Information Content of HyMap Hyperspectral Imagery

Cédric Bacour[1], Frédéric Baret[1], Stéphane Jacquemoud[2]
[1]*INRA-CSE, site Agroparc, domaine Saint-Paul, 84914 Avignon Cedex 09*
[2]*LED-Université Paris 7, CP 7071 - 2, place Jussieu, 75251 Paris Cedex 05*
cbacour@avignon.inra.fr, baret@avignon.inra.fr, jacquemo@ccr.jussieu.fr

ABSTRACT- *Hyperspectral characteristics of the HyMap airborne instrument are used to determine the minimum number of wavebands useful for accurate retrieval of canopy biophysical variables. The information content of a reflectance spectrum indicates the number of independent variables that explain its variance. It is usually determined statistically and leads to the identification of the spectral regions the most sensitive to variations of these variables. Here, a sensitivity analysis of the PROSPECT+SAIL model is performed with the aim of determining the most informative HyMap spectral bands on the dynamics of the canopy biophysical variables. The relevance of such optimal wavelengths is then assessed in inverse mode, where the variables are estimated from real reflectance spectra acquired during the DAISEX 1999 (Digital AIrborne Spectrometer EXperiment) campaign. Emphasis is on the estimation of the leaf chlorophyll content* C_{ab} *and the leaf area index LAI.*

1 INTRODUCTION

Inversions of canopy reflectance models have spread during the last decade to estimate vegetation characteristics. In comparison with empirical or semi-empirical methods, physically-based models better account for the interdependence between canopy state variables. Nevertheless, the non-unicity of the solution turns out to be a limiting factor for reliable estimates. Recent efforts to develop computationally efficient inversion techniques such as neural networks and look-up tables, and to regularize the inverse problem by introducing prior information on the variables, have only partially overcome this issue. A further approach may take advantage of optimal sampling configurations which i) express the best adequacy between the models and "reality" (i.e., between the input variables and the "real" canopy state variables; between the simulated and the measured reflectances), and ii) carry as much information as possible.

The inverse problem consists here in determining the set of model variables such that the simulated reflectances comply the best with observations. The search for an optimal set of canopy variables, by all the acceptable solutions, implicitely supposes that there is a particular combination of reflectances associated to it. The aim of this study is to determine the best choice of *N* observations, among *M* available (*N* being smaller than *M*), that leads to the best estimation of the canopy biophysical variables. The determination of such optimal configurations of observation is in progress (Kimes et al., 2000) and is advanced by spatial agencies (CNES, ESA, NASA) to improve the quality of remote sensing products and the definition of new instruments.

In remote sensing, the concept of *information content* of a reflectance spectrum has been first introduced by the pioneers of imaging spectroscopy applied to soils and plant canopies, even though this issue was not considered for inversion purposes. It measures the number of independent variables that explain most of the observed variability. Its determination is inseparable from the identification of the spectral regions the most sensitive to these variables (Price, 1975). Typically, five dimensions satisfactorily described the variability of radiometric signals measured over vegetation (Price, 1992; Curran, 2001): two in the visible, one in the near infrared, and two in the middle infrared. The selection of a limited number of wavebands for the estimation of plant canopy characteristics is generally made statistically (multiple regression analysis for instance).

In this paper, we propose an alternative approach based on the sensitivity analysis of a canopy reflectance model (the issue of adequacy between model and reality is not considered here since we assume that the model *is* reality). The PROSPECT+ SAIL model is used in the observation configuration of HyMap to determine the best wavelengths for the estimation of canopy biophysical variables, in particular the leaf chlorophyll content C_{ab} and the leaf area index *LAI* that are the two most relevant biophysical variables that reveal vegetation state and functioning.

Then, inversions of the coupled canopy reflectance model on HyMap reflectance spectra acquired during the DAISEX99 campaign enable to validate these preliminary results.

