaf chlorophyll a+b concentrations	r=0.026 (P value 0.60)	r =-0.041 (P value 0.427)	r= 0.009 (P value 0.86)

The results indicate that light availability, elevation and slope do not have significant relationship with the leaf chlorophyll a+b concentrations in this landscape. The reason is that Norway spruce (*Picea abies* L. *Karst*.) is a shade-tolerant species which can tolerate in the low light levels (Metslaid et al., 2005). This result confirms that low chlorophyll concentrations were found only due to low soil nutrient availability in this landscape. This result further strengthens the understanding of this forest ecosystem. Similarly, Baltzer & Thomas, (2005) examined the light and soil nutrient availability in temperate deciduous trees forest of Great Lakes–St. Lawrence region of eastern Canada and found that nutrient availability significantly influences the pigment concentrations rather than light availability. Elevation effects on foliar biochemistry were not significant for Douglas fir species in the Victoria watershed forest, Canada (Goodenough et al., 2010).

3.12. Summary of research answers

- **Research question 1:**

What are spectral bands significantly predicted the chlorophyll a+b (*Cab*) and carotenoids (*Cars*) in SMLR, PLSR and BRT regression models from laboratory spectra and airborne hyperspectral spectra?

Table 3-19 shows the wavebands significantly predicted the chlorophyll a+b and carotenoids in SMLR, PLSR and BRT models at the leaf scale.

- **SMLR**

The SMLR model selected wavebands in the blue regions (472nm), green peak (549nm) and red edge of the reflectance (761nm) for the prediction of *Cab*. In the case of *Cars* estimation, the wavebands in the blue regions (472nm) and green peak (549nm) were selected by SMLR at the leaf scale from laboratory spectra. At the canopy scale, the SMLR model selected the wavebands in the 692nm (close to major absorption wavelengths) and 738nm (red edge) from CR HyMap spectra for the prediction of *Cab*.

- **PLSR**

In the PLSR model, the wavebands in the blue regions (472nm), green peak (533-549nm), major absorption wavelengths (650-670nm) and red edge of the reflectance (731nm) significantly predicted the chlorophyll a+b at the leaf scale. The wavebands in blue regions (472nm), green peak (533-549nm) and the major absorption wavelengths (670-680 nm) significantly predicted the carotenoids at the leaf scale. At canopy scale, the wavebands in the 661-677nm and in the red edge of reflectance (738-761nm) from CR HyMap spectra contributed for the prediction of *Cab*.

- **BRT**

The BRT model selected the wavebands along the red edge (738nm), 549nm and 563nm (green peak of reflectance) for the estimation of *Cab* and *Cars* at the leaf scale from laboratory spectra. At canopy scale, the BRT model was not developed due to the limited number of samples.

- **Comparisons in the selected wavebands at leaf and canopy scales**

In general, the wavebands in the green peak (549nm) and located along the red edge of the reflectance (738-761nm) were selected by all models to predict chlorophyll a+b and carotenoids (Table 3-19 below). The BRT model alone didn't select the blue spectral regions for estimating chlorophyll a+b and carotenoids. Specifically, the PLSR model alone was based on the wavebands in the major absorption wavelengths (650-670nm). The PLSR model combined the information from the spectral regions (blue, both side of the green peak, major absorption wavelengths (red spectrum) and the red edge position) to predict chlorophyll a+b. The PLSR model was found to be more reliable based on the selected wavebands. Except the BRT model, SMLR and PLSR didn't select the spectral bands in the red edge position to determine the carotenoids concentration.

Table 3-19 The significant wavebands to predict chlorophyll a+b and carotenoids at leaf scale.

Biochemical	SMLR	PLSR	BRT
Chlorophyll a+b	472, 549, 761 nm	472, 487, 503-563, 656-671, 731nm	738, 549, 563 nm
Carotenoids	472, 549 nm	472, 533-549, 670-680 nm	738, 549, 563 nm

At the canopy scale, both SMLR and PLSR models were based on the wavebands in the major absorption features (670-690nm) and along the red edge of the reflectance. The results of carotenoids predictions at the canopy scale were poor.

- **Research question 2:**

Which regression method estimates foliar *Cab* and *Cars* most accurately from laboratory spectra and airborne hyperspectral spectra?

