Investigation of Leaf Biochemistry by Statistics

S. Jacquemoud*, J. Verdebout, G. Schmuck, G. Andreoli, B. Hosgood, S.E. Hornig
Joint Research Centre / Institute for Remote Sensing Applications
Unit for Advanced Techniques - TP 272
21020 Ispra (Va), Italy
tel: (39)-332-785181 fax: (39)-332-785469 email: stephane.jacquemoud@cen.jrc.it

Abstract

The biochemical content (total proteins, cellulose, lignin, and starch) of about seventy plant leaves has been related to their optical properties through statistical relationships. Both fresh and dry plant material, leaves or 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 infinite reflectance values (stacked leaves + needles), on rough values and on transformations of them such as the first derivative or the logarithm of the inverse reflectance. They underscored good prediction performances for protein, cellulose, and lignin with squared multiple correlation coefficients (R^2) higher than 0.8 using ten wavelengths. Starch whose concentration in the leaf was smaller compared to the other components was retrieved with less accuracy. As expected, dry material and optically thick samples provided respectively better results than fresh material and individual leaves.

INTRODUCTION

Estimating leaf biochemistry 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). It has also implications in agriculture to follow crop development and yield predictions. The biochemical constituents of interest are lignin, nitrogen, starch, and cellulose. As leaves are the most important surfaces of a plant canopy, relating their optical properties to these constituents is a priority. Surprising as it may seem, many references on this subject concern experiments performed on dry leaves or needles, and the results provided sometimes ensue directly from vegetation powders (Card et al., 1988; Wessman et al., 1988, Gastellu-Etchegorry et al., 1994). Only a few studies deal with fresh and green leaves (Curran et al., 1992). Moreover, if we except the recent paper from Conel et al. (1993) who used a two-stream radiative transfer model, only statistical analyses have been performed to retrieve leaf biochemical components from leaf reflectance. We intend here to investigate the retrieval of protein, cellulose, lignin, and starch by statistics both on fresh and dry material, on individual leaves and on optically thick samples (stacked leaves + needles).

THE EXPERIMENT

A laboratory experiment has been organized in the Joint Research Centre during the summer of 1993: we have undertaken to build a unique data set associating VIS-IR spectra of vegetation elements (leaves, conifer needles, stems, bark) both with physical measurements (blade thickness, water content, equivalent water thickness, specific leaf area) and biochemical analyses (photosynthetic pigments, protein, cellulose, lignin, and starch concentration). In order 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 outdoors. About 40 species of woody and herbaceaous plants were obtained from trees and crops in the area of the JRC. Total proteins, cellulose, lignin, and starch were measured by two specialized European laboratories (French and Belgian) using standard wet chemical analyses: the comparison between the concentration values provided by the two laboratories gave us an idea of the precision of these analyses. Protein and cellulose measurements were quite consistent while lignin and starch measurements showed some discrepancy. In this paper, only results of the Belgian laboratory are presented (they are very similar to the French ones). A Perkin Elmer Lambda 19 spectrophotometer equipped with an integrating sphere allowed the measurement of the directional-hemispherical reflectance (R) and transmittance (T) of the upper faces of 5 leaves. 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 with 1 nm step and special attention has been paid to the calibration problems. Finally, some leaves and needles of each species were dried, and the above procedure was repeated to analyse the effects of a lack of water in the middle infrared. Table 1 summarizes the radiometric data acquired in the framework of this experiment.

type of sample	water status	spectra	type of spectrum
individual leaves	fresh	63	R, T, R^∞
	dry	57	
needles	fresh	10	R^∞
	dry	10	

Table 1. *Spectrophotometric measurements.*

THE ANALYSIS

A stepwise multiple regression analysis which consists in relating the biochemical content C with the reflectance (or other spectral property) at 1, 2,...,N wavelengths [$R(\lambda_1)$, $R(\lambda_2)$,...,$R(\lambda_N)$] was performed:

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

The regression model is constructed adding the independent variables (reflectances for instance) one at a time. The first step is to choose the single variable which is the best statistical predictor; the second independent variable to be added to the regression equation is that which provides the best fit in conjunction with the first variable. Further variables are then added in this recursive fashion, adding at each step the optimum variable, given the other variables already in the equation. The wavelength interval was increased from 1 nm to 5 nm in order to reduce the calculation time: in total, 421 narrow wavebands were available. We distinguished two groups of plant material: individual leaves (R, T, and A) and stacked individual leaves + needles (R^∞) which represent optically thick samples. Finally, 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.

