Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. *Karst.*) using imaging spectroscopy

Martin Schlerfa,*, Clement Atzbergerb, Joachim Hillc, Henning Buddenbaumc, Willy Wernerb, Gebhard Schülede

**A B S T R A C T**

The research evaluated the information content of spectral reflectance (laboratory and airborne data) for the estimation of needle chlorophyll (C_AB) and nitrogen (C_N) concentration in Norway spruce (*Picea abies* L. *Karst.*) needles. To identify reliable predictive models different types of spectral transformations were systematically compared regarding the accuracy of prediction. The results of the cross-validated analysis showed that C_AB can be well estimated from laboratory and canopy reflectance data. The best predictive model to estimate C_AB was achieved from laboratory spectra using continuum-removal transformed data (R^2cv = 0.83 and a relative RMSEcv of 8.1%, n = 78) and from hyperspectral HyMap data using band-depth normalised spectra (R^2cv = 0.90, relative RMSEcv = 2.8%, n = 13). Concerning the nitrogen concentration, we observed somewhat weaker relations, with however still acceptable accuracies (at canopy level: R^2cv = 0.57, relative RMSEcv = 4.6%). The wavebands selected in the regression models to estimate C_AB were typically located in the red edge region and near the green reflectance peak. For C_N, additional wavebands related to a known protein absorption feature at 2350 nm were selected. The portion of selected wavebands attributable to known absorption features strongly depends on the type of spectral transformation applied. A method called “water removal” (WR) produced for canopy spectra the largest percentage of wavebands directly or indirectly related to known absorption features. The derived chlorophyll and nitrogen maps may support the detection and the monitoring of environmental stressors and are also important inputs to many bio-geochemical process models.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Many bio-geochemical processes in terrestrial ecosystems are related to foliar biochemistry in leaves and needles, specifically to leaf pigmentation and leaf nitrogen concentrations (Melillo et al., 1982; Vitousek, 1982; Waring and Running, 1998). Developing methods to quantify pigment content by remote sensing could advance the understanding of photosynthetic processes and allow detection and monitoring of foliar conditions (Ustin et al., 2009).

Spectroscopic methods have been developed and applied with varying success to determine chlorophyll from leaf or needle spectra (Curran et al., 1992; Penuelas et al., 1994; Datt, 1998; Daughtry et al., 2000), from ground-measured canopy reflectance (Blackburn, 1998; Darvishzadeh et al., 2008), and from airborne measurements of canopy reflectance (Johnson et al., 1994; Jago et al., 1999; Sampson et al., 2003; Zarco-Tejada et al., 2004). However, at this time current methods have not delivered unambiguous results and operational methods for pigment retrieval (Ustin et al., 2009).

Most of the total foliar nitrogen (N) is contained in protein and less in chlorophyll. The absorption bands of proteins differ greatly from chlorophylls because of their dissimilar chemical structures (Weyer, 1985). Estimation of proteins from reflectance spectra can generally be achieved less reliably than pigments, as the absorption features associated with proteins are relatively weak (Jacquemoud et al., 1995). Therefore, based on the fact that the two variables are moderately correlated within and across ecosystems, nitrogen content has often been related roughly to and thus mapped from chlorophyll content (Ustin et al., 2004). However, ecosystems where N limitations are less important show a pronounced decoupling of leaf chlorophyll and total N (Kokaly et al., 2009). Hence, an alternative is to use protein absorption features in the mid-IR (1300–3000 nm) for nitrogen estimation (Kokaly, 2001). Methods that operate in the mid-IR may suffer...
from the strong influence of leaf water. A method that attempts to remove the confounding effect of leaf water could improve the reliability of nitrogen prediction models.

In the transition from the leaf to the canopy level, confounding effects are introduced, such as canopy architecture (e.g. leaf area index, leaf angle distribution, foliar clumping, and multiple canopy layers) and influence of soil and understory background (Asner, 1998). Airborne and spaceborne studies are affected by additional factors, such as BRDF effects, variable atmospheric conditions along with instrument characteristics (signal-to-noise ratio, spectral bandwidth, geometric field of view). Despite these well known problems, it has been attempted to derive biochemical variables at forest canopy scale from analysis of imaging spectrometer data using statistical techniques (Wessman et al., 1988; Gastellu-Etchegorry et al., 1995). However, wavebands selected through statistical approaches were often not consistent with the absorption features of the biochemicals within the leaves (Johnson et al., 1994; Curran et al., 2001) and inconsistent results were reported on randomised data sets (Grossman et al., 1996). Signal-to-noise ratio has been identified as one of the main limiting factors in reliably estimating foliar biochemicals from imaging spectrometer data (Peterson and Hubbard, 1992; Smith and Curran, 1996). The Australian HyMap sensor records data of good quality (Cocks et al., 1998) and to date it has seldom been employed for the estimation of foliar biochemicals in forests (Huang et al., 2004).

