ELSEVIER

Investigation of Leaf Biochemistry by Statistics

S. Jacquemoud, J. Verdebout,† G. Schmuck,† G. Andreoli,†
and B. Hosgood†*

The biochemical concentration (total protein, cellulose, lignin, and starch) of 73 plant leaves has been related to their optical properties through statistical relationships. Both fresh and dry plant material, leaves and needles, were used in this study. Stepwise multiple regression analyses have been performed on reflectance, transmittance, and absorptance values (individual leaves) as well as on reflectance values of optically thick samples (stacked leaves + needles), on measured values and on transformations of them such as the first derivative or the logarithm of the reciprocal of the reflectance. They underscored good prediction performances for protein, cellulose, and lignin with high squared multiple correlation coefficients (r^2) values. Starch, whose concentration in the leaf was smaller compared to the other components, was estimated with less accuracy. As expected, dry material and optically thick samples provided respectively stronger correlations than fresh material and individual leaves.

INTRODUCTION

Estimating leaf biochemistry and leaf water status with remote sensing data is a challenge for the years to come. In the framework of the Global Change Program, it should provide interesting information about the functioning of terrestrial ecosystems by extending ecological models to different scales (Ustin et al., 1991; Peterson and Hubbard, 1992). The biochemical constituents of interest are lignin, proteins (nitrogen), cellulose, starch, chlorophyll, and water (Peterson, 1991). The major processes involved in terrestrial ecosystems like photosynthesis, primary production, or litter decomposition can be related to these constituents. For example, according to Wessman (1994), the lignin / nitrogen ratio, which drives the decomposition rates of forest litter, could be a key input to general circulation models because of its role in the carbon cycle. As leaves are the most important surfaces of a plant canopy, relating their optical properties to these constituents is a priority. Two different approaches may be considered: The first one rests on a theoretical basis which consists in developing a leaf scattering and absorption model involving biochemistry. For the moment, only chlorophyll content and water content have been explicitly included in leaf optical properties models. Conel et al. (1993) recently proposed a two-stream radiative transfer model to analyze the influence of protein, cellulose, lignin, and starch on leaf reflectance, but this model has not been validated. Except for this work, only statistical analyses have been performed to estimate leaf biochemical components. They represent the second approach.

Near-infrared reflectance spectroscopy (NIRS) is a method routinely used in forage quality analyses. The common practice in this field is to establish a regression between the reflectance of the dried powdered material and the biochemical content, using the results of the wet chemical analysis performed on a restricted number of samples. Once this calibration is determined, NIRS is then used to analyze the bulk of the samples with the advantage of being as accurate as the wet chemical techniques and, above all, much faster (Williams and Norris, 1987). With the purpose of transferring this procedure to remote sensing, multiple stepwise regressions between reflectance spectra and biochemical concentration have been first established for dry leaves and needles of nonagricultural plants (Card et al., 1988; Wessman et al., 1988; McLellan et al., 1991; Gastellu-Etchegorry et al., 1994). The application to remote sensing also implies a need to work on fresh material. Calibration equations derived from dry material cannot be applied directly to fresh leaves and new wavelengths have to be selected. The situation is then less favorable

---

*Laboratoire Associé de Météorologie Physique, Université Blaise Pascal / CNRS, Aubière, France

†Joint Research Centre / Institute for Remote Sensing Applications, Unit for Advanced Techniques, Ispra, Italy

Address correspondence to Stephane Jacquemoud, Univ. of California, Dept. of Land, Air, and Water Resources, Davis, CA 95616.

Received 21 November 1994; revised 5 July 1995.

REMOTE SENS. ENVIRON. 54:180–188 (1995)

655 Avenue of the Americas, New York, NY 10010

0034-4257 / 95 / $9.50
SSDI 0034-4257(95)00170-6

because the dominant effect of absorption by water largely masks the signatures of the biochemical components beyond 1.0 μm. Correlations on fresh material could however be found, but with different combinations of wavelengths (Curran et al., 1992; Grossman et al., 1994; NASA, 1994).

We intend in this article to investigate the estimation of protein, cellulose, lignin, and starch both on fresh and dry material, both on individual leaves and optically thick samples (stacked leaves + needles). Water and chlorophyllian pigments, which absorb most of the radiation, respectively, in the visible and mid-infrared domains, will be ignored in this study. Nevertheless, the experiment and the data set are presented as a whole. After a brief description of the stepwise multiple regression analysis method, results are discussed.

THE EXPERIMENT

Measurements from a laboratory experiment organized in the Joint Research Centre during the summer of 1993 were used (Hosgood et al., 1995). We have undertaken to build a data set associating visible / infrared spectra of vegetation elements (leaves, conifer needles, stems, etc.), both with physical measurements and biochemical analyses. To have a wide range of variation of leaf internal structure, pigmentation, water content, and biochemical components, plant species with different types of leaves have been collected. Many species were sampled twice (once in early summer, once in early autumn) in order to assess the variability in time of the biochemical composition. For particular species like maize, which is the main crop in the region, plants at several stages of the growing cycle were collected. In total, about 70 leaf samples representing 50 species of woody and herbaceous plants (four gymnosperms, nine monocotyledons, and 37 dicotyledons) were obtained from trees and crops in the area of the JRC (Table 1). In addition, various substances such as powdered starch or proteins and vegetation material such as stems or bark were also included in this data set.