The PLSR regression method was found to predict chlorophyll a+b and carotenoids most accurately (low AIC value) than SMLR and BRT at the leaf scale. However, all the models predicted chlorophyll a+b and carotenoids with an accuracy of more than 85% at the leaf scale. Even the simple linear model (SMLR) predicts with the least error than the complex model (BRT) at the leaf scale. This research failed to compare models performance at the canopy scale.

This result failed to reject null hypothesis; H_0: BRT does not estimate foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) more accurately than PLSR and SMLR from laboratory spectra.

- **Research question 3:**

Are foliar chlorophyll a+b (*Cab*) and carotenoids (*Cars*) concentrations significantly different over the underlying soil substrate types?

The leaf chlorophyll a+b concentrations (LCC) of Norway spruce were significantly different over the underlying soil substrate (represented by geological classes) at 95% confidence level in this forest landscape. The LCC in the soil substrate (low soil fertility) was significantly lower than from the other soil substrate units at 95% confidence level. This result rejected null hypothesis; H_0: Foliar chlorophyll a+b (*Cab*) do not vary significantly over the underlying soil substrate types at the 95% confidence level.

3.13. Limitations of this study

- The limitation in number of samples at the canopy scale.
- Non availability of spatial information regarding soil substrate in the study site.

4. CONCLUSION

The PLSR model generally predicts the foliar biochemicals content most accurate at the leaf scale than SMLR and BRT. But, the simple SMLR model also predicts the foliar biochemical content in reasonable accuracy and it is easy for interpretation. Linear modelling is reasonable in the prediction of foliar biochemical content at the leaf scale using hyperspectral remote sensing. However, the multiple controlling factors on reflectance at the canopy scale allow nonlinear relationships between pigments and reflectance. The estimation of pigments chlorophyll and carotenoids separately in the overlapping absorption features at the canopy scale is still challenge. The results of carotenoids prediction at the canopy scale are poor. Further studies may consider on estimating *Cars: Cab* ratio in forest canopies. The limitation of this research in the number of samples failed for comprehensive model evaluation at the canopy scale. This creates opportunities for further studies to compare linear models with BRT, neural network and support vector machines at the canopy scale.

The SMLR model is used to derive the map of leaf chlorophyll a+b concentrations (LCC) in Norway spruce forest. The derived map of LCC reflects distinct spatial patterns due to the variation in soil nutrient availability. Low soil nutrient availability results in the low foliar chlorophyll a+b concentrations. Therefore, information concerning spatial variations of pigments can be valuable indicator of soil nutrient availability. This highlights that an understanding of the spatial patterns of leaf chlorophyll concentrations in relation to the underlying soil nutrients is a valuable input for the forest management. This allows mapping and managing the low foliar chlorophyll forest stands by possible fertilization. Soil nutrient availability greatly influences the foliar biochemical concentration rather than light availability, slope and elevation. Further studies may focus on understanding the spatial variation on the other nutrients such as nitrogen and lignin in canopies. This research finding suggests further studies to develop the maps of foliar biochemical content as a tool for understanding forest ecosystem response to various environmental factors to discover ecological principles. This research result may be very interesting for further studies to unravel the spatial and temporal variation of foliar biochemicals across a wide range of ecosystem. This research can be added claim among few examples exist in literature reflects the application and increasing value of imaging spectroscopy for understanding the forest ecosystem.

LIST OF REFERENCES

Andrew, D. R., Duigan, S. P., & P.Berlyn, G. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist*(153), 185-194.

Asner, G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance - the SAIL model. *Remote Sensing of Environment, 64*, 234-253.

Asner, G. P., & 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.

Asner, G. P., & Martin, R. E. (2009). Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. [Article]. *Frontiers in Ecology and the Environment, 7*(5), 269-276.

Asner, G. P., & Vitousek, P. M. (2005). Remote analysis of biological invasion and biogeochemical change. *Proceedings of the National Academy of Sciences of the United States of America, 102*(12), 4383-4386.

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment, 93*(1-2), 53-67.

Atzberger, C., Martine Guerif, Baret, F., & Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture, 73*(2), 165-173.

Baltzer, J. L., & Thomas, S. C. (2005). Leaf optical responses to light and soil nutrient availability in temperate deciduous trees. *American Journal Botany, 92*(2), 214-223.

