

Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data

Marie WEISS^{a*}, Frédéric BARET^a, Ranga B. MYNENI^b, Agnès PRAGNÈRE^a, Youri KNYAZIKHIN^b

^a INRA Bioclimatologie, Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France

^b Department of Geography, Boston University, Massachusetts, USA

(Received 6 April 1999; accepted 11 May 1999)

Abstract – The objective of this study was to assess the ability to estimate canopy biophysical variables from remote sensing data observed at the top of the canopy in several directions and wavebands within the visible-near infrared domain. The variables considered were the leaf area index, leaf chlorophyll content, the fraction of photosynthetically active radiation absorbed by the canopy and the cover fraction. The SAIL radiative transfer model was inverted using a simple technique based on look-up-tables. The size of the look-up-table, and the number of its elements selected to get a distribution of the solution were first determined. The nadir reflectance in the red and near-infrared bands was considered to evaluate the retrieval performances in terms of the distributions and co-distributions of the solutions. The optimal spectral and directional sampling to estimate the variables considered was investigated. Finally, the impact of spatial heterogeneity on the retrieval performances, the effect of the model assumptions used to generate the look-up table and the effect of radiometric noise were evaluated. These results were discussed in view of the definition of future satellites and the selection of the best measurement configuration for accurate estimation of canopy characteristics.

Remote Sensing / model inversion / look-up table / optimal sampling / biophysical variable

Résumé – Évaluation d'une technique d'inversion pour l'estimation des variables biophysiques de cultures à partir de données de réflectance spectrale et directionnelle. L'objectif de cette étude était d'évaluer l'intérêt des données de télédétection pour l'estimation des variables biophysiques des couverts végétaux. Les données considérées étaient acquises au niveau du couvert dans plusieurs directions, pour différentes longueurs d'onde du domaine du visible et du proche infrarouge. Les variables estimées étaient l'indice foliaire, le contenu en chlorophylle des feuilles, la fraction de rayonnement photosynthétiquement actif absorbé par le couvert, et le pourcentage de couverture du sol. Une technique simple d'inversion de données basée sur des tables de correspondance a été utilisée. Dans un premier temps, la taille de la table et le nombre d'éléments sélectionnés pour obtenir une distribution de la solution ont été

Communicated by Gérard Guyot (Avignon, France)

* Correspondence and reprints
weiss@avignon.inra.fr

déterminés. Les performances d'estimation des variables en termes de distributions et de co-distributions des solutions ont été évaluées dans le cas le plus simple en utilisant les réflectances rouge et proche infrarouge au nadir. L'échantillonnage spectral et directionnel optimal correspondant à la meilleure précision d'estimation de chaque variable considérée a également été déterminé. Enfin, les effets de l'hétérogénéité spatiale, des hypothèses du modèle de transfert radiatif utilisé pour générer la table de correspondance, ainsi que du bruit radiométrique, sur l'estimation des variables ont été quantifiés. Ces résultats sont discutés en vue de la définition des caractéristiques des capteurs à venir pour obtenir des estimations précises des caractéristiques des couverts végétaux.

Téledétection / inversion de modèle / table de correspondance / échantillonnage optimal / variable biophysique

1. Introduction

The continental biosphere is one of the main components of the Earth's system since it controls the energy and mass exchange at the soil-vegetation-atmosphere interface. The characterisation of land surfaces is thus a critical issue, for which satellite observations can largely contribute.

The accuracy of canopy characteristics estimation from satellite data depends on the nature of the radiometric data, i.e. the spectral and directional information content of the data acquired. Numerous studies have thus investigated the potential of many instruments to extract canopy biophysical variables in the visible [e.g. 5, 8, 15], thermal infrared [e.g. 40] and microwave domains [e.g. 33, 46]. These variables must be both strongly coupled with the radiative transfer in the canopy and pertinent with regards to process models describing ecosystem functioning and its relations to other components of the global system. Apart from the albedo which was investigated in many studies [e.g. 27, 31], four key variables can be considered in the visible-near infrared domain:

- The green leaf area index, LAI , that defines the size of the interface for exchange of energy and mass between the canopy and the atmosphere;
- The daily fraction of photosynthetically active radiation absorbed by the canopy, $fAPAR$, that allows to derive the biomass production from simple primary productivity models [38];
- The leaf chlorophyll content, C_{ab} , that can be used as an indicator of the nitrogen content [41];
- The cover fraction, $fCover$, that corresponds to the gap fraction in the nadir direction. It is used to

decouple vegetation and soil contribution in energy balance processes with particular attention to evapotranspiration.

