Estimating nitrogen in mixed forests from HyMap data using band-depth analysis and branch-and-bound algorithm

Huber, Silvia; Zimmermann, Niklaus E; Kneubühler, Mathias; Itten, Klaus I

Abstract: Spatial information of nitrogen concentration (Nc) is of great interest because of its role in photosynthesis, ecosystem productivity and thus influences global cycling of carbon and oxygen. Imaging spectroscopy offers a means to assess this compound. Nc was estimated in mixed forests in Switzerland from airborne HyMap data using band-depth analysis. Instead of stepwise regression, an exhaustive search algorithm has been applied to select significant wavebands in order to build relationships between transformed reflectance and fieldmeasured Nc. This study confirms that partitioning data into vegetation types yielded in higher R2. R2 was largest for the homogeneous coniferous sample. A preclassification of the HyMap images is therefore recommended. The tested branch-and-bound algorithm performed well in selecting important known nitrogen absorption wavebands. A comparison with other subset selection methods is planned.

Published Version

Originally published at:
Estimating Nitrogen in Mixed Forests from HyMap Data using Band-Depth Analysis and Branch-and-Bound Algorithm

Huber Silvia¹, Niklaus E. Zimmermann², Mathias Kneubühler¹, Achilleas Psomas² and Klaus Itten¹

¹Remote Sensing Laboratories, Dept. of Geography, University of Zurich, Switzerland
²Swiss Federal Institute WSL, Dept. of Landscape Research, Switzerland

Abstract: Spatial information of nitrogen concentration (Nc) is of great interest because of its role in photosynthesis, ecosystem productivity and thus influences global cycling of carbon and oxygen. Imaging spectroscopy offers a means to assess this compound. Nc was estimated in mixed forests in Switzerland from airborne HyMap data using band-depth analysis. Instead of stepwise regression, an exhaustive search algorithm has been applied to select significant wavebands in order to build relationships between transformed reflectance and field-measured Nc. This study confirms that partitioning data into vegetation types yielded in higher R². R² was largest for the homogeneous coniferous sample. A pre-classification of the HyMap images is therefore recommended. The tested branch-and-bound algorithm performed well in selecting important known nitrogen absorption wavebands. A comparison with other subset selection methods is planned.

1. Introduction

Leaf Nc is of great interest because of its role in photosynthesis, ecosystem productivity and thus influences global cycling of carbon and oxygen. Previous studies have reported that canopy Nc showed a strong linear correlation with soil carbon to nitrogen (C:N) ratios and consequently affects nitrification, lignin decomposition and soil respiration rates (Berg and Meentemeyer, 2002; Medlyn et al., 2005) For that reason, ecosystem models can be substantially improved with spatial information of canopy Nc (Martin and Aber, 1997; Pan et al., 2004). Therefore, it is important to have robust and reliable means to retrieve this spatial information of Nc over large forested areas.

A number of studies have found that the calculation of band depth from dried and ground leaf material allows to develop calibration equations to estimate various biochemicals with high accuracies (Kokaly and Clark, 1999; Curran et al., 2001). Among different spectral transformation methods, continuum-removal analysis provided the most promising results (Run-he et al., 2003). The extension from controlled laboratory conditions to field level has been successfully demonstrated in grasslands for different biochemicals (Mutanga et al., 2004; Mutanga et al., 2005). Instead of relating Nc to dried leaf spectra, field-measured spectral data were used. First attempts in forested areas using airborne imaging spectrometer data for calibration have been conducted in eucalypt stands (Huang et al., 2004). However, in naturally grown mixed forest this method has not been evaluated yet. Stepwise regression has been most often used to locate wavebands in the foliar spectra. Because inconsistencies in this waveband selection method have been revealed (Grossman et al., 1996), an exhaustive search algorithm was tested, which has not been applied to reflectance data so far.

2. Methods

Three study areas in the Swiss Plateau region covered by mixed forest (Vordenwald 7°53’ E, 47°16’N, Küttingen 8°02’ E, 47°25’N, Bettlachstock 7°25’E, 47°13’N) were selected for sampling, to cover a wide range of species composition and foliar chemistry. In all areas the canopy was composed of a more or less even mixture of needleleaf and broadleaf crowns, mainly of European beech, English oak, northern red oak, European ash, sycamore maple, European alder, silver fir, Norway spruce and Scots pine.