To enhance absorption features present in vegetation spectra, different types of spectral transformations have been proposed, such as first-derivative of reflectance (Demetriades-Shah et al., 1990; Cho et al., 2008), continuum-removed (CR) spectra, band-depth normalised (BD) spectra (Kokaly and Clark, 1999) and water removal (WR) through least square matching (Gao and Goetz, 1994, 1995). Within a spectral multiple linear regression framework, these transformations permit a better identification of absorption features related to the investigated biochemicals. Research that systematically addresses the different spectral transformations on their performance in estimating forest biochemicals from hyperspectral remote sensing data is rare.

The objectives of this research were (i) to establish predictive models between characteristic absorption features and concentrations of total chlorophyll and nitrogen at different spatial scales (laboratory and airborne), (ii) to test a method that attempts to remove the influence of leaf water from fresh vegetation spectra and to systematically compare the water removal method with other types of spectral transformations regarding the accuracy of the predictive models, (iii) to analyse the wavebands selected by stepwise multiple regression (SMR) for consistency in relation to known absorption features, and (iv) to derive maps of foliar chlorophyll concentration and nitrogen concentration.

The study used foliage sampled in stands of Norway spruce (Picea abies L. Karst.). This species was selected for two reasons: First, because it is the only native species of the genus Picea in Central Europe and there widespread throughout. Second, managed stands of Norway spruce occur over many different soil types present in the study area which makes it ideal to study remote sensing of foliar chemicals as a large soil nutrient variability might increase the range of biochemical concentrations to be found in the leaves.

2. Materials and methods

2.1. Study site

The Gerolstein test site (50°15′N, 6°40′E) is located in the central part of the Eifel Mountains, Germany. The mean annual precipitation is about 800–900 mm and the mean annual air temperature is 7 °C. Geologically the area is underlain by a variety of bedrock types; mainly lower Devonian quartzite-sandstone–schist formations, middle Devonian limestones, Triassic sandstones, and Tertiary and Quaternary basalts. In parts of the area the underlying bedrock is mixed with Pleistocene loess cover or periglacial layers. From the extreme diversity of bedrock types manifold soil types developed in the area. Managed stands of Norway spruce occur over all soil types with soils developed on bedrocks ranging from poor sandstone to rich limestone. Different soil types show a large variability in effective cation exchange capacity, base saturation and soil nitrogen content (Schwind and Schuler, 2001). The area is highly suitable to conduct the intended research due to its good road accessibility and nearby location to Trier University with necessary lab equipment.

2.2. Collection of foliar samples

At Gerolstein test site 13 stands of Norway spruce (Picea abies L. Karst.) were identified and sampled in August 2002. Stand ages range from 44 to 135 years. Foliar samples were obtained according to standardised procedures from the upper part of the crown of three randomly selected trees within each stand detailed in the guidelines of the German Environmental Specimen Bank (Paulus et al., 1996). Skilled forest rangers climbed the trunks up to the treetop and cut off three branches from the upper crown region of each tree (approximately 7th–12th whorl of the branches). Foliage from the sampled branches was removed separately by age class (first and third year) according to their position on the branch, bagged, labelled and placed in cold boxes in the field. The foliage samples were stored in refrigerators at +5 °C after transport to the nearby Trier University Geosciences Laboratory. In total, 78 samples were obtained (13 stands × 3 trees × 2 age classes). From these samples, sub-samples of approximately 20 g were selected randomly for reflectance measurements. After reflectance measurement the same needles were subjected to standard wet chemical analysis to obtain measures of the concentration of chlorophyll a and b, and nitrogen.

2.3. Chemical analysis

Chlorophyll a + b concentration of the foliage was determined after Lichtenthaler (1987) under cool and totally dark conditions. Approximately 0.5 g of fresh sample were weighed and ground by hand in acetone with the aid of quartz sand until no green color was left in the residual material. After the chlorophyll solution was decanted off the residue, acetone was added to a volume of 50 ml. The volume was centrifuged for 30 min, and then measured in a laboratory spectrophotometer (Shimadzu UV-160A; Shimadzu Corporation, Kyoto, Japan). Chlorophyll a was measured at its absorption maximum of 663 nm and chlorophyll b at 646 nm. Total nitrogen concentration of the foliage was determined through CHN elementary analysis (Wilson, 1990). The samples were combusted in an oxygen atmosphere at 980 °C and the formed gas was led over a copper granulate material where the nitrogen oxides were reduced to N2. Next, the gases were carried via a helium stream to a detector and compared to a known standard value. Results were reported in percent element by dry weight [100 × g g⁻¹ dry matter].

Among the different ways of scaling-up from leaf level to canopy level (Zarco-Tejada et al., 2001) a direct statistical relationships between ground-measured biochemical data and canopy-measured reflectance was established. A direct statistical relation was chosen because modelling the spectral contributions of leaf nitrogen by physical means is difficult—it was attempted in the original version of the PROSPECT leaf model (Jacquemoud and
Baret, 1990) but later abandoned due to inconsistencies in the retrieval of N via model inversion (Kokaly et al., 2009).