2 SENSITIVITY ANALYSIS

2.1 The model

The SAIL radiative transfer model (Verhoef, 1984, 1985) is widespread in the remote sensing community for the estimation of vegetation biophysical variables. It calculates the canopy reflectance, provided the leaf optical properties and a limited number of variables describing its architecture: the leaf area index *LAI*, the mean leaf inclination angle *ALA*, assuming an ellipsoidal distribution of foliage elements (Campbell, 1990), the hot spot parameter s_l, and a soil brightness parameter α_{soil}.

It is coupled with the PROSPECT model in order to account for the leaf optical properties. The version used here requires the leaf structure parameter *N*, the chlorophyll a and b content C_{ab} (µg.cm^{-2}), the equivalent water thickness C_w (cm), the dry matter content C_m (g.cm^{-2}), and the brown pigment concentration C_{bp} (Jacquemoud and Baret, 1990; Baret and Fourty, 1997), to simulate leaf reflectance and transmittance spectra in the optical domain.

2.2 Experimental design

Design of numerical experiments recently emerged in the field of remote sensing for sensitivity analyzes of complex computational models (Bacour et al., 2002). They allow better sampling of the parameter space in a limited number of simulations where all the input variables vary simultaneously (Benoist et al., 1994). We used such an method to study the influence of each variable of the model within its range of variation.

The space of canopy realizations is determined after a Hyper Graeco Latin Geometric sampling scheme, the resolution of which allows full investigation of all interactions between two variables. The companion experimental design is made of 2401 simulations corresponding to different combinations of the eight PROSPECT+SAIL input variables, each of them taking one over seven values equidistributed within its definition range (Table I). As the simulations are conducted in the principal plane where the hot spot parameter has very little influence, the latter is fixed to 0.25.

The observation configurations comply with those of the HyMap instrument used in the DAISEX99 campaign (see §3.1): sun zenith angle θ_s of 17°, view zenith angle θ_v varying from 0° to 25° with a 5° step, and relative azimuth angle ϕ equal to 100° and 280°. 89 over 128 wavebands are used. They cover the solar spectrum from 457 to 2271 nm.

		Lower bound	**Upper bound**
Leaf	*N*	1	3
	C_{ab}	1	100
	C_m	0.002	0.02
	C_w	0.04	0.04
	C_{bp}	0	1
Canopy	*LAI*	0	8
	ALA	30	85
	α_{soil}	0.5	3

Table I. Variation of the PROSPECT+SAIL input parameters used in the experimental design.

2.3 Sensitivity analysis

The effects and contributions of the model variables are assessed from the whole set of simulations (for more details, see Bacour et al., 2002). The mean effect of a variable *v* represents the distance between the mean values of the model responses when *v* is on level *n*, $\bar{\rho}_{vn}$, and the general mean $\bar{\rho}$. They are expressed as a percentage:

$$E_{vn} = \frac{\bar{\rho}_{vn} - \bar{\rho}}{\bar{\rho}} \times 100 \qquad (1)$$

Since the results present only a slight dependence with the view zenith angle, we will deal with directional averaged values hereafter. The spectral sensitivity of the model variables is summarized in Figure I: given a wavelength, the tangent to the effect surface expresses the ability of the model to link variations of reflectance levels to variations of the biophysical variables: a positive (respectively, negative) slope means that increasing the value of a variable translates into an increase (respectively, decrease) of the reflectance; moreover, the sharper the slope is, the more sensitive the reflectance is to variations of the considered variable. One can observe a quasi-exponential decrease of C_{ab}, C_w, and *LAI* effects on their definition range, with a noticeable reversal for the leaf area index in the green (where increasing biomass tends to increase the reflectance). On the other hand, effects of the other variables turn out to be almost linear.