Both $fAPAR$ and $fCover$ are secondary variables, i.e. they are combinations of primary variables such as LAI and C_{ab} that are state variables of the canopy and describe the structure or the optical properties of the elements.

Empirical approaches based on spectral indices have been developed since the early exploitation of satellite data to estimate canopy characteristics from remote sensing data. However, they are suffering from severe limitations due to the lack of physics introduced in the retrieval technique and the small amount of radiometric information they can exploit. Alternative approaches based on radiative transfer model inversion techniques have been recently investigated, in particular to exploit the richer data provided by the new generation of sensors such as POLDER, VEGETATION, MODIS, MISR and MERIS.

The accuracy of the biophysical variable estimation from satellite data mainly depends on the quantity and quality of the information used:

- The quantity of radiometric information, which resumes to the sampling of the spectral, directional and spatial dimensions;
- The quality of the radiometric information that is mainly driven by the radiometric resolution, including noise and inaccuracies coming from radiometric, spectral, spatial, and directional domains. Knowledge of the structure of the inaccuracies is thus mandatory to better control the inversion process;
- The quality of the radiative transfer model used, i.e. its pertinence with regards to the type of

canopy it will apply to, with particular attention on the way canopy structure is represented and the way the physical problem of the radiative transfer is solved;

- The a priori knowledge of the distribution of some canopy characteristics that will better constrain the inversion process. This could be partly derived from knowledge of the type of vegetation observed;

- The inversion technique used. Two groups of inversion techniques can be defined. The first one includes iterative optimisation techniques [16, 21, 25] and/or the use of simulated Look-Up Tables (LUT) [23, 24]. They both consist in minimising a distance between the simulated and the measured reflectance. These two methods are multivariate since the solution corresponds to the whole set of primary input variables of the model. The secondary variables can be then derived in a second step. The second group of inversion techniques is based on empirical relationships developed over radiative transfer model simulation and using spectral indices [3, 29] or neural networks [6, 22, 45]. They both consist, during the calibration step, in minimising a distance between the simulated and actual variable of interest by adjusting the coefficients of the relationship between the reflectance and the variable. Conversely to the previous approaches, these latter are able to focus on one single variable (primary or secondary) of interest. One advantage of these techniques is their low computer resources requirement when calibrated and used in routine applications.

The objective of this study was to evaluate the performances of canopy biophysical variable estimation from remote sensing data. The four key biophysical variables LAI , C_{ab} , $fAPAR$ and $fCover$ will be considered. Previous studies [7, 32] have shown that neural networks and LUT were generally performing the best. We used here the LUT method since the learning phase of neural networks requires larger computer time and thus becomes difficult to handle when repeated several times as we did. The study was restricted to top of canopy reflectance data, and did not address the atmospheric correction problem. Further, for simplicity,

the sun position was set to 45° zenith angle and no diffuse radiation was considered.

The inversion process and the associated performances in terms of the distribution and co-distribution of the solutions were first analysed using the simple case where only red and near infrared bands and nadir observations were considered. The optimal measurement configuration was then defined in terms of the set bands and view directions that gave the better estimation performances. Finally, the effect of the pixel heterogeneity, the impact of radiometric noise and that of the radiative transfer model assumptions used to generate the look-up table were evaluated.