To link canopy reflectance to biochemical measurements the biochemical data that had been obtained on leaf level (first and third year needles) were averaged and aggregated to stand level by computing the stand mean of both age classes. Averaging the two needle age classes was justified by the fact that both concentrations are positively correlated. Correlations between the two age classes were positive on a stand level (e.g. for $C_N: r = 0.83, \sigma < 0.01$) indicating that forest growth conditions affect different needle age classes in a similar way. It is also more logic to assume that the canopy reflectance is determined by the combined effect of needle optical properties of both age classes. For the models developed on laboratory spectra, the samples of the two age classes were pooled together to obtain age class independent models.

2.4. Laboratory spectral measurements

Needle optical properties were measured in the laboratory from optically dense needle layers following a procedure described by Despan and Jacquemoud (2004) using an ASD Field-Spec-2 spectrometer (www.asdi.com). The target was illuminated by a 1000 W halogen light source at a zenith angle of 45° and a distance of approximately 60 cm. The optical head of the instrument (FOV: 25°) was positioned at nadir at a distance of 10 cm above the target so that the instrument’s field of view lay well within the target’s perimeter (Φ: 8 cm). Needles were arranged side-by-side (1 cm thickness) in a glass cuvette over a matt black background. Relative spectral radiances between 350 and 2500 nm were recorded at 1 nm steps with a spectral resolution of approximately 3–10 nm. Each sample measurement comprised 50 single measurements of the target and was repeated four times thereby rotating the target by approximately 70° to average out any possible directional effects. A reference panel with known reflectance properties made of polytetrafluoroethylene (Spectralon) was measured before and after the target to calculate the target reflectance. The five spectra recorded for each sample were averaged to produce a single spectrum. The resulting spectra were smoothed using a 31 nm wide moving Savitzky-Golay filter to reduce instrument noise (polynomial degree: 3) and resampled to the HyMap wavebands.

2.5. Hyperspectral image data

The hyperspectral HyMap sensor with 126 spectral channels was flown over the Gerolstein test site on 14 July 2003. It was assumed that the relative differences in foliar concentrations did not change over one year between the field measurements and the image acquisition, as both data sets were recorded at similar phenological stages. The HyMap data were recorded at 12:40 h local time at an average flying height of 3400 m above ground level. The resulting ground resolution was about 7 m with a full scene covering about 4 km × 13 km. The flight line was run in an almost N–S direction. The acquired scene was cloud free. Prior to analysis, 16 out of the 126 HyMap spectral channels were identified as “bad” bands and removed from the dataset. Radiometric pre-processing consisted of an across track illumination correction (first order polynomial) to remove view angle effects. A combined sensor calibration and atmospheric correction was applied to derive top-of-canopy (TOC) reflectances from the recorded digital numbers. For this purpose, the aerosol optical depth was derived from estimates of the horizontal visibility provided by three meteorological stations located in the region (mean horizontal visibility: 35 km). On the day of the overflight several reference targets with different surface characteristics were measured at the ground with an ASD Field-Spec-2 spectrometer. Ground reflectance measurements and aerosol characteristics together allowed converting the digital counts into TOC spectra (for details see Schlerf et al., 2005).

The image data were geometrically corrected to the local coordinate system (Gauss Krüger, Zone 2, Ellipsoid: Bessel 1841, Datum: Potsdam) using the parametric geocoding software PARGE (Schläpfer et al., 1998) and a DEM of the area (source: State Survey Office of Rheinland-Pfalz, Koblenz, Germany) (Schlerf, 2006). The root mean squared error (RMSE) of the geometric correction was 4.9 m in the x-direction and 5.6 m in the y-direction. With an image pixel size of 7 m, the geometric correction was accurate to within a pixel.

As many of the investigated forest stands in the study area were formed by just a relatively small number of image pixels, a growing region approach was applied using the image processing software ENVI (version 4.0) to determine the pixels representative for a given stand. At each field plot in the image, four pixels were defined as a “region of interest” (ROI). Growing of ROIs was permitted in four directions based on the standard deviation computed in HyMap band 29 (near-IR) resulting in ROIs of 8–31 image pixels. For further analysis, only the average signatures form the selected pixels were used.

2.6. General workflow

The general workflow of the research is outlined in Fig. 1. Using stepwise multiple linear regression (SMR) cross-validated regression models between spectral data at canopy/needle scale and needle biochemical concentrations of $C_{SB}$ and $C_N$ were established. Following approaches by Kokaly and Clark (1999) and Curran et al. (2001), five broad absorption features (Table 1) were extracted from the laboratory and airborne spectra. The reflectance data within these five broad spectral windows were subjected to four spectral transformations: first-difference (FD), continuum-removal (CR), band-depth normalisation (BD) and least square matching (WR). For comparison, we also analysed the original (un-transformed) data (RF). The five data sets were analysed sequentially with SMR with biochemical concentrations as dependent variables. The samples of the two age classes were pooled together to develop age class independent SMR models ($n = 78$). With the spectral inputs derived from the HyMap image, only the averages of the two needle age classes were used ($n = 13$).