A Perkin Elmer Lambda 19 Spectrophotometer equipped with an integrating sphere allowed the measurement of the hemispherical reflectance (R) and transmittance (T), the probe light beam of the instrument being incident on the upper face of the leaves with an angle of 8 degrees. The absorptance (A) was derived from R and T through the simple relationship: $A = 1 - (R + T)$. Moreover, the reflectance of optically thick samples (R_∞) was obtained with needles and by stacking leaves in order to magnify the radiometric signal and minimize the leaf to leaf variability. Spectra were scanned over the 400–2500 nm wavelength interval in 1-nm steps. The spectral resolution varied from 1 nm to 2 nm in the visible / near-infrared (400–1000 nm) and from 4 nm to 5 nm in the mid-infrared (1000–2500 nm). The calibration of the instrument was performed using Spectralon reflectance and wavelength calibration standards. For each sample, the optical properties of five representative fresh leaves were averaged to reduce the noise of the spectrophotometer and to smooth out the small but not negligible leaf to leaf variability. All the above procedure was repeated on dried leaves and needles to analyze the influence of water, which is known to obscure the biochemical information in the mid-infrared region. The wavelength interval was increased from 1 nm to 5 nm in order to reduce the noise, the number of data, and consequently the calculation time. In total, 421 narrow wavebands were used in the regression. The first derivative was calculated for each spectrum: noise in spectra was low enough to avoid the use of smoothing techniques. Table 4 summarizes the radiometric data acquired in the framework of this experiment.

Parallel to the radiometric acquisitions, many physical and biological measurements were performed on leaf samples. Leaf blade thickness was measured with a calliper rule (five measurements per leaf). We immediately measured the fresh weight of a 4.10 cm^2 disc taken on each leaf using a cork borer. Then, the disc was placed in a drying oven at 85°C for 48 hr and reweighed to determine the water content (WC = water mass over fresh mass), the equivalent water thickness or water depth (EWT = water mass per unit leaf area), and the specific leaf area (SLA = dry weight per unit leaf area). Special attention has been paid to the measurement of SLA because biochemical concentrations used in leaf optical properties models are generally expressed in weights per unit leaf area; in this way, we will be able in the future to compare outputs from models with outputs from statistical relationships. Remaining leaf samples were frozen for later biochemical analysis: the photometric determination of photosynthetic pigments (chlorophyll *a*, *b*, and total carotenoids) was performed with a UV-2001 PC Spectrophotometer in 100% acetone using the equations of Lichtenthaler (1987). With regard to the other biochemical constituents, about 250 g of fresh material was placed in a drying oven and was sent to two independent laboratories[1], #1 and #2, which were in charge of the wet chemical analyses of protein, cellulose, lignin, and starch. Protein was expressed as 6.25 times the Kjeldahl nitrogen concentration, and cellulose as the percentage of crude fiber using the Weend proximate analysis procedure (Barton, 1987). Percent lignin was determined by the Van Soest and Wine (1968) method, and starch by a classical polarimetric method (Ewerts method).

---

[1] Analyses were conducted by the "Centre de Recherches Agronomiques" of Libramont, Belgium (#1) and by "Europe Sols" in Toulouse, France (#2).

Table 1. Latin Names of Plant Leaves Used in This Study and Corresponding Biochemistry (g g^{-1}) as Measured by the Two Laboratories (#1 and #2)