2. Data sets and methods

The look up table is one of the simplest technique used to invert models. It consists first in generating a table in \mathfrak{R}^p , the space of canopy realisation that contains p input variables of the radiative transfer model. Then, the radiative transfer model is used to generate the corresponding reflectance table in \mathfrak{R}^n , where n is the number of measurement configuration, i.e. the number of bands and directions considered. In the application phase, for a given reflectance measurements in the n bands and directions considered, the elements of the reflectance table in \mathfrak{R}^n that are the closest to the measurement are selected, and the corresponding set of input variables in \mathfrak{R}^p represents the solution of the problem. In this study, the distance criterion, that corresponds to the cost function in optimisation techniques, was defined as the relative root mean square error ($RMSE^*$) between the measured reflectance ρ and the estimated reflectance $\hat{\rho}$ found in the look-up-table:

$$RMSE^* = \sqrt{\frac{1}{n_d \cdot n_b} \sum_{i=1}^{n_d} \sum_{j=1}^{n_b} \left(\frac{\rho_{ij} - \hat{\rho}_{ij}}{\rho_{ij}} \right)^2} \quad (1)$$

where n_d was the number of viewing directions and n_b the number of wavebands. The relative $RMSE^*$ was preferred to the use of the classical $RMSE$ since it did not put the emphasis on bands or direc-

tions that have the largest absolute reflectance values. We should note also here that we did not impose particular constraints on the input variables to be estimated such as often done in iterative optimisation techniques [36].

The LUT technique is defined by:

- The radiative transfer model used to derive the reflectance table from the table of input variables;
- The way the space of canopy realisation was sampled. It is driven by the distribution and co-distribution laws used for each input variable, and by the number of cases (sets of input variables) considered that defines the size of both the input variable and reflectance tables;
- The way the solution was extracted from the table, i.e. how and how many cases were selected from the reflectance table (and after, from the corresponding input variable table) to get a distribution of the possible solutions. This was investigated as a preliminary step before describing the actual results.

These issues will be first discussed, before describing the data sets used to evaluate the LUT technique.

2.1. Radiative transfer model

The choice of the radiative transfer model must satisfy at least two constraints:

- it must allow a fair representation of the radiative transfer in the canopy, both from the physics and from the description of canopy architecture;
- it must also be associated with a rather limited number of input variables and small computational requirements to facilitate the study.

We have chosen a simple turbid medium radiative transfer model, SAIL, [42, 43] that offers a good compromise between the computation time, the realism of the simulations and the limited number of input variables associated with the degree of complexity of canopy architecture description. The hot spot correction [26] was implemented to better describe the bi-directional variation of the reflectance. We used an ellipsoidal leaf angle dis-

tribution function [10], which presents the advantage to be a continuous function. The SAIL model has been validated over a large range of crops such as soybean [4, 17], orchards [18], maize [28] or sugar beet [1].

The canopy structure was described by the leaf area index (LAI), the average leaf inclination angle ALA , and the hot spot parameter that represents the apparent size of the leaves. The leaf optical properties were simulated thanks to the PROSPECT model [14, 19], assuming that the leaf was bi-lambertian. The input variables of the PROSPECT model are the leaf chlorophyll, dry matter and water contents (C_{ab} , C_{dm} and C_w) and a structure parameter N . The dry matter and water contents were tied together, assuming a constant relative water content of 80%, which is generally observed for green active leaves. Departure from this situation will not affect seriously the simulations since water and dry matter contribute only marginally to the absorption in the visible and near infrared domains. The soil was assumed to be lambertian, with a typical soil reflectance spectrum, times a brightness parameter b_s .

2.2. Sampling the space of canopy realisation

The space of canopy realisation was generated by randomly drawing each radiative transfer model input variables within distribution laws. The input variables were assumed independent because no information was available about their possible covariance. The distribution law of each variable was selected so that the density of probability was set to be proportional to the sensitivity of the reflectance to the variable considered. It allowed to better sample domains where the reflectance is more sensitive to the considered variables. This was achieved by applying transformations F for each variable x such as the sensitivity of the reflectance to the transformed variable was about constant: $dp/dF(x) \approx cste$. The transformations were selected using a trial and error process applied to the simple case of red and near infrared nadir observations. Then, uniform random drawing were completed over the transformed variables.

Table I. Transformation used to generate the distribution of the variables. The distribution laws of the transformed variables were uniform.