2.7. Spectral transformations

2.7.1. First-difference (FD)

The first-derivative was approximated with first-difference, i.e., by calculating differences in reflectance between adjacent wavebands. Compared to reflectance spectra, derivative spectra are in principle less sensitive to variation in soil background, illumination, and surface albedo (Demetriades-Shah et al., 1990; Cho et al., 2008).

2.7.2. Continuum-removal (CR) and band-depth (BD) normalisation

The main purpose of continuum-removal (CR) and band-depth (BD) normalisation is the minimization of effects that extraneous factors may have on reflectance spectra to highlight shape differences of absorption features (Kokaly and Clark, 1999). CR is based on a continuum line approximated by linear segments. The continuum-removed reflectance at a certain wavelength was calculated by dividing the original reflectance value by the values of the continuum line at the corresponding wavelength. Then, the band depth was computed by subtracting the continuum-removed reflectance from 1. Finally, the band depth was normalised by dividing the band depth of each waveband by the band depth at the band centre.
2.7.3. Water removal (WR) for protein absorption features

Leaf water is known to obscure the weak protein absorption features. To remove the obscuring effects of leaf water, Gao and Goetz (1994, 1995) developed the least squares spectral matching technique. We used their technique (hereafter named water removal, WR) to enhance the information about proteins. WR was applied only to the three spectral features related to proteins (see Table 1).

The water removal technique assumes that a spectrum can be modelled using the absorption spectra of water and protein:

$$R_{\text{mod}}(\lambda) = (A + B\lambda)e^{-(C_Wk_W(\lambda)+C_PK(\lambda))}$$  \hspace{1cm} (1)

where $R_{\text{mod}}(\lambda)$ is the modelled spectrum, $k_W(\lambda)$ and $k_P(\lambda)$ are absorption coefficients of water and protein, respectively, $C_W$ is the concentration of water, $C_P$ is the concentration of protein, and the term $(A + B\lambda)$ represents the background level of the calculated spectrum, assumed to be linear over a relatively short wavelength interval (Gao and Goetz, 1994, 1995).

For a given protein absorption feature, the information relevant to leaf protein can be enhanced by eliminating the leaf water influence. The contribution of leaf water to the total fresh leaf reflectance is modelled by fitting the three unknowns ($A$, $B$ and $C_W$) in Eq. (1) while setting $C_P$ to zero. Then, the residual spectra between measured reflectance $R_{\text{mes}}(\lambda)$ and modelled reflectance $R_{\text{mod}}(\lambda)$ is computed which is called the water removed spectrum $WR(\lambda)$:

$$WR(\lambda) = \frac{R_{\text{mes}}(\lambda) - R_{\text{mod}}(\lambda)}{R_{\text{mes}}(\lambda)}$$  \hspace{1cm} (2)

$WR(\lambda)$ contains the information relevant to leaf protein (i.e., the term $C_PK$ in Eq. (1)). To parameterise the model, the absorption coefficients of liquid water ($k_W(\lambda)$) and protein ($k_P(\lambda)$) were taken from the LIBERTY model (Dawson et al., 1998). The optimisation of the three unknowns was implemented in Matlab (www.mathworks.com) using the Nelder-Mead simplex method.

### Table 1

<table>
<thead>
<tr>
<th>Denotation of absorption feature (nm)</th>
<th>Relation to chemical</th>
<th>Start (nm)</th>
<th>End (nm)</th>
<th>Causal absorption bands (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>670</td>
<td>Chlorophyll $a+b$</td>
<td>548</td>
<td>760</td>
<td>640, 660</td>
</tr>
<tr>
<td>1200</td>
<td>Lignin, water</td>
<td>1080</td>
<td>1270</td>
<td>1120, 1230</td>
</tr>
<tr>
<td>1800</td>
<td>Protein</td>
<td>1680</td>
<td>1820</td>
<td>1690, 1730</td>
</tr>
<tr>
<td>2100</td>
<td>Protein</td>
<td>2100</td>
<td>2200</td>
<td>2130, 2180</td>
</tr>
<tr>
<td>2300</td>
<td>Protein</td>
<td>2240</td>
<td>2370</td>
<td>2240, 2300, 2350</td>
</tr>
</tbody>
</table>

Fig. 1. The general workflow of the research.

2.8. Regression models

After transformation the five broad absorption features were combined into single data sets, z-transformed (band wise mean of 0 and standard deviation of 1), and subjected to regression analysis. Forward SMR was used to develop relationships between chemical concentration and the spectral transformations.