	Laboratory #1				*Laboratory #2*			
	Protein	*Cellulose*	*Lignin*	*Starch*	*Protein*	*Cellulose*	*Lignin*	*Starch*
				Gymnosperms				
*Picea abies**	7.2	28.5	13.5	0.6	7.0	25.9	13.4	0.0
*Pinus contorta**	9.5	34.7	13.0	0.8	8.4	33.7	12.6	0.0
Pinus wallichiana	12.4	28.6	16.3	2.4	11.3	25.0	17.0	1.5
*Pseudotsuga menziesii**	8.7	26.9	16.6	0.0	8.5	23.5	13.2	0.0
				Monocotyledons				
*Bambusa acundinacea**	16.5	31.7	6.0	0.0	16.1	27.0	6.1	0.0
Chamaerops humilis	11.3	37.2	9.7	0.8	12.4	25.8	8.8	0.2
Iris germanica L.	15.7	27.6	3.5	0.0	15.3	24.0	4.1	0.0
Musa ensete	20.8	21.9	4.5	1.2	21.3	18.2	3.3	1.2
*Oryza sativa**	16.1	28.2	3.9	0.0	17.3	23.9	5.3	0.0
Phleum pratense L.	18.3	29.8	3.2	0.4	19.6	24.6	4.7	0.0
Phragmites communis	16.3	25.6	5.6	0.0	19.8	21.0	6.1	0.0
Sorghum halepense	23.7	30.0	3.6	0.4	24.2	24.9	3.5	0.0
*Zea mays L.**	19.3	26.8	3.1	0.1	19.1	20.2	2.9	1.9
				Dicotyledons				
*Acer pseudo-platanus L.**	17.6	20.0	9.5	0.6	18.5	17.1	8.7	0.8
Alnus glutinosa	22.2	15.8	14.1	1.6	22.0	12.5	10.0	1.7
*Armeniaca vulgaris**	15.5	13.5	13.9	3.3	19.8	10.1	4.6	4.4
Beta vulgaris L.	34.5	10.6	10.8	2.6	35.5	5.6	3.4	2.3
Betula alba L.	14.9	21.4	13.6	1.6	15.0	17.5	14.5	2.6
Brassica oleracea L.	33.2	16.4	6.0	0.0	35.5	11.7	2.6	0.0
*Castanea sativa**	12.2	27.2	11.9	2.9	12.4	21.4	13.2	4.4
*Corylus avellana L.**	16.7	18.8	14.2	2.7	13.6	14.7	11.0	1.2
Fagus silvatica L.	17.2	26.6	17.1	4.5	17.3	22.9	23.2	0.6
Ficus carica L.	19.7	14.1	9.6	0.8	20.8	10.9	0.0	4.4
*Fraxinus excelsior L.**	18.7	14.6	19.1	2.6	19.4	11.2	10.6	0.3
Hedera helix L.	14.3	23.1	13.6	0.0	15.1	19.2	13.1	0.0
Helianthus annuus L.	34.9	9.1	12.5	0.8	35.8	8.3	3.3	0.0
Juglans regia L.	20.1	15.3	16.2	3.3	21.9	13.2	6.9	0.5
Lactuca sativa	35.6	16.8	1.6	2.3	35.5	12.4	3.9	2.7
Laurus nobilis L.	11.8	26.8	16.8	2.0	10.5	22.6	18.4	6.4
Lycopersicum esculentum	26.6	12.9	7.2	0.8	28.1	9.6	3.1	1.1
*Medicago sativa L.**	34.7	12.8	4.8	5.0	32.4	6.1	3.1	1.5
*Morus alba L.**	21.7	13.5	5.8	1.9	23.2	10.6	3.2	0.9
Platanus acerifolia	3.4	39.1	34.9	3.3	3.6	28.8	4.0	1.6
Populus canadensis	17.5	16.5	10.6	1.8	18.5	13.8	10.0	0.0
Populus tremula L.	15.2	21.6	12.1	0.2	15.5	22.9	10.9	0.0
*Prunus laurocerasus**	8.6	18.0	24.4	5.6	8.3	15.6	12.5	1.7
Prunus serotina	17.9	12.7	6.3	0.0	18.5	10.7	6.0	0.0
*Quercus pubescens**	15.4	23.6	15.6	2.1	13.7	22.2	15.4	0.5
Quercus rubra	13.6	23.8	12.7	1.7	13.4	23.9	13.3	3.6
*Robinia pseudoacacia L.**	26.1	16.6	14.4	5.2	25.8	13.4	12.0	2.7
Salix alba L.	14.4	15.2	9.5	1.4	14.4	12.1	9.2	1.4
Salvia officinalis L.	19.2	20.7	13.2	2.4	21.6	15.9	12.5	2.7
Soja hispida	31.6	16.8	2.9	3.3	31.1	14.9	4.2	2.1
Solanum tuberosum L.	30.3	14.5	1.1	3.7	31.9	11.0	2.6	1.4
*Tilia platyphyllos**	18.9	17.3	7.6	0.6	20.0	14.2	6.9	0.7
Trifolium pratense L.	31.4	15.8	2.2	2.4	31.7	12.1	3.04	0.0
Ulmus glabra	16.6	17.5	9.3	0.8	16.9	14.5	0.1	0.0
Urtica dioica L.	26.3	12.9	10.3	2.5	27.9	12.0	3.6	0.0
*Vitis silvestris**	15.3	12.1	22.12	4.3	16.0	9.7	4.7	5.9
Vitis vinifera L.	19.1	10.7	27.5	4.1	19.7	10.5	5.4	2.8

[a] Stars indicate that several samples were available for the species: average values of concentrations are then provided.

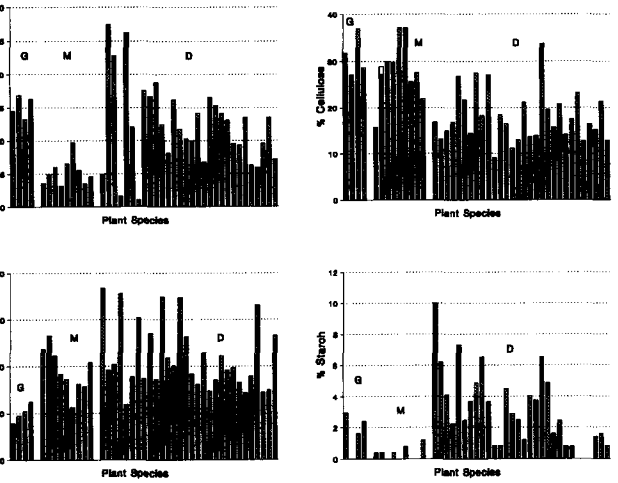


Figure 1. Distributions of lignin, cellulose, protein, and starch concentrations for gymnosperms (G), monocotyledons (M), and dicotyledons (D). The different values represented in one bar correspond to the various samples of the same plant species.

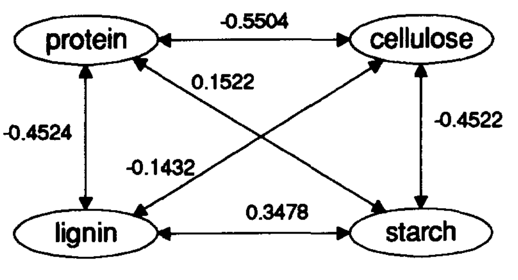


Figure 2. Correlation coefficient for relationships among chemicals expressed on a dry weight basis (g g^{-1}).