Barry, K. M., Newnham, G. J., & Stone, C. (2009). Estimation of chlorophyll content in Eucalyptus globulus foliage with the leaf reflectance model PROSPECT. *Agricultural and Forest Meteorology, 149*(6-7), 1209-1213.

Bartley, G. E., & Scolnik, P. A. (1995). Plant Carotenoids: Pigments for Photoprotection,Visual attraction,and human health. *The Plant cell*(7), 1027-1038.

Bastien, P., Vinzi, V. E., & Tenenhaus, M. (2005). PLS generalised linear regression. *Computational Statistics & Data Analysis, 48*(1), 17-46.

Bauer, G., Schulze, E. D., & Mund.M. (1996). Nutrient contents and concentrations in relation to growth of Picea abies and Fagus sylvatica along a European transect. *Tree Physiology, 17*, 777-786.

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

Blackburn, G. A. (1998a). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing, 19*(4), 657 - 675.

Blackburn, G. A. (1998b). Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. [Article]. *Remote Sensing of Environment, 66*(3), 273-285.

Blockeel, H., & Struyf, J. (2002). Efficient Algorithms for Decision Tree cross validation. *Journal of Machine Learning Research, 3*, 621-650.

Boyle, J. R., & Powers, R. F. (2000). Forest soils and ecosystem sustainability. *Forest Ecology and Management, 138*(1-3), 1-1.

Bring, J. (1994). How to Standardize Regression Coefficients. *The American Statistician, 48*(3), 209-213.

Britton, G., S, L.-J., & Pfander.H. (2008). Natural functions, *Carotenoids* (Vol. 4, pp. 202-208). Basel, Switzerland: Springer Science,.

Burger, J. A. (2009). Management effects on growth, production and sustainability of managed forest ecosystems: Past trends and future directions. [Review]. *Forest Ecology and Management, 258*(10), 2335-2346.

Carter, G. A., & Knapp, A. K. (2001). Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *American Journal of Botany, 88*(4), 677-684.

Cawley, G. C., & Talbot, N. L. C. (2003). Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognition, 36*(11), 2585-2592.

Cawley, G. C., & Talbot, N. L. C. (2004). Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Network, 17*(10), 1467-1475.

Chan, J. C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment, 112*(6), 2999-3011.

Chen, L., Huang, J. F., Wang, F. M., & Tang, Y. L. (2007). Comparison between back propagation neural network and regression models for the estimation of pigment content in rice leaves and panicles using hyperspectral data. *International Journal of Remote Sensing, 28*(16), 3457-3478.

Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation, 9*(4), 414-424.

Clark, R. N. (1999). Spectroscopy of rocks and minerals,and principles of spectroscopy. In A. N. Rencz (Ed.), *Manual of Remote sensing* (Vol. 3, pp. 3-58). Newyork: John Wiley and Sons.

Clark, W. A. V., & Hosking, P. L. (1986). *Statistical methods for geographers*. New York.: Wiley & Sons.

Clemens, G., Fiedler, S., Cong, N. D., Van Dung, N., Schuler, U., & Stahr, K. (2010). Soil fertility affected by land use history, relief position, and parent material under a tropical climate in NW-Vietnam. *CATENA, 81*(2), 87-96.

Coops, N. C., marie-Louise smith, Mary E. Martin, & Scott V.Ollinger. (2003b). Prediction of Eucalypt Foliage Nitrogen Content From Satellite-Derived Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing, 41*(6), 1338-1346.

Coops, N. C., Stone, C., Culvenor, D. S., Chisholm, L. A., & Merton, R. N. (2003a). Chlorophyll content in eucalypt vegetation at the leaf and canopy scales as derived from high resolution spectral data. *Tree Physiology, 23*(1), 23-31.

Cote, B., Fyles, J. W., & Djalilvand, H. (2002). Increasing N and P resorption efficiency and proficiency in northern deciduous hardwoods with decreasing foliar N and P concentrations. *Annals of forest science, 59*(3), 275-281.

Covelo, F., Rodríguez, A., & Gallardo, A. (2008). Spatial pattern and scale of leaf N and P resorption efficiency and proficiency in a *Quercus robur* population. *Plant and Soil, 311*(1), 109-119.

