

# Estimating Nitrogen in Mixed Forests from HyMap Data using Band-Depth Analysis and Branch-and-Bound Algorithm

Huber Silvia<sup>1</sup>, Niklaus E. Zimmermann<sup>2</sup>, Mathias Kneubühler<sup>1</sup>, Achilleas Psomas<sup>2</sup> and Klaus Itten<sup>1</sup>

<sup>1</sup>Remote Sensing Laboratories, Dept. of Geography, University of Zurich, Switzerland

<sup>2</sup>Swiss 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

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

Table 1. Pre-selected wavelength ranges and their associated absorption features for nitrogen (Curran, 1989).

Pre-selected wavelength ranges [nm]	Known nitrogen absorption features [nm]
558-756	640, 660
877-1069	910, 1020
1273-1673	1510
1794-2205	1940, 1980, 2060, 2130-2180
2223-2420	2240, 2300, 2350

the continuum lines (Clark and Roush, 1984). The continuum-removed spectra were 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) was calculated by dividing the band depth of each channel by the band depth at the band center (Eq. 1):

$$BNC = \frac{1 - (R/R_i)}{1 - (R_c/R_{ic})}, \quad (1)$$

where  $R$  is the reflectance of the sample at the waveband of interest,  $R_i$  is the reflectance of the continuum line at the waveband of interest,  $R_c$  is the reflectance of the sample at the absorption feature center and  $R_{ic}$  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 Nc 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 Nc 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, because there was insufficient data to split it into a training and a validation group (Hastie et al., 2001).

### 3. Results

Foliar Nc determined in the laboratory varied more across deciduous than coniferous species. Deciduous foliar Nc (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 Nc of conifers ranged from 0.93 % to 1.55 % with a mean of 1.21 % and a standard deviation of 0.14 %. A lower Nc is characteristic for conifers compared to deciduous species.

Table 2 shows the results of multiple linear regression between HyMap BNC values and foliar Nc. 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  $\pm 12$  nm of a known N absorption waveband. One selected waveband fell within the red edge range. The  $\pm 12$  nm range was defined by Curran et al. (2001) to indicate causal chemical absorption. A seven-term regression equation was derived for the homogeneous deciduous sample. Only one of the selected wavebands was within the  $\pm 12$  nm range. Despite of selecting a complex model and the smaller data range compared to the combined sample, the lowest  $R^2$  was obtained investigating this 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 fell within 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 calibration (SEC) were

Table 2. Linear regression results between Nc and BNC.

Sample	n	Wavelength selected [nm]	$R^2$	SEC $_{\dagger}$	SEP $_{\dagger\dagger}$
Combined Deciduous/ Coniferous	122	711 <sup>d</sup> , 944, 1504 <sup>a</sup> , 1544, 1972 <sup>a</sup> , 2188 <sup>a</sup>	0.52	0.41	0.42
Homogeneous Deciduous	72	695, 928 <sup>b</sup> , 1329, 2028, 2047 <sup>b</sup> , 2154 <sup>c</sup> , 2240 <sup>a</sup>	0.38	0.25	0.26
Homogeneous Coniferous	50	726 <sup>d</sup> , 741 <sup>d</sup> , 912 <sup>a</sup> , 1023 <sup>a</sup> , 2083 <sup>c</sup>	0.62	0.09	0.10

<sup>a</sup> Within 12 nm, <sup>b</sup> within 20 nm and <sup>c</sup> within 25 nm of known absorption features.

<sup>d</sup> Red edge slope.  $\dagger$  Standard error of calibration;  $\dagger\dagger$  Standard error of prediction.

obtained examining the coniferous sample. Generally, the standard error of prediction (SEP) was nearly identical to the SEC and increasing with an increase of the data range.

### 4. Discussion and Outlook

$R^2$  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 of more complex canopy architecture. Nevertheless,  $R^2$  was higher for the combined sample than the homogeneous deciduous sample, even though comprising the largest data range. 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  $R^2$  (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  $R^2$  than the combined sample. Therefore, a pre-classification of the HyMap image is