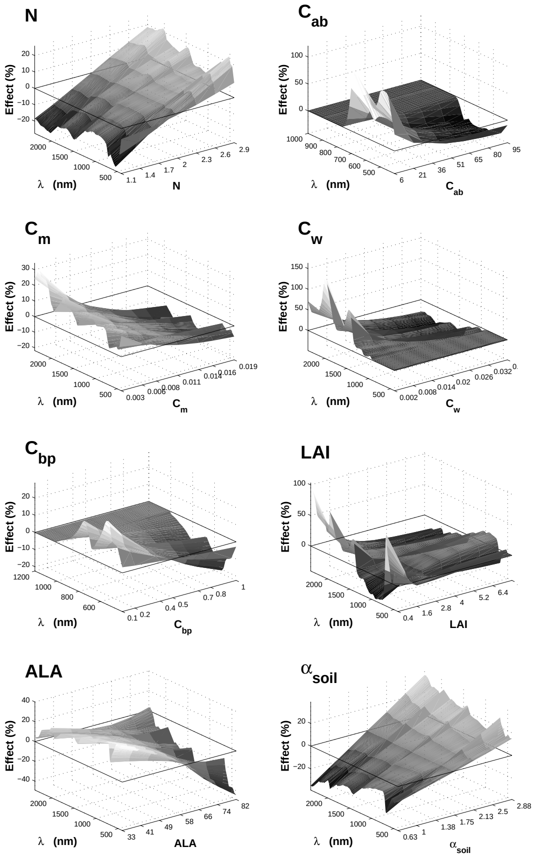

Figure I. Mean effects of the PROSPECT+SAIL variables as a function of the wavelength: N, C_{ab}, C_m, C_w, C_{bp}, LAI, ALA, and α_{soil}.

Whereas the study of the variable effects reveals their sensitivity, the determination of their relative contribution helps organizing their own influence on the canopy reflectance into a hierarchy. For each variable, the contribution index C_v characterizes the variance it explains:

$$C_v = \frac{\frac{N}{n}\sum_{1\le m\le n}\left[\rho_{vm}-\bar{\rho}\right]^2}{\sum_{1\le k\le N}\left[\rho_k-\bar{\rho}\right]^2}\times 100 \qquad (2)$$

for N = 2401 simulations and m = 7 levels taken by each variable.

As illustrated by Figure II, the total contribution of the variables and their interactions almost equal 100%; the residues are attributed to inherent computational errors and should be regarded as noise. Figure II clearly shows the spectral influence of each variable: in the visible, the chlorophyll content drives about 50% of the reflectance variations, with a weaker contribution near 550 nm; in the near-infrared, the most important variables are the leaf angle parameter and the leaf area index; the middle infrared confirms the strong influence of light absorption by the leaf water content around 1450 and 1940 nm.

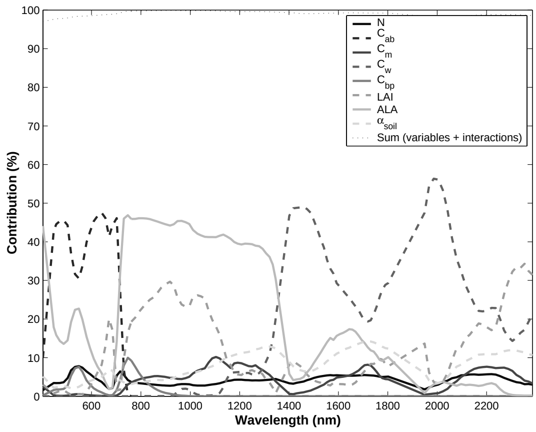

Figure II. Spectral variation of the contribution of the PROSPECT+SAIL variables.