RESULTS AND DISCUSSION

Each chemical component (protein, cellulose, lignin, starch) was treated independently of the others, and so each yields an independent set of regression coefficients. Goodness of fit is

* Permanent affiliation: LAMP/OPGC, Université Blaise Pascal, 63177 Aubière, France

0-7803-1497-2/94 $4.00

			Proteins						Lignin					
			N=1		N=5		N=10		N=1		N=5		N=10	
			R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse	R^2	rmse
individual leaves	**fresh (63)**	**R**	0.1787	6.30	0.4998	4.91	0.8030	3.08	0.1870	5.77	0.4403	4.79	0.6847	3.59
		∂R/∂λ	0.3915	5.42	0.7074	3.76	0.8634	2.57	0.3023	5.34	0.6386	3.85	0.8200	2.71
		T	0.0897	6.63	0.5557	4.63	0.8132	3.00	0.0784	6.14	0.4946	4.55	0.6630	3.71
		∂T/∂λ	0.4148	5.32	0.7806	3.25	0.8836	2.37	0.3094	5.32	0.6087	4.00	0.7808	3.00
		A	0.2212	6.13	0.4976	4.93	0.8267	2.89	0.2044	5.71	0.3980	4.96	0.6089	4.00
		∂A/∂λ	0.4096	5.34	0.7775	3.28	0.9008	2.19	0.3475	5.15	0.6800	3.62	0.8492	2.49
	dry (57)	**R**	0.3636	5.50	0.8100	3.01	0.8935	2.25	0.1865	5.66	0.4812	4.52	0.6089	3.92
		∂R/∂λ	0.7489	3.46	0.8745	2.45	0.9400	1.69	0.3943	4.88	0.7405	3.20	0.8500	2.43
		T	0.2974	5.78	0.7271	3.61	0.8623	2.56	0.1003	5.95	0.6838	3.53	0.7974	2.82
		∂T/∂λ	0.7605	3.38	0.8870	2.32	0.9468	1.59	0.4092	4.82	0.7161	3.34	0.8497	2.43
		A	0.4632	5.06	0.7517	3.44	0.9050	2.13	0.1751	5.70	0.5605	4.16	0.7477	3.15
		∂A/∂λ	0.7337	3.56	0.8744	2.45	0.9454	1.61	0.4102	4.82	0.7220	3.31	0.8360	2.54
stacked leaves + needles	**fresh (73)**	**R∞**	0.4790	5.49	0.7420	3.86	0.8432	3.01	0.1402	5.69	0.3921	4.78	0.5700	4.02
		∂R∞/∂λ	0.6159	4.71	0.8011	3.39	0.9099	2.28	0.2606	5.28	0.5897	3.93	0.7526	3.05
		log(1/R∞)	0.4528	5.63	0.7556	3.76	0.8518	2.93	0.1482	5.66	0.3782	4.84	0.5598	4.07
	dry (67)	**R∞**	0.5860	4.83	0.8568	2.84	0.8977	2.40	0.1535	5.50	0.4294	4.51	0.6873	3.34
		∂R∞/∂λ	0.7868	3.46	0.8884	2.51	0.9420	1.81	0.4031	4.62	0.7225	3.15	0.8708	2.15
		log(1/R∞)	0.5325	5.13	0.7926	3.41	0.8752	2.65	0.1235	5.60	0.4044	4.61	0.6450	3.56

Table 2. *Results of stepwise regression on leaf spectral characteristics: [chemical concentration] = $a_0+a_1.R(\lambda_1)+a_2.R(\lambda_2)+...+a_N.R(\lambda_N)$*

measured by R^2, the squared multiple correlation coefficient which can be interpreted as the fraction of total sum of squares explained by the regression. The root mean square error *rmse* is calculated as:

$$rmse = \sqrt{\frac{(C-C^*)^2}{n}}$$

where C and C* are respectively the measured and estimated concentrations, and n is the number of individuals. Let us take the case of total proteins: Figure 1 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 five wavelengths and a saturation effect when the number of wavelengths approaches ten. The other constituents present the same trends (results not shown).