The concentrations expressed in g g^{-1} spanned a broad range of values (Table 1). A cross-laboratory comparison gave us an idea of the precision of the analyses: Protein and cellulose were quite consistent while lignin and starch differed significantly (Table 2). In the following, only results of #1 are presented. Figure 1 shows the distribution of chemistry data for the leaf samples: The lignin concentration of needles and dicotyledons which mainly correspond to woody plants respectively amounts to about 14.3% and 12.1% of the dry weight while that of monocotyledons is three times lower (4.3%). The distribution is rather different for cellulose: Its concentration in needles and monocotyledons respectively amounts to 29.5% and 28.3% while that in dicotyledons is 10% lower. Lignin and cellulose are structural components of cell walls: They constitute altogether around 30.8% of the dry weight of plant leaves and 43.8% of the dry weight of needles, but they are physiologically and biochemically inactive (Kramer and Kozlowski, 1979). In contrast, the concentration of protein which has essential roles in biochemical and physiological processes is twice as high as in plant leaves (19.4%) than in needles (8.5%). Finally, the starch concentration is very small except for dicotyledons where it equals 2.5% of the dry weight. The relationships between the chemical concentrations are shown in Figure 2: Protein, cellulose, and lignin are negatively correlated one to each other; starch is positively correlated to protein and lignin, negatively to cellulose.

THE ANALYSIS

For determining predictive spectral wavebands of chemical concentrations, stepwise multiple regression analyses which consist in relating the concentration C with the reflectance (or other spectral property) at 1, 2, . . . ,N wavelengths [$R(\lambda_1)$, $R(\lambda_2)$, . . . ,$R(\lambda_N)$] were performed:

$$C = a_0 + a_1 \cdot R(\lambda_1) + a_2 \cdot R(\lambda_2) + \cdots + a_N \cdot R(\lambda_N). \quad (1)$$

The regression model was constructed adding the independent variables (e.g., reflectances) one at a time. The first step was to choose the single variable which was the best statistical predictor; the second independent variable to be added to the regression equation was that which provided the best fit in conjunction with the first variable. Further variables were then added in this recursive fashion, adding at each step the optimum variable, given the other variables already in the equation. Goodness of fit was measured by two criteria: the squared multiple correlation coefficient r^2, which could be interpreted as the fraction of total sum of squares explained by the regression, and by the root mean square error rmse calculated as

Table 2. Summary Statistics for the Biochemical Concentrations of the Leaf Samples[a]

	Laboratory #1			*Laboratory #2*			*Comparison #1, #2*	
	Range	*Mean*	*Std*	*Range*	*Mean*	*Std*	*Bias*	*Std*
Protein	6.4–36.8	18.4	7.7	6.1–35.8	18.8	8.1	−0.45	1.29
Cellulose	9.1–37.2	21.0	7.0	2.1–37.5	17.5	6.8	3.57	2.18
Lignin	1.1–27.5	10.8	6.2	0.0–23.2	8.1	5.1	3.27	6.56
Starch	0.0–10.0	1.8	2.1	0.0–9.4	1.4	2.1	0.47	1.65

[a] Values are expressed in percentage of dry weight (g g^{-1}).

Table 3. Squared Multiple Correlation Coefficient r^2 Obtained for Protein and Reflectance (R) on 50 Regression Equations Established on Same-Sized Subsamples[a]

			r^2			
		Training	Training	Validation	Training	Validation
Fresh leaves	Number of Samples	63	42	21	21	42
	Mean	0.50	0.65	0.31	0.77	0.18
	Min		0.39	0.04	0.51	0.00
	Max		0.78	0.54	0.92	0.50
Dry leaves	Number of Samples	57	38	19	19	38
	Mean	0.81	0.82	0.66	0.90	0.48
	Min		0.65	0.14	0.75	0.06
	Max		0.91	0.87	0.97	0.78

[a] The number N of selected wavelengths equals 5.

$$\mathrm{rmse}=\sqrt{\frac{\sum(C-C^*)^2}{n-k}}, \tag{2}$$

where C and C^* are, respectively, the measured and estimated concentrations, n is the number of individuals, and k the number of regressors in the calibration equation including the constant. k is closely related to the size of the data set: As discussed by Hruschka (1987), a minimum of 10 samples for each regressor is usually required in standard NIRS techniques to avoid overfitting problems. In this study, 57–73 samples were available so we decided to restrict the number of selected wavelengths to five.

RESULTS

When running stepwise regressions, the whole data set can be used, but it is also quite common to divide it in two parts, a calibration or training set and a validation set in proportion of two thirds / one third (Card et al., 1988; Wessman et al., 1988; Curran et al., 1992). In that case, samples are generally randomly selected. Although this method is generally sound, it may confuse the issue in terms of result interpretation if there are few samples available. As evidence for this, about 50 regression equations were established on same-sized subsamples of our data set (calibration set) and were tested on the remainder (validation set). The number of selected wavebands was fixed at five in that test performed on protein concentrations and reflectance spectra. We first considered a calibration set made up of two thirds of the data set (42 and 38 samples, respectively, for fresh and dry leaves), and second a calibration set made up of one third (21 and 19 samples). Table 3 shows that, in both cases, r^2 values corresponding to the calibration sets are rather similar while those derived from the validation sets range over a wide field. Selected wavelengths also change according to the drawing lots, and differences are all the more important since the calibration set is small. Due to the difficulties in choosing an appropriate calibration set, we decided to run the stepwise regressions on all the samples.