Craine, J. M., Elmore, A. J., Aidar, M. P. M., Bustamante, M., Dawson, T. E., Hobbie, E. A., et al. (2009). Global patterns of foliar nitrogen isotopes and their relationships with climate, mycorrhizal fungi, foliar nutrient concentrations, and nitrogen availability. *New Phytologist, 183*(4), 980-992.

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

Curran, P. J., Dungan, J. L., & Peterson, D. L. (2001). Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: Testing the Kokaly and Clark methodologies. *Remote Sensing of Environment, 76*(3), 349-359.

Curran, P. J., Kupiec, J. A., & Smith, G. M. (1997). Remote sensing the biochemical composition of a slash pine canopy. *IEEE Transactions on Geoscience and Remote Sensing, 35*(2), 415-420.

Curran, P. J., Robert Windham, W., & Gholz, H. L. (1995). Exploring the relationship between reflectance red edge and Chlorophyll Concentration in Slash Pine leaves. *Tree Physiology, 15*(3), 203-206.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing, 63*(4), 409-426.

Dash, J., Curran, P. J., & Foody, G. M. (2008). Remote sensing of terrestrial chlorophyll content [Book section]. In A. P. Cracknell, C. A. Varotsos, V. F. Krapivin, J. Dash, P. J. Curran & G. M. Foody (Eds.), *Global Climatology and Ecodynamics* (pp. 77-105). Berlin: Springer Praxis Books.

Davies, K. (2004). *Plant Pigments and their Manipulation: Annual Plant Reviews* (Vol. 14): Blackwell Publishing Ltd,Oxford,UK.

Davis, M. R., Coker, G., Parfitt, R. L., Simcock, R., Clinton, P. W., Garrett, L. G., et al. (2007). Relationships between soil and foliar nutrients in young densely planted mini-plots of Pinus radiata and Cupressus lusitanica. *Forest Ecology and Management, 240*(1-3), 122-130.

Dawson, T. P., & Curran, P. J. (1998). Technical note A new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing, 19*(11), 2133 - 2139.

Death, G. (2007). Boosted Trees for Ecological Modeling and Prediction. *Ecology, 88*(1), 243-251.

Demetriades-Shah, T. H., Steven, M. D., & Clark, J. A. (1990). High resolution derivative spectra in remote sensing. *Remote Sensing of Environment, 33*(1), 55-64.

Demmig-Adams, B., Gilmore, A. M., & Adams, W. A. (1996). In Vivo functions of carotenoids in higher plants. *The Federation of American societies for Experimental Biology, 10*, 403-412.

Dougherty, C. (2007). Properties of the Regression Coefficients and Hypothesis Testing. In C. Dougherty (Ed.), *Introduction to Econometrics* (Third Edition ed.). London: Oxford University Press.

Duchemin, B. (1999). NOAA/AVHRR Bidirectional Reflectance: Modeling and Application for the Monitoring of a Temperate Forest. *Remote Sensing of Environment, 67*(1), 51-67.

Durbha, S. S., King, R. L., & Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment, 107*(1-2), 348-361.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jacknife. *Annuals of statistics, 7*, 1-26.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap,the jackknife, and cross-validation. *The American Statistician, 37*(1), 36-48.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*, 802-813.

ESRI. (2011). Area solar radiation. Online Tutorial of Arc GIS Package; http://webhelp.esri.com/arcgiSDEsktop/9.3/index.cfm?TopicName=Area_Solar_Radiation.

Fang, H., Liang, S., & Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. *Remote Sensing of Environment, 85*(3), 257-270.

Fox, J. (1997). *Applied regression analysis, linear models and related methods*. London. : Sage.

Freedman, D. A., & Peters, S. C. (1984). Bootstrapping a Regression Equation: Some Empirical Results. *Journal of the American Statistical Association, 76*(385), 97-106.

Friedman, J. H. (2001). Greedy function approximation:A gradient boosting machine. *The Annuals of statistics, 29*(5), 1189-1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics and Data Analysis, 38*(4), 367-378.

Ganesh, S. (2010). Multivariate Linear Regression. In P. Penelope, B. Eva & M. Barry (Eds.), *International Encyclopedia of Education* (pp. 324-331). Oxford: Elsevier.

Gastelluetchegorry, J. P., Zagolski, F., Mougin, E., Marty, G., & Giordano, G. (1995). An assessment of canopy chemistry with AVIRIS - a case study in the landes forest, south-west France [Article]. *International Journal of Remote Sensing, 16*(3), 487-501.