In order to determine Nc in the laboratory, foliar material was sampled in late July to coincide with the peak of the growing season. Leaf samples were excised by a tree climber and stored in cool environment. To obtain a statistically representative sample, each sample consisted of several leaves from three different upper sunlit canopy branches. As a consequence the first three needle-years were pooled for the coniferous samples. In the laboratory, the samples were dried and ground before a high-temperature, dry combustion method was applied to determine the N concentration. In total, 122 samples were analyzed, whereof 50 were recorded as coniferous samples. To pinpoint the sampled tree crown later in the HyMap images, the stem position of each sampled tree was measured with a Trimble GeoXT GPS receiver. The positional accuracy was improved by applying a post processing differential correction on the recorded data in the Pathfinder Office software. The average horizontal positional error among all trees and study areas was 2.8 m, while the average radius of a deciduous and a coniferous tree crown were found to be 4.8 m and 2.9 m, respectively.

On 29 July 2004, imaging spectrometer data were registered for all study areas using Hyperspectral Mapping Imaging Spectrometer (HyMap). The HyMap instrument is flown at an altitude of 3000 m asl and measures upwelling radiance in 126 wavebands, from 0.45 to 2.48 μm, with a
spectral resolution of 15–20 nm and a spatial resolution of 5 m. HyMap at-sensor radiance data were transformed to apparent surface reflectance using the ATCOR4 software. The orthorectification based on the parametric geocoding procedure PARGE.

Pixels of each sampled tree were identified in the HyMap images. For each tree one pixel was extracted. The continuum removal analysis was applied at five pre-selected wavelength ranges according to known nitrogen absorption features (Tab. 1). The continuum is simply an estimate of the other absorptions present in the spectrum, not including the one of interest. Straight-line segments were used to approximate the continuum lines (Clark and Roush, 1984). The continuum-removed spectra are calculated by dividing the original reflectance values by the corresponding values of the continuum line (Kokaly and Clark, 1999). To minimize external influences the continuum-removed reflectance spectra were normalized. The band depth normalized to the center (BNC) is calculated by dividing the band depth of each channel by the band depth at the band center (Eq. 1):

\[
BNC = \frac{I - \left( \frac{R}{R_c} \right)}{I - \left( \frac{R}{R_c} \right)_c},
\]

where \( R \) is the reflectance of the sample at the waveband of interest, \( R_c \) is the reflectance of the continuum line at the waveband of interest, \( R_R \) is the reflectance of the sample at the absorption feature center and \( R_c \) is the reflectance of the continuum line at the absorption feature center. The band center is the minimum of the continuum-removed absorption feature (Kokaly and Clark, 1999). With multiple linear regression analysis relationships between BNC values, representing the explanatory variables, and field-measured \( N_e \) were investigated. In order to see whether a robust calibration equation can be developed among different plant functional types, data from all three study areas were combined in one sample \((n=122)\). A second and third sample contained only the coniferous \((n=50)\) and deciduous data \((n=72)\) of all sites, respectively. In the five pre-selected wavelength ranges all possible subsets of wavebands were exhaustively searched through for each sample using an efficient branch-and-bound algorithm (Miller, 2002). The basis of branch-and-bound algorithms is a ranking function. As a result we retrieved a number of wavelengths correlated with the dependent variable and a linear equation combining the values of the independent data set at these wavelengths with coefficients established by the regression. In order to avoid overfitting, an equation for predicting \( N_e \) was chosen according to the size of the analyzed sample, the Bayesian Information Criterion (BIC) and the adjusted \( R^2 \), which adjusts for the degrees of freedom in the model. The statistical methods were implemented by \( R \), a free software environment for statistical computing and graphics (Ihaka and Gentleman, 1996). The effectiveness of the models was tested with a 6-fold cross-validation method, since there is insufficient data to split it into a training and a validation group (Hastie et al., 2001).

3. Results

Foliar \( N_e \) determined in the laboratory varied more across deciduous than coniferous species. Deciduous foliar \( N_e \) (percent by dry weight) ranged from 1.80 % to 3.44 % with a mean and standard deviation of 2.28 % and 0.30 %, respectively. Foliar \( N_e \) of conifers ranged from 0.93 % to 1.55 % with a mean of 1.21 % and a standard deviation of 0.14 %. A lower \( N_e \) is characteristic for conifers compared to deciduous species.