Each chemical component (protein, cellulose, lignin, starch) was treated independently of the others, and so each yields an independent set of regression coefficients. We distinguished two groups of measurements: individual leaves (R, T, and A) and stacked individual leaves + needles (R_∞), which represented optically thick samples. First derivatives—respectively $\partial R/\partial\lambda$, $\partial T/\partial\lambda$, $\partial A/\partial\lambda$, and $\partial R_\infty/\partial\lambda$—and $\log(1/R_\infty)$ were also investigated. In each group, fresh leaves were separated from dry leaves. Tables 4 and 5 detail statistical outputs of the stepwise multiple regression analysis for proteins, cellulose, lignin, and starch. From a general point of view, correlations are better for optically thick samples than for individual leaves, and better for dry samples than for wet samples. Best r^2 and rmse values are obtained for first derivatives. As for individual leaves, $\partial R/\partial\lambda$, $\partial T/\partial\lambda$, and $\partial A/\partial\lambda$ provide very similar results. As for optically thick samples, $\log(1/R_\infty)$, which can be considered as a measure of absorptance, does not lead to significant improvement of correlations, and, as noticed earlier by Gastellu-Etchegorry et al. (1994), it is surprisingly a worse predictor of leaf biochemistry than the raw infinite reflectance. In general, protein, cellulose, and lignin are reasonably estimated with r^2 higher than 0.70 for five wavelengths. Starch, whose content in the leaves is very small ($<3\%$ of dry matter), is retrieved with more difficulty.

Let us detail these results with the case of protein. Figure 3 shows the variation of r^2 as terms are added to the regression equation; curves are very similar to that obtained by Card et al. (1988). One can note a strong increase of r^2 for the first four wavelengths and a beginning of saturation when the number of wavelengths approaches five. In that situation, raw spectra and first derivatives tend to join. The other constituents present the same trends (results not shown). Inversely, we could have plotted the variation of the root mean square error as a function of the selected wavebands, and we would have noticed that the more variables included, the smaller would be the rmse.

Table 4. Results of Stepwise Multiple Regression Analysis on Protein and Lignin for One and Five Regressors[a]

			Protein				*Lignin*			
			$N=1$		$N=5$		$N=1$		$N=5$	
			r^2	*rmse*	r^2	*rmse*	r^2	*rmse*	r^2	*rmse*
Individual leaves	Fresh	R	0.179	6.4	0.500	5.2	0.187	5.9	0.440	5.0
	(63)	$\partial R / \partial \lambda$	*0.392*	*5.5*	*0.707*	*4.0*	*0.302*	*5.4*	*0.639*	*4.0*
		T	0.090	6.7	0.556	4.9	0.078	6.2	0.495	4.8
		$\partial T / \partial \lambda$	0.415	5.4	0.781	3.4	0.309	5.4	0.609	4.2
		A	0.221	6.2	0.498	5.2	0.204	5.8	0.398	5.2
		$\partial A / \partial \lambda$	0.410	5.4	0.778	3.4	0.348	5.2	0.680	3.8
	Dry	R	0.364	5.6	0.810	3.2	0.187	5.8	0.481	4.8
	(57)	$\partial R / \partial \lambda$	0.749	3.5	0.875	2.6	0.394	5.0	0.741	3.4
		T	0.297	5.9	0.727	3.8	0.100	6.1	0.684	3.7
		$\partial T / \partial \lambda$	0.761	3.4	0.887	2.5	0.409	4.9	0.716	3.5
		A	0.463	5.1	0.752	3.6	0.175	5.8	0.561	4.4
		$\partial A / \partial \lambda$	0.734	3.6	0.874	2.6	0.410	4.9	0.722	3.5
Stacked leaves + needles	Fresh	R_∞	0.479	5.6	0.742	4.0	0.140	5.8	0.392	5.0
	(73)	$\partial R_\infty / \partial \lambda$	0.616	4.8	0.801	3.5	0.261	5.4	0.590	4.1
		$\log(1 / R_\infty)$	0.453	5.7	0.756	3.9	0.148	5.7	0.378	5.1
	Dry	R_∞	0.586	4.9	0.857	3.0	0.154	5.6	0.429	4.7
	(67)	$\partial R_\infty / \partial \lambda$	0.787	3.5	0.888	2.6	0.403	4.7	0.723	3.3
		$\log(1 / R_\infty)$	0.533	5.2	0.793	3.6	0.124	5.7	0.404	4.8

[a] The number of samples used to estimate the chemical concentrations is shown in parentheses.