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.

Geladi, P., & Dabakk, E. (2009). Computational Methods and Chemometrics in Near Infrared Spectroscopy. In L. John (Ed.), *Encyclopedia of Spectroscopy and Spectrometry* (pp. 386-391). Oxford: Academic Press.

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology, 160*(3), 271-282.

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letter, 32*(8), L08403.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. *Photochemistry and Photobiology, 75*(3), 272-281.

Goodenough, D. G., Niemann, K. O., Quinn, G. S., Gross, A., & Lang, K. (2010, 25-30 July 2010). *Environmental controls on forest chemistry: evaluating and refining foliar chemistry from imaging spectroscopy.* Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International.

Goodenough, D. G., Niemann, K. O., Quinn, G. S., & Liu, J. (2009, 26-28 Aug. 2009). *Mapping spatio-temporal variation in Douglas-fir (Pseudotsuga menziesii) foliar biochemistry.* Paper presented at the First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS '09., Grenoble.

Grossman, Y. L., Ustin, S. L., Jacquemoud, S., Sanderson, E. W., Schmuck, G., & Verdebout, J. (1996). Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of Environment, 56*(3), 182-193.

Guang-sheng, C., De-hul, Z., & Fu-sheng, C. (2004). Concentrations of foliar and surface soil in nutrientsPinus spp. plantations in relation to species and stand age in Zhanggutai sandy land, northeast China. *Journal of Forestry Research, 15*(1), 11-18.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment, 81*(2-3), 416-426.

Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment, 86*(4), 542-553.

Hattenschwiler, S. (2005). Effects of Tree Species Diversity on Litter Quality and Decomposition. In M. Scherer-Lorenzen, C. Körner & E.-D. Schulze (Eds.), *Forest Diversity and Function* (Vol. 176, pp. 149-164): Springer Berlin Heidelberg.

Helmi Zullhaidi, M. S., & Nasrulhapiza, H. (2009). Hyperspectral Imagery for Mapping Disease Infection in Oil Palm Plantation Using Vegetation Indices and Red Edge Techniques. *American journal of Applied Sciences, 6*(6), 1031-1035.

Hobbie, S., & Gough, L. (2002). Foliar and soil nutrients in tundra on glacial landscapes of contrasting ages in northern Alaska. *Oecologia, 131*(3), 453-462.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., & Foley, W. J. (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sensing of Environment, 93*(1-2), 18-29.

Huber, S., Kneubühler, M., Psomas, A., Itten, K., & Zimmermann, N. E. (2008). Estimating foliar biochemistry from hyperspectral data in mixed forest canopy. *Forest Ecology and Management, 256*(3), 491-501.

Im, J., & Jensen, J. R. (2008). Hyperspectral Remote Sensing of Vegetation. *Geography Compass, 2*(6), 1943-1961.

Im, J., R.Jensen, J., Coleman, M., & Nelson, E. (2009). Hyperspectral remote sensing analysis of short rotation woody crops grown with controlled nutrient and irrigation treatments. *Geocarto International, 24*(4), 293-312.

Imanishi, J., Nakayama, A., Suzuki, Y., Imanishi, A., Ueda, N., Morimoto, Y., et al. (2010). Nondestructive determination of leaf chlorophyll content in two flowering cherries using reflectance and absorptance spectra. *Landscape and Ecological Engineering, 6*(2), 219-234.

Ismail, R., & Mutanga, O. (2010). A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation, 12*(Supplement 1), S45-S51.

Jacquemoud, S., Verdebout, J., Schmuck, G., & Andreoli, G. (1995). Investigation of Leaf Biochemistry by statistics. *Remote Sensing of Environment, 54*, 180-188.

Jensen, J. R. (2007). *Remote sensing of the environment : an earth resource perspective* (Second edition ed.). Upper Saddle River: Prentice Hall.

Johnson, L. F., Hlavka, C. A., & Peterson, D. L. (1994). Multivariate analysis of AVIRIS data for canopy biochemical estimation along the oregon transect. *Remote Sensing of Environment, 47*(2), 216-230.

Keiner, L. E., & Yan, X.-H. (1998). A Neural Network Model for Estimating Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery. *Remote Sensing of Environment, 66*(2), 153-165.