Table 2 shows the results of multiple linear regression between HyMap BNC values and foliar \( N_e \). The selected wavebands varied among the three tested samples. A six-term relationship was developed on the BNC values of the combined sample. Three out of six explanatory variables occurred within ±12 nm of a known \( N \) absorption waveband. One selected waveband lies in the red edge range. The ±12 nm range was defined by Curran et al. (2001) to indicate causal chemical absorption. A seven-term regression equation was derived from the homogeneous deciduous sample. Only one of the selected wavebands was within the ±12 nm range. Interestingly, the lowest \( R^2 \) was obtained from this sample, even though a complex model was selected and the smaller data range compared to the combined sample. Four of the five selected wavebands in the regression developed on the homogeneous coniferous sample were either directly attributable to the \( N \) absorption features or lied in the red edge range. It has been reported that the red edge slope is sensitive to a variation in foliar chemistry (Yoder and Pettigrew-Crosby, 1995; Mutanga et al., 2005). The highest \( R^2 \) and the lowest standard error of

<table>
<thead>
<tr>
<th>Sample</th>
<th>Wavelength selected [nm]</th>
<th>( R^2 )</th>
<th>SEC (_T)</th>
<th>SEP (_{11})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>711, 944, 1504, 1544, 1972, 2188</td>
<td>0.52</td>
<td>0.41</td>
<td>0.42</td>
</tr>
<tr>
<td>Deciduous/</td>
<td>695, 928, 1329, 2028, 2047, 2154, 2240</td>
<td>0.38</td>
<td>0.25</td>
<td>0.26</td>
</tr>
<tr>
<td>Coniferous</td>
<td>726, 741, 912, 1023, 2083</td>
<td>0.62</td>
<td>0.09</td>
<td>0.10</td>
</tr>
</tbody>
</table>

* Within 12 nm, † within 20 nm and †† within 25 nm of known absorption features.

* Red edge slope, † Standard error of calibration; †† Standard error of prediction.

The standard error of prediction was nearly identical to the SEC and increasing with an increase of the data range.
4. Discussion and Outlook

R² was largest for the homogeneous coniferous sample and smallest for the homogeneous deciduous sample. A possible explanation for this might be the smaller data range of the coniferous sample compared to the other two samples, despite more complex canopy architecture. However, higher R² was achieved for the combined sample that had the largest data range than the homogeneous deciduous sample. The correlation might be influenced by the coniferous data since the two plant functional types were clearly distinguishable in the residuals map of the model fitted to the combined sample. Previous studies showed that partitioning data into vegetation types yielded in higher R² (Serrano et al., 2002; Mutanga et al., 2004). This study clearly confirms this finding for the homogeneous coniferous sample. However, the homogeneous deciduous sample has to be partitioned even among species to yield higher R² than the combined sample. Therefore, a pre-classification of the HyMap image is recommended.

Generally, the tested search algorithm retrieved satisfying results and a comparison to other subset selection methods is planned. Important known N absorption bands were selected. Again, best results were achieved with the homogeneous coniferous sample. The accurate geo-location of a single tree in a naturally grown mixed forest is problematic due to GPS accuracy and HyMap distortions. This can lead to wrong pixel extraction from the HyMap Images. With the attempt to sample purer canopy spectra another extraction method will be assessed which accounts for the tree crown radius and GPS horizontal errors. Further research should be done to investigate nonlinear relationships between transformed spectra and biochemical concentration.

Acknowledgment

The authors would like to thank the Swiss Federal Institute WSL, DLR for scheduling the fight campaign and the Swiss National Science Foundation. We also thank the many individuals who have helped with data collection and analysis.

References


Address:
Silvia Huber, Mathias Kneubühler and Klaus Itten, Winterrhurerstrasse 190, 8057 Zurich, Switzerland. (Email: shuber@geo.unizh.ch; kneub@geo.unizh.ch; itten@geo.unizh.ch)
Niklaus E. Zimmermann and Achilleas Psomas, Zuercherstrasse 111, 8903 Birmensdorf, Switzerland. (niklaus.zimmermann@wsl.ch; achilleas.psomas@wsl.ch)