Table 2 details statistical outputs of the stepwise multiple regression analysis for total proteins and lignin. We chose these two constituents because, according to Wessman (1994), the lignin / nitrogen ratio which drives the decomposition rates of forest litter is an indicator of ecosystem processes. 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, ∂R/∂λ, ∂T/∂λ, and ∂A/∂λ provide very similar results. As for optically thick samples, log(1/R∞) which can be considered as a measure of absorptance, is surprisingly a worse predictor of leaf biochemistry than the raw infinite reflectance. In general, total proteins, cellulose, and lignin are reasonably estimated with R^2 higher than 0.70 with 5 wavelengths, and R^2 higher than 0.85 with 10 wavelengths. Starch whose content in the leaves is very small (<10% of dry matter) is retrieved with more difficulty.

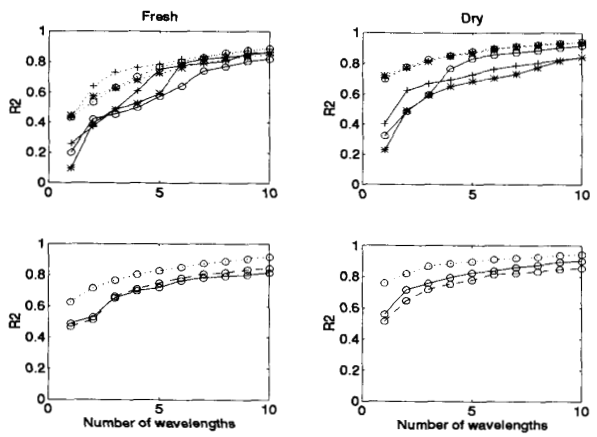

Figure 1.
Multiple squared correlation coefficient R^2 versus the number of terms in the regression equation for the stepwise regression of protein concentration. Plots at the top represent the individual leaves [o reflectance, * transmittance, + absorptance, — raw values, ··· first derivative], plots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

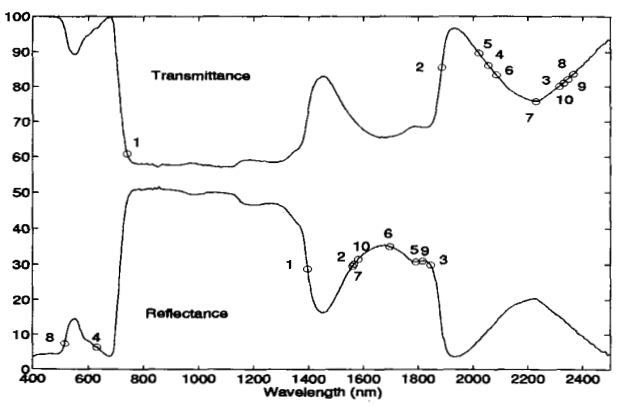

Figure 2.
Reflectance and transmittance spectra of a typical fresh leaf. Wavelengths selected by the stepwise multiple regression analysis to explain protein are represented by circles.

The distribution of the wavelengths stemming from the stepwise multiple regression analysis is quite amazing. To illustrate it, we have plotted the reflectance and transmittance spectra both of a typical fresh and dry leaf together with the position of the first, second, third, until tenth wavelength selected in the estimation of protein (Figures 2 and 3). 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 independent. One can also notice groups of wavelengths very close together. 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 occur precisely at known stretching frequencies: 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.