2.4 Definition of optimal wavebands

In order to determine optimal wavebands for the estimation of C_{ab} and LAI, let us define sensitivity S_v and optimality I indices which combine both the sensitivity and contribution of the variables:

$$S_v = \sum_{1\le m\le n}\sqrt{\Delta(E_{vm})^2} \qquad (3)$$

and

$$I = (S_{Cab}+S_{LAI})\times\left(C_{Cab}+C_{LAI}-\sum_{i=1}^{m} I^{i}_{Cab/LAI}\right) \qquad (4)$$

S_v expresses the cumulative magnitude of the sensitivity of the variable v (i.e. the slope of its mean effects) on the range of variation. In the definition of the optimality index, the impact of the interactions where C_{ab} and/or LAI are involved, has been taken into account so as to reduce possible compensations between variables during inversions. Each wavelength has been attributed such indices. Then, the selection of optimal spectral bands is only based on the values of I. Because of the distribution of the latter in the solar spectrum – the maximum values are all located in the near infrared – we decided to split it into three domains: the visible (400-700 nm), the near infrared (700-1300 nm), and the middle infrared (1300-2300 nm). Also, another hierarchy has been established to account for correlations between the wavelengths: for each wavelength λ_0, the optimality index I is weighted by a factor that depends of the value of the correlation of λ_0 with the other wavebands comprised within a given correlation length (that is, every spectral band λ_i verifying $corr(\lambda_0,\lambda_i) \geq 1/e$). The goal was to spread the information of adjacent wavebands on the solar spectrum with respect to physical assumptions (Figure III).

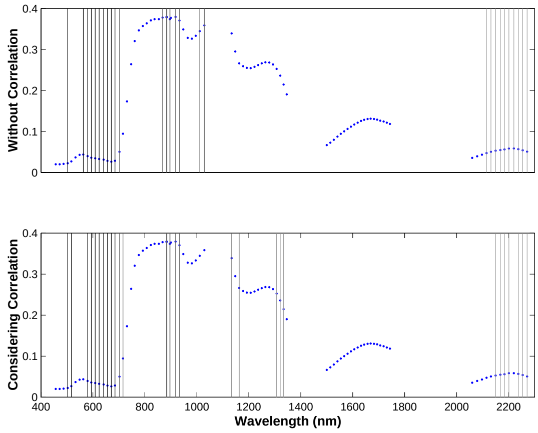

Figure III. Position of the eight most optimal wavelengths in the visible (400-700 nm), near infrared (700-1300 nm), and middle infrared (1300-2300 nm) domains, together with an experimental wheat reflectance spectrum, whether the correlation between wavebands is taken into account (top) or not (bottom).

3 CANOPY BIOPHYSICAL VARIABLE ESTIMATION

The relevance of the optimality indices determined previously is appraised with HyMap measurements, and illustrated with the leaf area index for which the ground truth was available. The aim is to assess whether the use of the selected wavebands determined as above improves the estimation of *LAI* or not, by comparison with inversions performed with the whole radiometric information.

3.1 The DAISEX99 campaign

The scientific goal of DAISEX (Berger et al., 2001; Müller et al., 2001; Moreno et al., 2001) was to demonstrate the retrieval of geo- bio-physical variables from imaging spectrometers. In this context, the Barrax (Spain) test site – an agricultural flat area of 3 km × 3 km centered on 39°3'N, 2°5'W – was monitored the 3rd of June 1999 with HyMap and *in situ* to determine some biophysical variables including the leaf area index of corn (*Zea mays* L.), sugar beet (*Beta vulgaris* L.), and barley (*Hordeum vulgare* L.).

3.2 Inversions

Inversions of PROSPECT+SAIL are conducted by means of a quasi-Newton algorithm to minimize the misfit function S^2 that characterizes the gap between measured R_{mes} and simulated R_{mod} reflectances:

$$S^2 = \sum_{i=1}^{N} \left[R^i_{mes} - R^i_{\mathrm{mod}} \right]^2 \qquad (5)$$

The routine E04JYF of the Numerical Algorithm Group library allows fixing the upper and lower bounds of variation of the nine variables estimated simultaneously (the eight ones defined in Table I plus the *hot spot* parameter). Also, 10 sets of variables, with values drawn randomly according to a uniform distribution, are used as initial guess of the inversion process. In the following, the estimated *LAI* values are averages of these 10 estimates.