The distribution of the wavelengths stemming from the stepwise multiple regression analysis is quite amazing. To illustrate it, the reflectance and transmittance spectra both of a typical fresh and dry leaf have been plotted together with the position of the first, second, ..., fifth wavelength selected in the estimation of protein (Figs. 4 and 5). First, one can notice great differences from one case to another: In particular, for fresh leaves, spectral regions sensitive to the protein concentration are completely different according to the type of measurement used. According to Card et al. (1988), it is difficult to associate particular chemical bonds with the wavelengths selected by stepwise regression since the latter depend on many factors such as the kind of data chosen (reflectance, transmittance, or absorptance). Moreover, leaves contain several constituents each with a number of absorption peaks, causing peak broadening and shifting, and so selected wavelengths do not always

Table 5. Results of Stepwise Multiple Regression Analysis on Cellulose and Starch for One and Five Regressors[a]

			Cellulose				*Starch*			
			$N=1$		$N=5$		$N=1$		$N=5$	
			r^2	*rmse*	r^2	*rmse*	r^2	*rmse*	r^2	*rmse*
Individual leaves	Fresh	R	0.040	6.4	0.495	4.8	0.046	2.1	0.359	1.8
	(63)	$\partial R / \partial \lambda$	0.237	5.7	0.616	4.2	0.156	2.0	0.537	1.5
		T	0.057	6.3	0.350	5.4	0.065	2.1	0.344	1.8
		$\partial T / \partial \lambda$	0.287	5.5	0.628	4.1	0.284	1.8	0.555	1.5
		A	0.082	6.2	0.459	4.9	0.047	2.1	0.329	1.8
		$\partial A / \partial \lambda$	0.433	4.9	0.715	3.6	0.233	1.9	0.500	1.6
	Dry	R	0.137	6.0	0.606	4.2	0.040	1.9	0.408	1.6
	(57)	$\partial R / \partial \lambda$	0.410	4.9	0.820	2.8	0.240	1.7	0.639	1.2
		T	0.037	6.3	0.686	3.7	0.042	1.9	0.575	1.3
		$\partial T / \partial \lambda$	0.388	5.0	0.844	2.6	0.265	1.7	0.684	1.1
		A	0.130	6.0	0.539	4.5	0.046	1.9	0.359	1.6
		$\partial A / \partial \lambda$	0.402	5.0	0.816	2.9	0.248	1.7	0.722	1.1
Stacked leaves + needles	Fresh	R_∞	0.248	6.1	0.596	4.6	0.129	1.9	0.357	1.7
	(73)	$\partial R_\infty / \partial \lambda$	0.419	5.4	0.699	4.0	0.229	1.8	0.478	1.5
		$\log(1 / R_\infty)$	0.244	6.1	0.598	4.6	0.128	1.9	0.401	1.6
	Dry	R_∞	0.352	5.7	0.614	4.5	0.080	1.8	0.373	1.5
	(67)	$\partial R_\infty / \partial \lambda$	0.474	5.1	0.812	3.1	0.246	1.6	0.640	1.2
		$\log(1 / R_\infty)$	0.363	5.6	0.544	4.9	0.090	1.8	0.443	1.5

[a] The number of samples used to estimate the chemical concentrations is shown in parentheses.

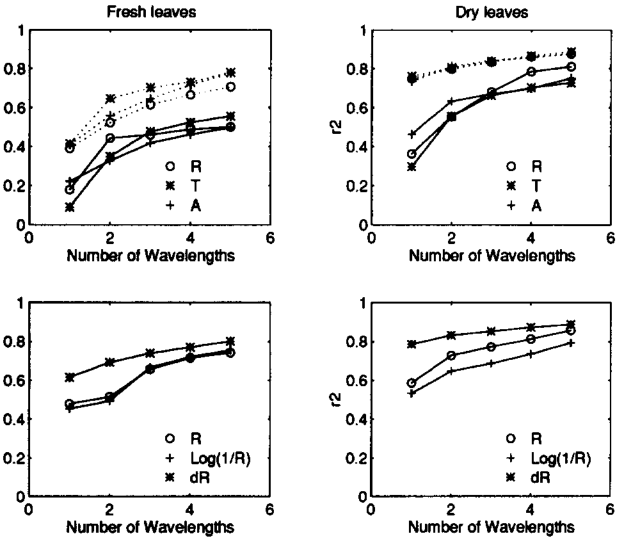


Figure 3. Multiple squared correlation coefficient r^2 versus the number of terms in the regression equation for the stepwise regression of protein concentration. Plots at the top represent the individual leaves [(O) reflectance; (*) transmittance; (+) absorptance; (—) raw values; (···) first derivative], plots at the bottom the optically thick samples (stacked leaves + needles) [infinite reflectance, (o) raw values; (+) first derivative; (*) log(1 / *R*)].

occur precisely at known stretching frequencies (Curran, 1989). In fact, in this study, they are radically different! By simulating the effects on spectral reflectance of hypothetical mixes of protein, cellulose, lignin, and starch, Conel et al. (1993) underscored such effects which complicate the interpretation of leaf optical properties in terms of leaf biochemistry.

DISCUSSION AND CONCLUSION

This investigation of leaf biochemistry by regression analysis provided results which were in good agreement with the literature; we showed that information on leaf biochemistry—protein, cellulose, lignin, and starch in this study—was attainable with five selected wavelengths. We arrived at this conclusion with very different plant species at various physiological states. The analysis of the optical properties of fresh individual leaves contributes to the originality of this study: When using vegetation powders or dry stacked leaves, the radiometric signal is enhanced, and regression analyses are known to work relatively well; the extension of such an approach to green individual leaves was not obvious due to the weak spectral signature generated by the above leaf biochemical components. The strength of this paper also lies in the analysis of reflectance and transmittance spectra of single leaves. The discrepancy between the selected wavebands in reflectance and transmittance spectra is quite curious and contradicts the idea that the specific biochemical causes produce consistent effects.