As the number of samples was limited, cross-validation was used to assess the accuracy of the models yielding statistically unbiased results (Huang et al., 2004). The results were assessed in terms of three cross-validated statistics: coefficient of determination ($R^2_{\text{cv}}$), root mean squared error ($\text{RMSE}_{\text{cv}}$) and normalised $\text{RMSE}_{\text{cv}}$ ($\text{nRMSE}_{\text{cv}}$). To avoid model overfitting (Curran, 1989), two control measures were implemented: (1) Cross-validated values of RMSE were plotted against the number of bands selected. The number of bands included into the model was controlled by the $p$-values for entry.
and removal. The RMSEcv value above the break in slope was used to define the maximum number of bands. (2) Control runs were done with biochemical data that had been brought into random order. When a control run for a particular model achieved values of \( R^2cv \) larger than zero, \( p \)-values for entry and removal were decreased until no correlation between spectral data and randomised biochemical data was observed any longer. After taking into account both control measures, models typically included between 2 and 7 wavebands for needle spectra \((n = 78)\) and between 1 and 4 wavebands for canopy spectra \((n = 13)\).

### 2.9. Interpretation of selected wavebands

To interpret the relation of the selected wavebands to foliar constituents a list of absorption bands related to leaf chlorophyll and leaf nitrogen was compiled from Curran (1989), Fourtette et al. (1996) and Lucas and Curran (1999), namely: 640 and 660 nm (chlorophyll), 1200 nm (water), 1690, 1730, 2130, 2180, 2240, 2300, and 2350 nm (protein) (Table 1). Following Curran et al. (2001) a selected waveband was considered as being directly related when it occurred within 30 nm around the central absorption band of the chemical of interest. A selected waveband was considered as being indirectly related when it occurred within 30 nm around the central absorption band of a chemical with which the chemical of interest was correlated.

### 3. Results and discussion

#### 3.1. Biochemical concentrations

Summary statistics (Table 2) indicate that chlorophyll \( a + b \) concentration \((C_{AB})\) has variability two times larger than nitrogen concentration \((C_N)\). Despite the efforts in selecting stands having a large variation in soil nutrient availability, variability of foliar nitrogen concentration in the study area was relatively small \((\text{coefficient of variation about 10\%})\). As expected, first year needles have lower \( C_{AB} \) than third year needles. \( C_N \) does not show a marked difference between first and third year needles. Linear correlation coefficients between \( C_{AB} \) and \( C_N \) in needle samples \((n = 78)\) indicate a significant but weak positive relation \((r = 0.37, \sigma < 0.01)\). On stand level \((n = 13)\) we found a stronger positive relation \((r = 0.78, \sigma < 0.01)\).

#### 3.2. Spectral transformations

#### 3.2.1. Laboratory spectra

With laboratory spectra cross-validated \( R^2 \) between measured and estimated concentrations for different spectral transformations were uniformly large \((>0.75)\) for \( C_{AB} \) and uniformly moderate \((about 0.50)\) for \( C_N \) (Table 3). The best predictive model to estimate \( C_{AB} \) at the laboratory level yielded an \( R^2cv \) of 0.81 and a relative RMSEcv of 8.7\% \((n = 78)\). The lower coefficient of determination for \( C_N \) as compared to \( C_{AB} \) can be attributed to the small dynamic range of this variable (Table 2). Relative errors \((\text{nRMSEcv})\) for the two biochemicals were between 7\% \((C_N)\) and 10\% \((C_{AB})\). This shows that under controlled laboratory conditions both biochemicals can be estimated accurately from fresh (optical thick) needle spectra. The various spectral transformations had only a limited impact probably because the laboratory setting reduced significantly the perturbing factors that affect the spectral signatures at canopy scale.

#### 3.2.2. Canopy spectra

With image spectra, differences between the spectral transformations were more pronounced (Table 4). Chlorophyll concentration \((C_{AB})\) was well estimated by all methods except the untransformed reflectance spectra \((RF)\). This demonstrates the positive effects of the selected transformation techniques to cope with the perturbing factors introduced at the canopy scale. The best predictive model to estimate \( C_{AB} \) was achieved by band-depth normalisation \((BD)\) with a cross-validated coefficient of determination \((R^2cv)\) of 0.90 and a relative RMSE of 2.8\%. Results obtained from the continuum-removed spectra \((CR)\) were also very accurate \((\text{relative RMSEcv}: 4\%)\), and used only two instead of three wavebands in the model.