3.3 Studied cases

Different combinations of optimal wavelength selections have been studied for the estimation of the leaf area index (Table II). In each case, the estimations were made with the previously selected wavelengths, considering or not the correlation, as well as with randomly drawn wavelengths.

The use of such a limited radiometric information for infering biophysical variables by inversion rises the question of determination of the optimization problem. Theoretically, solving the inverse problem imposes the system to be over-determined, i.e. the number of variables M to estimate must be at most equal to the number of measurements N; the solution is therefore not unique. However, from a statistical point of view, the higher the number of avalaible data is (considering they are noise free), the more consistent the estimation is. The ratio N/M is therefore an indicator of the confidence one can expect in the estimates.

	Number of wavebands				Ratio
	VIS	NIR	MIR	Total	*N / M*
# 1	16	36	37	89	9.9
# 2	16	36	—	52	5.8
# 3	8	10	—	18	2
# 4	10	10	—	20	2.2
# 5	6	6	6	18	2

Table II. Different cases tested for the estimation of the leaf area index with HyMap.

Results of the *LAI* estimations are gathered in Figure IV. Reducing the spectral information generally leads to debase the estimates of low *LAI* values (below 1), as illustrated by #2 to #5. This trend may result from different soil conditions between the model and real crops: in the experimental designed simulations, the soil is assumed Lambertian; in the inversions, it is anisotropic and characterized by the SOILSPECT model (Jacquemoud et al., 1992) whose input parameters were estimated on a bare soil from HyMap spectra. For higher *LAI* values, the estimates better match the *in situ* measurements, even when few radiometric data are available (#3, #4, and #5). The introduction of spectral correlations in the determination of the optimal wavelengths does not improve the estimations. Moreover, inversions performed at the "optimal" wavelengths led to better results, with regard to the ground truth, than when using random ones, pointing out a certain relevance of the concept of optimal configuration.

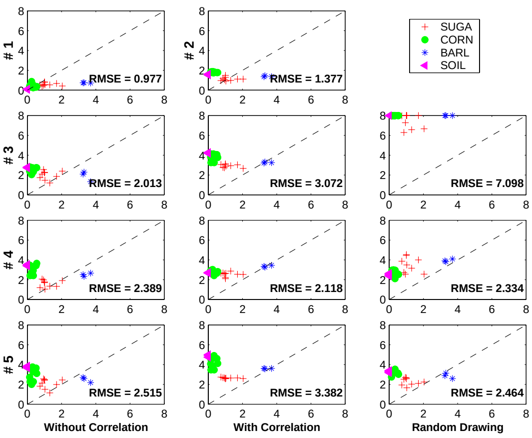

Figure IV. Estimated values of the leaf area index (y-axis) against the *in situ* measurements (x-axis) for different selections of the wavebands (see Table II).

4 CONCLUSION

Improved estimation of canopy biophysical variables through inversion of a radiative transfer model may be obtained by taking advantage of optimal sampling configurations i) carrying as much information as possible, and ii) providing the best adequacy between the models and "reality". The first point was investigated in this paper with the PROSPECT+SAIL model.

In a first step, a sensitivity analysis of the model enables to point out some spectral features where the influence of the leaf area index and the chlorophyll a plus b content is prominent as compared to the other variables. On this basis, different sets of optimal wavelengths were determined. These were then employed in inversion on HyMap hyperspectral reflectance spectra acquired during the DAISEX99 campaign. It appeared that, even for limited radiometric information, high LAI values were consistently estimated when compared to the ground truth (results were even better than when using all the reflectances). Conversely, lower *LAI* values (<1) were poorly retrieved, probably because of different soil characteristics between simulations and actual canopies. The hypothesis of a homogeneous canopy is also no more valid in that case.