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., & Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment, 113*(Supplement 1), S78-S91.

Kokaly, R. F., & Clark, R. N. (1999). Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sensing of Environment, 67*(3), 267-287.

Kumar, L., & Skidmore, A. K. (2000). Radiation-Vegeatation Relationshipe in a Eucalyptus Forest. *Photogrammetric Engineering & Remote Sensing, 66*(2), 193-204.

LaCapra, V. C., Melack, J. M., Gastil, M., & Valeriano, D. (1996). Remote sensing of foliar chemistry of inundated rice with imaging spectrometry. *Remote Sensing of Environment, 55*(1), 50-58.

Lawerence, R., & Labus, M. (2003). Early Detection of Douglas-Fir Beetle Infestation with Subcanopy Resolution Hyperspectral Imagery. *Western Journal of Applied Forestry, 18*, 202-206.

Lawerence, R. L., Wood, S. D., & Sheley, L. R. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment, 100*(3), 356-362.

Lawrence, R. (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment, 90*(3), 331-336.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand:an analysis using boosted regression trees. *Marine Ecology Progress series, 321*, 267-281.

Lee, D. W., Oberbauer, S. F., Johnson, P., Krishnapilay, B., Mansor, M., Mohamad, H., et al. (2000). Effects of irradiance and spectral quality on leaf structure and function in seedlings of two Southeast Asian Hopea (Dipterocarpaceae) species. *American Journal Botany, 87*(4), 447-455.

Li, B., Morris, J., & Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 64*(1), 79-89.

Lopez-Serrano, F. R., de las Heras, J., Gonzalez-Ochoa, A. I., & García-Morote, F. A. (2005). Effects of silvicultural treatments and seasonal patterns on foliar nutrients in young post-fire Pinus halepensis forest stands. *Forest Ecology and Management, 210*(1-3), 321-336.

Lucas, N. S., & Curran, P. J. (1999). Forest ecosystem simulation modelling: the role of remote sensing. *Progress in Physical Geography, 23*(3), 391-423.

Magill, A. H., Aber, J. D., Currie, W. S., Nadelhoffer, K. J., Martin, M. E., McDowell, W. H., et al. (2004). Ecosystem response to 15 years of chronic nitrogen additions at the Harvard Forest LTER, Massachusetts, USA. *Forest Ecology and Management, 196*(1), 7-28.

Majeke, B., van Aardt, J. A. N., & Cho, M. A. (2008). Imaging spectroscopy of foliar biochemistry in forestry environments. [Article]. *Southern Forests, 70*(3), 275-285.

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.

McNeil, B. E., Read, J. M., & Driscoll, C. T. (2007). Foliar Nitrogen Responses to Elevated Atmospheric Nitrogen Deposition in Nine Temperate Forest Canopy Species. *Environmental Science & Technology, 41*(15), 5191-5197.

McNeil, B. E., Read, J. M., Sullivan, T. J., McDonnell, T. C., Fernandez, I. J., & Driscoll, C. T. (2008). The Spatial Pattern of Nitrogen Cycling in the Adirondack Park, New York. *Ecological Applications, 18*(2), 438-452.

Metslaid, M., Llisson, T., Vincente, M., Nikinmaa, E., & Jogiste, K. (2005). Growth of advance regeneration of Norway spruce after clear-cutting. *Tree Physiology, 25*, 793-801.

Miller, D. E., & Watmough, S. A. (2009). Soil acidification and foliar nutrient status of Ontario's deciduous forest in 1986 and 2005. *Environmental Pollution, 157*(2), 664-672.

Moore, D. S., & McCabe, G. P. (2006). *Introduction to the practice of statistics* (Fifth edition ed.). New York: W.H Freeman and Company.

Moorthy, I., Miller, J. R., & Noland, T. L. (2008). Estimating chlorophyll concentration in conifer needles with hyperspectral data: An assessment at the needle and canopy level. *Remote Sensing of Environment, 112*(6), 2824-2838.

Moulton, L. H., & Zeger, S. L. (1991). Bootstrapping generalized linear models. *Computational Statistics & Data Analysis, 11*(1), 53-63.

Mutanga, O., & Skidmore, A. K. (2003). *Continuum-removed absorption features estimate tropical savanna grass quality in situ*. Paper presented at the 3rd EARSEL workshop on Imaging spectroscopy,Herrsching,13-16 May 2003.