The promising results that were obtained for chlorophyll may be explained by the fact that chlorophyll is a strong absorber of radiation and that the chlorophyll absorption feature is not masked by other components.
by leaf water. Using radiative transfer models it is easy to show that the pigment concentration has a large influence on canopy reflectance (Schlerf and Atzberger, 2006). The same factors operating conversely may explain the weaker relations between reflectance and the concentration of nitrogen ($R^2_{cv} = 0.57$, relative RMSEcv = 4.6%). In fact, protein as one of the main nitrogen bearing leaf constituents is only a weak absorber of radiation (Dawson et al., 1998; Jacquemoud et al., 1995). The already weak absorption feature is further masked by leaf water. Using radiative transfer models it is easy to show that the pigment concentration has a large influence on canopy reflectance (Schlerf and Atzberger, 2006).

The derived results confirm the findings of comparable airborne hyperspectral studies. For instance, using narrowband spectral indices, Zarco-Tejada et al. (2001) mapped the chlorophyll content in closed forest canopies with RMSEs in the range 3.0–5.5%. Chlorophyll $a + b$ concentration estimation of conifer needles using coupled leaf and canopy models was shown to be feasible with a RMSE of $5.3 \mu g cm^{-2}$, for a pigment range of 25.7–45.9 $\mu g cm^{-2}$ (Moorthy et al., 2008). Nonetheless lower accuracies were obtained using the red edge inflection point for predicting the chlorophyll $a$ concentration in conifer forests (relative RMSE: 24%) (Blackburn, 2002).

No statistically significant relationship could be established between leaf nitrogen and RF spectra, whereas transformed canopy spectra (FD, CR, BD and WR) yielded statistically significant models with cross-validated $R^2$ between 0.34 and 0.57. This highlights again the positive benefits of the investigated transformation techniques when dealing with canopy spectra. Nitrogen concentration was moderately well estimated by CR, FD and WR spectra using three wavebands with a relative RMSEcv below 5%. The results are comparable to accuracies obtained in other airborne hyperspectral studies. For instance, Huang et al. (2007) estimated nitrogen in Eucalypt foliage from HyMap data using PLSR with a relative RMSE of 4.8% (absolute RMSE of 0.69 mg g$^{-1}$ and mean nitrogen concentration of 14.35 mg g$^{-1}$). Over a wide range of forest ecosystems foliar N concentration was predicted to within 7–15% of the mean field measured values using PLSR (Martin et al., 2008). In savannah grass nitrogen concentrations was mapped with a relative RMSE of 8.3% of the mean observed nitrogen concentration using continuum-removal (Mutanga and Skidmore, 2004).

### 3.3. Selected wavebands

The analysis of the selected wavebands from canopy spectra revealed that only a small number are related directly to the central absorption wavelengths tabulated in Table 1. As the stands were heterogeneous with respect to structural attributes like leaf area index, crown coverage, etc., this result is understandable. In fact, at stand level, SMR has to select spectral wavebands that correct for differences in stand attributes to establish useful correlations with the (biochemical) variables of interest. The wavebands chosen by SMR to determine the leaf chlorophyll concentration (Table 4) although not attributable to main absorption wavelengths were often located in spectral regions that are known to be important to detect chlorophyll. For instance, the predictive model to estimate total chlorophyll concentration found with CR spectra was based on wavebands located along the red edge (738 nm) and along the edge of the green peak (585 nm). This was also the case with the regression models developed on laboratory spectra (not shown). The finding that maximum correlation with chlorophyll concentration occurs on the edge of absorption features is in full agreement with the results of other studies where differences in reflectance between healthy and stressed vegetation due to changes in pigment levels have been detected in the green peak and along the red edge (Gitelson and Merzlyak, 1996; Haboudane et al., 2002). For $C_N$, an additional waveband related to a known protein absorption features at 2350 nm was selected. This shows that the estimation of $C_N$ does not rely solely on the correlation between $C_{AB}$ and $C_N$. The WR transformation for $C_N$ yielded accuracies (both in $R^2_{cv}$ and in RMSEcv) comparable to the two best performing transformation techniques (FD and CR) using the same number of wavebands. From the three selected wavebands, however, two could be directly related to protein absorption bands, whereas FD and CR models gave only one out of three.

Wavebands close to protein features at 1730, 2180, and 2350 nm that have often been selected in the literature are also selected in this research. As in the present work, a combined selection of wavebands related to both chlorophyll and protein

### Table 4

Cross-validated results obtained at stand level from HyMap imagery ($n = 13$). The two biochemicals were estimated by SMR using 5 different spectral transformations. Units of RMSE for $C_{AB}$ [mg g$^{-1}$ dry matter] and for $C_N$ [% dry matter]. Bands (total–direct–indirect) indicate the total, directly or indirectly related number of bands selected by SMR.