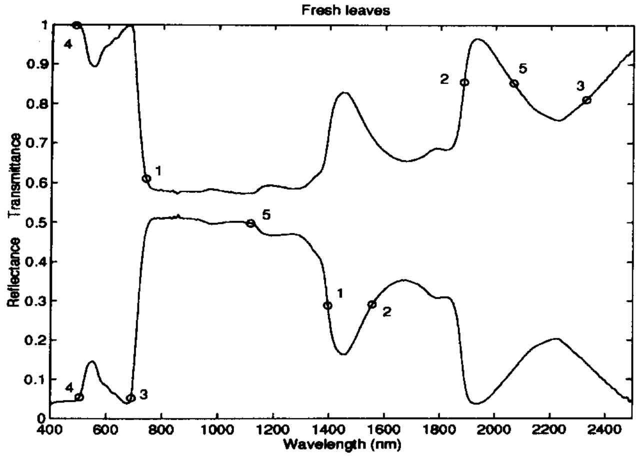
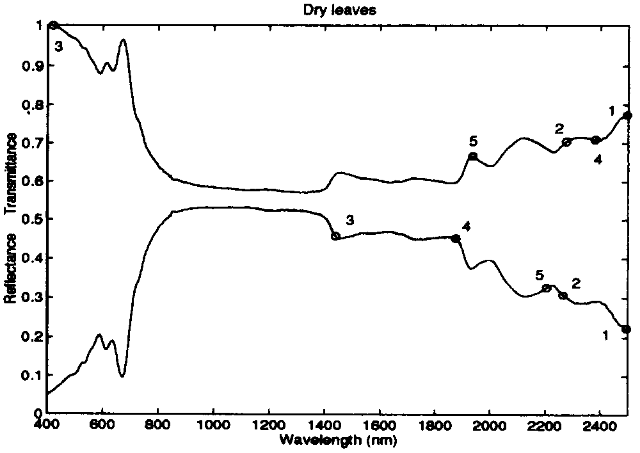


Figure 4. Reflectance and transmittance spectra of a typical fresh leaf. Wavelengths selected by the multiple stepwise regression analysis on protein are represented by circles in the order of the selection.

Concerning the methodological approach, there are several ways to run stepwise multiple regression analyses; First, in the choice of the independent variables, that is, the wavelengths (backward or forward selection, hybrid techniques); second, in the choice of the training set used to establish the statistical relationships. As pointed out before, results may be strongly dependent on this choice. In this article, the small number of samples (63 and 57, respectively, for fresh and dry individual leaves; 73 and 67, respectively, for fresh and dry optically thick samples) prevented us from proceeding in that fashion. However, in most articles, this separation is performed on the same data set which questions the validity of such an approach. In principle, it should be better to use two independent data sets, as Grossman et al. (1995) did recently, because results gain in generality.

Figure 5. Reflectance and transmittance spectra of a typical dry leaf. Wavelengths selected by the multiple stepwise regression analysis on protein are represented by circles in the order of the selection.



Protein seems to be the component that comes out best. As protein and nitrogen contents are essentially proportional, this result is in accordance with many published works. Cellulose and lignin, which stand for the carbon pool of the leaf, nevertheless, fit next best with good correlation coefficients. This is very encouraging for the prospect of the C / N ratio calculation dear to plant physiologists. However, this ratio generally approximated by the cellulose + lignin / nitrogen ratio is mainly useful at canopy level. To what extent results obtained at leaf level can be transposed at canopy level? The answer is unambiguous: Calibration equations previously established at leaf level are not directly usable with field, airborne, or spaceborne remote sensing data because a plant canopy cannot be assimilated from a pile of plant leaves. In particular, variations of canopy architecture generally parameterized by the leaf orientation and the leaf area index induce spectral distortions as shown by Jacquemoud (1993). These distortions may prevent relationships derived from fresh optically thick samples (stacked leaves + needles) to be used on semiinfinite vegetated areas. When working at a smaller scale, say with AVIRIS (Airborne Visible / Infrared Imaging Spectrometer) spectra, it is consequently necessary to establish calibration relationships on known areas, and this must be repeated if the target drastically changes with new vegetation types and different canopy structures, or if atmospheric conditions also change (Martin and Aber, 1993). In the end, it appears that regression results cannot be easily reproduced mainly due to the extraordinary variability of nature and environmental conditions.

To overcome this problem, it may be judicious to analyze plant biochemical content by using models, first at leaf level and second at canopy level. This is what we intend to do in the future by extending present leaf optical properties models to important constituents other than chlorophyll or water, for example, protein, cellulose, lignin, etc., which are potentially attainable by remote sensing as a result of specific absorption bands observed in the middle infrared. It should help us to understand their specific effects on the radiometric signal.

This research was undertaken at JRC / Institute for Remote Sensing Applications / Unit for Advanced Techniques, Ispra, Italy. We are grateful to A. Pedrini of the AT Unit for his help in the calibration of the spectrophotometer data, and to S. E. Hornig for his assistance during the experiment.