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. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing, 62*(1), 34-42.

Mutanga, O., van Aardt, J., & Kumar, L. (2009). Imaging spectroscopy (hyperspectral remote sensing) in southern Africa: an overview. *South African Journal of Science, 105*, 193-198.

Neff, J., Reynolds, R., Sanford, R., Fernandez, D., & Lamothe, P. (2006). Controls of Bedrock Geochemistry on Soil and Plant Nutrients in Southeastern Utah. *Ecosystems, 9*(6), 879-893.

Nguyen, H. T., & Lee, B.-W. (2006). Assessment of rice leaf growth and nitrogen status by hyperspectral canopy reflectance and partial least square regression. *European Journal of Agronomy, 24*(4), 349-356.

Niemann, K. O., Goodenough, D. G., & Bhogal, A. S. (2002). Remote sensing of relative moisture status in old growth Douglas fir. *International Journal of Remote Sensing, 23*(2), 395 - 400.

NOAA. (2011). Remote Sensing: The Logistics Accessed on 25/1/11 from http://www.csc.noaa.gov/products/sccoasts/html/rsdetail.htm.

O'Neill, A. L., Kupiec, J. A., & Curran, P. J. (2002). Biochemical and reflectance variation throughout a Sitka spruce canopy. *Remote Sensing of Environment, 80*(1), 134-142.

Ollinger, S., Smith, M. L., Martin, M. E., Hallett, R. A., Goodale, C. L., & Aber, J. D. (2002). Regional Variation in Foliar Chemistry and N Cycling among Forests of Diverse History and Composition. *Ecology, 83*(2), 339-355.

Omrad, D. P., Lesser, V. M., Olszyk, D. M., & Tingrey, D. T. (1999). Elevated Temperature and carbon Dioxide Affect chlorophylls and Carotenoids in Douglas -Fir seedlings. *International Journal of Plant Science, 160*(3), 529-534.

Pardo, L. H., McNulty, S. G., Boggs, J. L., & Duke, S. (2007). Regional patterns in foliar 15N across a gradient of nitrogen deposition in the northeastern US. *Environmental Pollution, 149*(3), 293-302.

Penuelas, J., Baret, F., & Fillellia, A. (1995). Semi-empirical indices to assess Carotenoids/chlorophyll *a* ratio from leaf spectral reflectance. *Photosynthetica, 31*(2), 221-230.

Rea, L. M., & Parker, R. A. (2005). *Designing and conducting survey research : a comprehensive guide*. San Francisco: Jossey-Bass.

Richardson, A. D., & Berlyn, G. P. (2002). Spectral reflectance and photosynthetic properties of Betula papyrifera (Betulaceae) leaves along an elevational gradient on Mt. Mansfield, Vermont, USA. *American Journal of Botany., 89*(1), 88-94.

Richardson, A. J., Everitt, J. H., & Gausman, H. W. (1983). Radiometric estimation of biomass and nitrogen content of Alicia grass. *Remote Sensing of Environment, 13*(2), 179-184.

Ruiliang, P., Peng, G., Biging, G. S., & Larrieu, M. R. (2003). Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Transactions on Geoscience and Remote Sensing, 41*(4), 916-921.

Run-he, S., Da-fang, Z., & Zheng, N. (2003). *Effects of spectral transformations in statistical modeling of leaf biochemical concentrations*. Paper presented at the IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003.

Sampson, P. H., Zarco-Tejada, P. J., Mohammed, G. H., Miller, J. R., & Noland, T. L. (2003). Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods. [Article]. *Forest Science, 49*(3), 381-391.

Sari, M., Sonmez, N. K., & Kurklu, A. (2005). Determination of seasonal variations in solar energy utilization by the leaves of Washington navel orange trees (*Citrus sinensis* L. Osbeck). *International Journal of Remote Sensing, 26*(15), 3295 - 3307.

Sariyildiz, T., & Anderson, J. M. (2003). Interactions between litter quality, decomposition and soil fertility: a laboratory study. *Soil Biology and Biochemistry, 35*(3), 391-399.

Sariyildiz, T., & Anderson, J. M. (2005). Variation in the chemical composition of green leaves and leaf litters from three deciduous tree species growing on different soil types. *Forest Ecology and Management, 210*(1-3), 303-319.