<table>
<thead>
<tr>
<th>Spectral transformation</th>
<th>$C_{AB}$</th>
<th>$C_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>$R^2_{cv}$</td>
<td>0.46 $^*$</td>
</tr>
<tr>
<td></td>
<td>RMSEcv (nRMSEcv)</td>
<td>0.22 (6.9%)</td>
</tr>
<tr>
<td></td>
<td>Bands (total–direct–indirect)</td>
<td>3–0–0</td>
</tr>
<tr>
<td>FD</td>
<td>$R^2_{cv}$</td>
<td>0.79 $^*$</td>
</tr>
<tr>
<td></td>
<td>RMSEcv (nRMSEcv)</td>
<td>0.13 (4.0%)</td>
</tr>
<tr>
<td></td>
<td>Bands (total–direct–indirect)</td>
<td>3–0–0</td>
</tr>
<tr>
<td>CR</td>
<td>$R^2_{cv}$</td>
<td>0.80 $^*$</td>
</tr>
<tr>
<td></td>
<td>RMSEcv (nRMSEcv)</td>
<td>0.13 (4.0%)</td>
</tr>
<tr>
<td></td>
<td>Bands (total–direct–indirect)</td>
<td>2–0–0</td>
</tr>
<tr>
<td>BD</td>
<td>$R^2_{cv}$</td>
<td>0.90 $^*$</td>
</tr>
<tr>
<td></td>
<td>RMSEcv (nRMSEcv)</td>
<td>0.09 (2.8%)</td>
</tr>
<tr>
<td></td>
<td>Bands (total–direct–indirect)</td>
<td>3–0–0</td>
</tr>
<tr>
<td>WR</td>
<td>$R^2_{cv}$</td>
<td>n.a.</td>
</tr>
<tr>
<td></td>
<td>RMSEcv (nRMSEcv)</td>
<td>0.058 (4.9%)</td>
</tr>
<tr>
<td></td>
<td>Bands (total–direct–indirect)</td>
<td>3–2–0</td>
</tr>
</tbody>
</table>

$^*$ $p_{in} = 0.10$, $p_{out} = 0.15$.

$^*$ $p_{in} = 0.05$, $p_{out} = 0.10$. 
absorption features in a single model to estimate nitrogen concentration is also reported from other studies (Johnson et al., 1994; Curran et al., 2001; Serrano et al., 2002). When analysing canopy reflectance spectra, it is also clear that selected wavebands will have to respond to the investigated biochemical, but will have to counterbalance at the same time the perturbing effects introduced by forest structure and background reflectance (Atzberger, 2004).

3.4. Predictive equations and maps of foliar biochemicals

The models developed on CR data were among the most accurate spectral transformations for canopy spectra (Table 4). The continuum-removal is also very easy to implement using standard image processing software. It was hence decided to use the CR models for mapping purposes. For chlorophyll prediction the CR model used only 2 wavebands (Table 4):

\[
C_{AB} = -41.781 \cdot CR_{738} + 16.539 \cdot CR_{585} + 33.304
\]

where \(C_{AB}\) is the estimated chlorophyll \(a + b\) concentration and \(CR_{l}\) denotes the continuum-removed value of reflectance in the specified waveband (in nm). Similarly, for leaf nitrogen and CR data a 3-band model was obtained:

\[
C_{N} = -5.354 \cdot CR_{738} + 2.799 \cdot CR_{1157} - 8.983 \cdot CR_{2373} + 12.724
\]

where \(C_{N}\) is the estimated nitrogen concentration.

The negative regression coefficient of \(CR_{738}\) (Eqs. (3) and (4)) agrees with the general understanding that with increasing chlorophyll/nitrogen concentration a shift of the red edge inflection point to longer wavelengths occurs (i.e., a red shift). This is accompanied by reduced reflectance in the red edge region (e.g. at 738 nm) and ultimately by reduced continuum-removed reflectance. In a similar way, the negative sign of \(CR_{738}\) in the protein absorption feature (Eq. (4)) can be explained. Conversely, a positive relation between chlorophyll and continuum-removed reflectance is expected around the green peak, e.g. \(CR_{585}\). The positive coefficient at 1157 nm in the \(C_{N}\) model probably reflects the negative relation between leaf nitrogen and lignin concentrations. In fact, the main lignin absorption feature is located at 1120 nm (Table 1), and thus relatively close to the selected band at 1157 nm. The positive coefficient at 1157 nm indicates thus a lower lignin concentration which can be indicative of higher nitrogen concentration in our sample.

The scatter plots (Fig. 2) show linear relationships between measured and estimated values of \(C_{AB}\) and measured and estimated values of \(C_{N}\) at the laboratory and canopy scales. The cross-validated points in the scatter diagrams are close to the 1:1 line over the entire range of values indicating similar model performance over the full range of concentrations encountered in this study.

Predictive models developed using CR spectra (Eqs. (3) and (4)) were applied to the entire HyMap image to compute maps of \(C_{AB}\) and \(C_{N}\) (not shown) with all non-conifer pixels masked out.