These preliminary results point out that the relevance of the "optimal" wavelengths derived in this study still remains ambiguous. Their validation in the inversion process is limited by the number of canopies considered and the lack of information on the experimental errors (for the radiometric part as well as for the *LAI* ground truth). It also appears that most of the wavebands retained are contiguous, i.e. their information content is correlated. Other choices in the empirical determination of the optimality indices may have lead to different results: more emphasis could have been put on either the sensitivity or the contribution, different weights could have been laid to the parameter interactions, etc. Finally, their use in the prospect of estimation of canopy biophysical variables with model inversion is restricted by the determination of the inverse problem (i.e. the number of variable to estimate vs. the number of available data). Quite obviously, the potentiality of exploiting optimal wavelengths should be enhanced by addition of directional information.

Acknowledments: The authors greatly thank Yves Tourbier, Renault - Direction de la Recherche, who provided the Hyper Graeco Latin Geometric experimental design used in the study.

References

Bacour, C., Jacquemoud, S., Tourbier, Y., Dechambre, M., Frangi, J.-P., 2002, Design and analysis of numerical experiments to compare four canopy reflectance models. *Remote Sensing of Environment*, **79**, 72-83.

Baret, F., Fourty, T., 1997, Radiometric estimates of nitrogen status in leaves and canopies. In Diagnosis of the Nitrogen Status in Crops, edited by G. Lemaire (Springer, Berlin), pp. 201–227.

Benoist, D., Tourbier Y., Germain-Tourbier, S., 1994, Plans d'expériences : construction et analyse (Tec & Doc, Lavoisier, Paris), 700 pp.

Berger, M., Moreno, J., Mueller, A., Beisl, U., Richter, R., Schaepman, M., Strub, G., Stoll, M. P., Nerry, F., Leroy, M., Rast, M., Wursteisen, P., Attema, E., 2001, The DAISEX campaigns in support of a future land surface processes mission. *ESA Bulletin*, accepted for print.

Campbell, G. S., 1990, Derivation of an angle density function for canopies with ellipsoidal leaf angle distribution. *Agricultural and Forest Meteorology*, **49**, 173-176.

Curran, P. J., 2001, Imaging spectrometry for the study of vegetation canopies. In Proceedings of the International Workshop on Spectroscopy Application in Precision Farming, Freising-Weihenstephan, Germany, on January 16-18 2001, pp. 7-8.

Jacquemoud, S., Baret, F., 1990, PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, **34**, 75-91.

Jacquemoud, S., Baret, F., Hanocq, J.-F., 1992, Modeling spectral and bidirectional soil reflectance. *Remote Sensing of Environment*, **41(2-3)**, 123-132.

Kimes, D. S., Knyazikhin, Y., Privette, J. L., Abuelgasim, A. A., Gao, F., 2000, Inversion methods for physically-based models. *Remote Sensing Reviews*, **18**, 381-439.

Moreno, J., Casalles, V., Martinez-Lozano, J. A., Melia, J., Sobrino, J., Calera, A., Montero, F., Cisneros, J. M., 2001, The measurement programme at Barrax. In Proceedings of the Final Results Workshop on DAISEX, ESTEC, Noordwijk, the Netherlands, on March 15-16 2001, pp. 43-52.

Müller, A., Hausold, A., 2001, The airborne imaging spectrometer data acquisition programme in 1998, 1999, 2000. In Proceedings of the Final Results Workshop on DAISEX, ESTEC, Noordwijk, the Netherlands, on March 15-16 2001, pp. 7-12.

Price, J. C., 1975, Information content of Iris spectra. *Journal of Geophysical Research*, **80(15)**, 1930-1936.

Price, J. C., 1992, Variability of high-resolution crop reflectance spectra. *International Journal of Remote Sensing*, **13(14)**, 2593-2610.

Verhoef, W., 1984, Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sensing of Environment*, **16**, 125-141.

Verhoef, W., 1985, Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment*, **17**, 165-178.