REFERENCES

Barton, F. E., II (1987), Analytical applications to fibrous foods and commodities, in *Near-Infrared Technology in the Agricultural and Food Industries* (Williams and Norris, Eds.), American Association of Cereal Chemists, St. Paul, MN, pp. 169–183.

Card, D. H., Peterson, D. L., and Matson, P. A. (1988), Prediction of leaf chemistry by the use of visible and near infrared spectroscopy, *Remote Sens. Environ.* 26:123–147.

Conel, J. E., van den Bosch, J., and Grove, C. I. (1993), Application of a two-stream radiative transfer model for leaf lignin and cellulose concentrations from spectral reflectance measurements, in *Summaries 4th Annual JPL Airborne Geoscience Workshop, Volume 1. AVIRIS Workshop* (R. O. Green, Ed.), JPL Publication 93-26, Jet Propulsion Laboratory, Pasadena, CA, pp. 39–51.

Curran, P. J. (1989), Remote sensing of foliar chemistry, *Remote Sens. Environ.* 30:271–278.

Curran, P. J., Dungan, J. L., Macler, B. A., Plummer, S. E., and Peterson, D. L. (1992), Reflectance spectroscopy of fresh whole leaves for the estimation of chemical concentration, *Remote Sens. Environ.* 39:153–166.

Gastellu-Etchegorry, J. P., Zagolski, F., Mougin, E., Marty, G., and Dubuisson, P. (1994), Forest canopy chemistry with high spectral resolution—an evaluation of the atmospheric influence, in *Proc. 6th Int. Symp. Physical Measurements and Signatures in Remote Sensing,* CNES, pp. 261–268.

Grossman, Y. L., Sanderson, E. W., and Ustin, S. L. (1994), Relationships between leaf chemistry and reflectance for plant species from Jasper Ridge Biological Preserve, California, in *Proc. 14th Int. Geosci. Remote Sens. Symp. (IGARSS'94),* Pasadena, CA, 8–12 August.

Grossman, Y. L., Ustin, S. L., Jacquemoud, S., Sanderson, E. W., Schmuck, G., and Verdebout, J. (1995), Examination of several approaches for extraction of leaf biochemistry information from leaf reflectance data, *Remote Sens. Environ.* forthcoming.

Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, G., and Schmuck, G. (1995), Leaf Optical Properties EXperiment 93 (LOPEX93), Report EUR-16095-EN, European Commission, Joint Research Centre, Institute for Remote Sensing Applications, Ispra, Italy.

Hruschka, W. R. (1987), Data analysis: wavelength selection methods, in *Near-Infrared Technology in the Agricultural and Food Industries* (Williams & Norris, Eds.), American Association of Cereal Chemists, St. Paul, MN, pp. 35–55.

Jacquemoud, S. (1993), Inversion of the PROSPECT + SAIL canopy reflectance model from AVIRIS equivalent spectra: theoretical study, *Remote Sens. Environ.* 44:281–292.

Kramer, P. J., and Kozlowski, T. T. (1979), *Physiology of Woody Plants,* Academic, London.

Lichtenthaler, H. K. (1987), Chlorophylls and carotenoids: pigments of photosynthetic biomembranes, *Methods Enzymol.* 148:350–382.

Martin, M. E., and Aber, J. D. (1993), Measurements of canopy chemistry with AVIRIS data at Blackhawk Island and Harward Forest, in *Summaries 4th Annual JPL Airborne Geoscience Workshop, Volume 1. AVIRIS Workshop* (R. O. Green, Ed.), JPL Publication 93-26, Jet Propulsion Laboratory, Pasadena, CA, pp. 113–116.

McLellan, T. M., Aber, J. D., Martin, M. E., Melillo, J. M., and Nadelhoffer, K. J. (1991), Determination of nitrogen, lignin, and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy, *Can. J. For. Res.* 21:1684–1688.

NASA (1994), Accelerated Canopy Chemistry Program Final Report to NASA-EOS-IWG, NASA, Washington, DC, 33 pp.

Peterson, D. L. (1991), Report on the workshop Remote Sensing of Plant Biochemical Content: Theoretical and Empirical Studies, NASA Ames Research Center, MS 242-4, Moffet Field, CA, 50 pp.

Peterson, D. L., and Hubbard, G. S. (1992), Scientific issues and potential remote-sensing requirements for plant biochemical content, *J. Imaging Sci. Technol.* 36(5):446–456.

Ustin, S. L., Wessman, C. A., Curtiss, B., Kasischke, E., Way, J. B., and Vanderbilt, V. C. (1991), Opportunities for using the EOS imaging spectrometers and synthetic aperture radar in ecological models, *Ecology* 72:1934–1945.

Van Soest, P. J., and Wine, R. J. (1968), Determination of lignin and cellulose in acid detergent fiber with permanganate, *J. Assoc. Off. Anal. Chem.* 51:780–785.

Wessman, C. A. (1994), Estimating canopy biochemistry through imaging spectrometry, in *Imaging Spectrometry—A Tool for Environmental Observations* (J. Hill and J. Mégier, Eds.), ECSC, EEC, EAEC, Brussels and Luxembourg, pp. 57–69.

Wessman, C. A., Aber, J. D., Peterson, D. L., and Melillo, J. M. (1988), Foliar analysis using near infrared spectroscopy, *Can. J. For. Res.* 18:6–11.

Williams, P., and Norris, K. (1987), *Near-Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, St. Paul, MN, 330 pp.