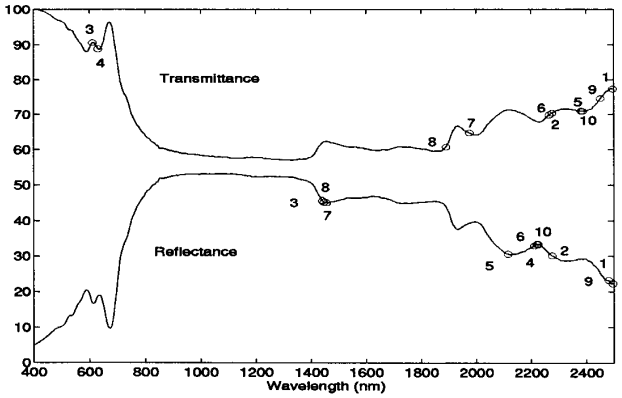

Figure 3.
Reflectance and transmittance spectra of a typical dry leaf. Wavelengths selected by the stepwise multiple regression analysis to explain protein are represented by circles.

Nethertheless, Figure 4 shows potential for predicting chemical concentrations over a wide range of conditions: the plots all indicate a good distribution for the biochemical components and a low bias in the regression as shown by symmetry of points about the 45° line through the origin.

CONCLUSION

This investigation of leaf biochemistry by statistics provided results which are in good agreement with the literature; the analysis of the optical properties of fresh individual leaves contributes to the originality of this study. We have shown that information on leaf biochemistry was attainable with only a few selected wavelengths. Proteins seem to be the biochemical component that come out the best. However, the calibration equations previously established are not directly usable with remote sensing data. On the fringe of this study, we have also simulated AVIRIS equivalent spectra and repeated the multiple stepwise regression analyses with a set of wavelengths carefully chosen outside the atmospheric absorption regions. Althought a plant canopy cannot be assimilated to a pile of leaves, relationships obtained with fresh optically thick samples (stacked leaves + needles) might be used on semi-infinite vegetated areas. Finally, in the future, we intend to analyse plant biochemical content by using models. The extension of present leaf optical properties models to important constituents other than chlorophyll or water, which are potentially attainable by remote sensing as a result of specific absorption bands observed in the middle infrared (cellulose, lignin, sugar, nitrogen...etc), should help us to understand their specific effects on the radiometric signal.

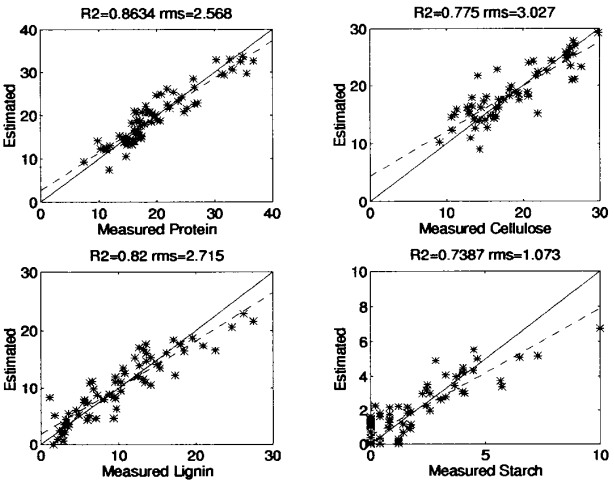

Figure 4.
Protein, cellulose, lignin, and starch predicted concentrations (fresh individual leaves, $\partial R/\partial\lambda$, 10 wavelengths) versus concentrations according to wet chemical analyses.

ACKNOWLEDGMENTS

We are grateful to A. Pedrini of the Advanced Techniques Unit of the JRC fot his help in the calibration of the spectrophotometer data.

REFERENCES

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

Conel, J.E., van den Bosch, J., Grove, C.I. "Application of a two-stream radiative transfer model for leaf lignin and cellulose concentrations from spectral reflectance measurements, in Proc. 4th Annual JPL Geoscience Workshop, Vol. 1: AVIRIS Workshop (R.O. Green, Ed), Washington DC, 25-29 October 1993, pp. 39-51.

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

Gastellu-Etchegorry, J.P., Zagolski, F., Mougin, E., Marty, G., Dubuisson, P. "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, Val d'Isère (France), 17-21 January 1994, in print.

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

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

Wessman C.A. "Estimating canopy biochemistry through imaging spectrometry" in *Imaging Spectrometry as a Tool for Environmental Observations* (J. Hill, Ed), Kluwer Academic Publishers (Dordrecht), 1994, in print.