Schlerf, M. (2006). *Determination of structural and chemical forest attributes using hyperspectral remote sensing data - Case studies in Norway spruce forests, http://ubt.opus.hbz-nrw.de/volltexte/2006/369/*. Unpublished Ph.D Dissertation, Trier University.

Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment, 100*(3), 281-294.

Schlerf, M., Atzberger, C., Hill, J., Buddenbaum, H., Werner, W., & Schuler, G. (2010). Retrieval of chlorophyll and nitrogen in Norway spruce (Picea abies L. Karst.) using imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation, 12*(1), 17-26.

Schmidt, K. S., & Skidmore, A. K. (2001). Exploring spectral discrimination of grass species in African rangelands. *International Journal of Remote Sensing, 22*(17), 3421 - 3434.

Schonlau, M. (2005). Boosted regression (boosting) : An introductory tutorial and a stata plugin. *The stata Journal, 5*(3), 330-354.

Sherrod, P. H. (2010). DTREG- Predictive Modeling software. from www.dtreg.com on November 2010

Skidmore, A. K., Ferwerda, J. G., Mutanga, O., Van Wieren, S. E., Peel, M., Grant, R. C., et al. (2010). Forage quality of savannas -- Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. *Remote Sensing of Environment, 114*(1), 64-72.

Smith, M. L., Martin, M. E., Plourde, L., & Ollinger, S. V. (2003). Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between an airborne (AVIRIS) and a spaceborne (Hyperion) sensor. *IEEE Transactions on Geoscience and Remote Sensing, 41*(6), 1332-1337.

Solberg, S., Rindal, T. K., & Ogner, G. (1998). Pigment composition in Norway spruce needles suffering from different types of nutrient deficiency. *Trees - Structure and Function, 12*(5), 289-292.

Suat, S., & Dervis, T. (2007). Bootstrap and Jacknife Resampling Algorithms for Estimation of Regression Parameters. *Journal of Applied Quantitative Methods, 2*(2), 188-199.

Tedeschi, L. O. (2006). Assessment of the adequacy of mathematical models. *Agricultural Systems, 89*, 225-247.

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., Legg, C., & De Dieu, M. J. (2004). Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sensing of Environment, 90*(1), 23-43.

Townsend, P. A., Foster, J. R., Chastain, R. A., Jr., & Currie, W. S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Transactions on Geoscience and Remote Sensing, 41*(6), 1347-1354.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., et al. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment, 113*(Supplement 1), S67-S77.

Van der Meer, F. D., & De Jong, S. M. (2001). *Imaging spectrometry : basic principles and prospective applications*. Dordrecht Kluwer Academic.

Wang, G. G., & Klinka, K. (1997). White spruce foliar nutrient concentrations in relation to tree growth and soil nutrient amounts. *Forest Ecology and Management, 98*(1), 89-99.

Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock-recruitment relationships. *Fisheries Research, 77*(2), 220-225.

Wessman, C. A. (1992). Imaging spectrometry for remote sensing of ecosystem processes. *Advances in Space Research, 12*(7), 361-368.

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

Xiaobo, Z., Jiewen, Z., Hanpin, M., Jiyong, S., Xiaopin, Y., & Yanxiao, L. (2010). Genetic Algorithm Interval Partial Least Squares Regression Combined Successive Projections Algorithm for Variable Selection in Near-Infrared Quantitative Analysis of Pigment in Cucumber Leaves. *Applied Spectroscopy, 64*(7), 786-794.

Yoder, B. J., & Pettigrew-Crosby, R. E. (1995). Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. *Remote Sensing of Environment, 53*(3), 199-211.

Zarco-Tejada, P. J., Miller, J. R., Harron, J., Hu, B., Noland, T. L., Goel, N., et al. (2004). Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies. *Remote Sensing of Environment, 89*(2), 189-199.

Zas, R., & Serrada, R. (2003). Foliar nutrient status and nutritional relationships of young Pinus radiata D. Don plantations in northwest Spain. *Forest Ecology and Management, 174*(1-3), 167-176.

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*(7), 3234-3247.

Zhang, Y., Chen, J. M., & Thomas, S. C. (2007). Retrieving seasonal variation in chlorophyll content of overstory and understory sugar maple leaves from leaf-level hyperspectral data. *Candian Journal of Remote sensing, 33*(5).