After integrating the biochemical forest maps to a Forest-GIS they were compared to maps of geology and stand age. In Fig. 3, distinct patterns related to the underlying geology can be observed. At locations A (quaternary loams and colluvium), B (triasic sands), and C (Devonian clay loam) prevailing colors are blue-cyan (large \(C_{AB}\)), red-yellow (small \(C_{AB}\)) and cyan-yellow (intermediate \(C_{AB}\)), respectively. The variation in base saturation in

![Fig. 2. Estimated against measured leaf total chlorophyll concentration (\(C_{AB}\)) and estimated against measured leaf nitrogen concentration (\(C_{N}\)) using laboratory spectra (A) and image spectra (B). In A, dots and crosses represent first year and third year needles, respectively.](image-url)
the area (Schwind and Schüler, 2001) seems to have a clear effect on chlorophyll concentrations. Conversely, we observed a limited effect of stand development on leaf chlorophyll in the study area. A more detailed interpretation of the two maps was beyond the scope of the present study.

At the Gerolstein test site, observed minimum nitrogen concentration (stand mean) is 1.08% of dry matter and the observed maximum is 1.31%. According to a classification scheme for nitrogen proposed by Hüttl (1992), 12 out of 13 sampled conifer stands have very small nitrogen concentrations. Compared to the within ecosystem nitrogen variation reported from tall grass prairie (Hobbs et al., 1991) or the variation in foliar nitrogen reported from forests across North America which spans a range from 1 to 2.5% (Yin, 1993), the within ecosystem variation in Norway spruce stands is considerably smaller with a range of only 0.23% nitrogen. The large variation in base saturation present in the soils at the Gerolstein area apparently is not equalled by corresponding variations in leaf nitrogen concentrations.

4. Conclusions

This study systematically evaluated different types of spectral transformations regarding the accuracy of the estimation of chlorophyll ($C_a$) and nitrogen ($C_n$) concentrations in Norway spruce canopies and the consistency of selected wavebands in relation to known absorption features. The upscaling effect has been investigated by running the analysis on laboratory spectra and on HyMap airborne spectra.
In this research, predictive models between characteristic absorption features and concentrations of total chlorophyll and nitrogen at different spatial scales (laboratory and airborne) were established. The results indicate that variations in chlorophyll and nitrogen concentrations can be mapped from laboratory and airborne spectra with accuracies (cross-validated relative RMSE) between 5 and 10%. The obtained accuracy is similar to slightly better to what has been reported in the literature, with however a limited sample size.

We systematically compared different spectral transformations regarding the accuracy of the predictive models. Any of the investigated transformation techniques provided accuracies superior to the un-transformed data. The positive effects of the various transformations are more apparent in the case of the airborne data where many factors influence the spectral signature. Presumably, at least some of the confusing factors can be minimized by using appropriate spectroscopic techniques. The wavebands selected in the regression models to estimate C_{AB} were typically located in the red edge region and near the green reflectance peak. For C_{G}, additional wavebands related to a known protein absorption feature at 2350 nm were selected.

Compared to the other investigated transformation techniques, the water removal approach (WR) for mapping of protein features did not increase the accuracy of the estimations. With WR, however, a higher number of wavebands could be related to known absorption features. This increases confidence in looking on basic chemical-spectroscopic linkages and eases portability to other tree species. Albeit simple, the WR technique can be seen as a (strongly simplified) physically based radiative transfer model. Physically based canopy reflectance models for forests (e.g. Schlerf and Atzberger, 2006) are known for their potential to produce more accurate and consistent predictions of leaf composition and canopy structural variables because they are based in physics and use the full spectrum rather than individual bands. However, they usually require detailed inputs to work properly.

In future studies, a larger variation of nitrogen has to be considered by extending the study to other ecosystems, such as deciduous forest, vine yards, orchards or grasslands. However, when dealing with samples made up of different species one has to consider the existence of cross-species co-variances between vegetation structure and chemistry. This has been avoided in the present study by focussing solely on Norway spruce as the most widespread and typical conifer species in Europe.

The derived chlorophyll and nitrogen maps represent the status of the Norway spruce forests of the Gerolstein study site in 2003. Future studies may use these maps as a reference to detect and monitor environmental stressors and to the advance understanding of photosynthetic processes (Ustin et al., 2009). Maps of canopy chemistry are also important prerequisites to many bio-geochemical process models (Kokaly et al., 2009). Ongoing research will use these maps to initialize and parameterize these spatially distributed dynamic models (e.g. Schlerf et al., 2004).

Acknowledgments

This research was financially supported by the German Research Community (DFG, reference number HI 703/1-1) and the research fund of Trier University. Forest stands were carefully selected with support of Dr. W. Schwind, Gerolstein. The authors thank Samuel Bärisch, Eva Pohl, Lena Lieckfeld (Trier University), H.-J. Mack (Trippstadt) and three forest rangers for assistance during fieldwork at Gerolstein test site. Organisation of the flight campaign by Andreas Müller and Stefanie Holzwarth (German Aerospace Centre, Oberpfaffenhofen) is highly acknowledged.

References