	SAIL Variables	Units	Minimum	Maximum	Transformed variables
LAI	Leaf area index	Unitless	0	8	$e^{-0.5LAI}$
ALA	Average leaf angle	$^{\circ}$	20	75	$\cos(ALA)$
h	Hot-spot parameter	Unitless	0.05	1	e^{-3h}
C_{ab}	Leaf chlorophyll content	$\mu\text{g}\cdot\text{cm}^{-2}$	20	100	$e^{-0.01C_{ab}}$
C_w	Leaf water content	cm^{-1}	0.005	0.025	e^{-50C_w}
N	Leaf structure index	Unitless	1	2.5	N
b_s	Soil brightness	Unitless	0.5	1.5	b_s

Table I presents the distribution laws finally selected. One important consequence of this process is that low LAI , C_{ab} , C_w and hot spot values were more densely sampled. A total of 280000 simulations were generated using Monte Carlo drawings.

2.3. Test data sets

Two data sets, combining top of canopy bi-directional reflectance and corresponding biophysical variables were used to evaluate the look-up-table retrieval performances.

First, a reference data set (called 3D) was generated by applying the PARCINOPY ray tracing code [11] to a 3D model of maize canopy architecture [13]. The combination of these models allowed to get very realistic reflectance simulations both from the physics of the radiative transfer and from the canopy structure description. Each simulation provided the bi-directional reflectance values covering the whole hemisphere by 5° zenith and azimuth steps, for a given waveband. Small noise associated to the Monte-Carlo process was observed (relative noise of 2.5%), although 3 million rays were traced.

The second test data set (called 1D) was generated with the SAIL model. This made the 1D test data set fully consistent with the look-up-table. To be able to compare results obtained on 1D and 3D data sets, the 1D data set was also perturbed with 2.5% relative Gaussian noise.

Eighteen cases, corresponding to six development stages and variation in leaf chlorophyll content and soil type were considered for the two test data sets (Tab. II).

Since ray tracing technique used for the 3D test case required large computer resources, a limited set of wavebands were simulated in the visible and near infrared domains which is the more relevant for LAI , leaf chlorophyll content, canopy structure and $fAPAR$ estimation [2]. Nine wavebands, 10 nm width, were considered: 500, 562, 630, 692, 710, 740, 795, 845, and 882 nm.

The directional sampling was reduced to 32 view directions, assuming a solar zenith angle at 45° , with no diffuse radiation (Fig. 1). Considering

Table II. The 18 combinations of LAI , leaf chlorophyll content and soil type used to generate the two test data sets. For the 18 cases, $ALA = 56^{\circ}$, $h = 0.1$, $C_w = 0.014 \text{ cm}^{-1}$, $N = 1.4$.

LAI	Dry soil, $b_s = 1.4$		Wet soil $b_s = 0.6$	
	C_{ab}	C_{ab}	C_{ab}	C_{ab}
0.25	30	50	70	50
0.86	-	50	-	-
1.64	30	50	70	50
2.34	-	50	-	-
3.01	30	50	70	50
6.25	30	50	70	50

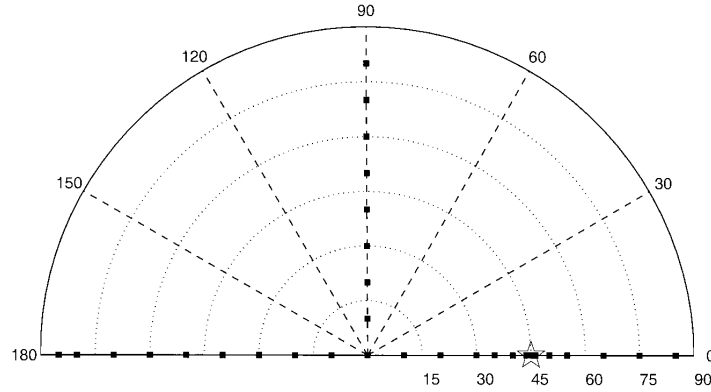


Figure 1. Polar representation of the directional sampling (■ 32 directions) used in this study. The sun direction is indicated by a ☆ and is set at 45° zenith angle.

the symmetry with respect to the solar plane, only half the hemisphere was investigated. A particular attention was given to the principal plane, and the sampling was refined near the solar position, since the large reflectance variation are observed around

the hot spot [37]. Some regularly spaced points in the perpendicular plane were added, following the results of Roujean and Bréon [39] who estimated $fAPAR$ using a bi-directional vegetation index.

Figure 2 shows the difference between the reflectance values simulated with the two models (1D and 3D), for all wavelengths, all directions of

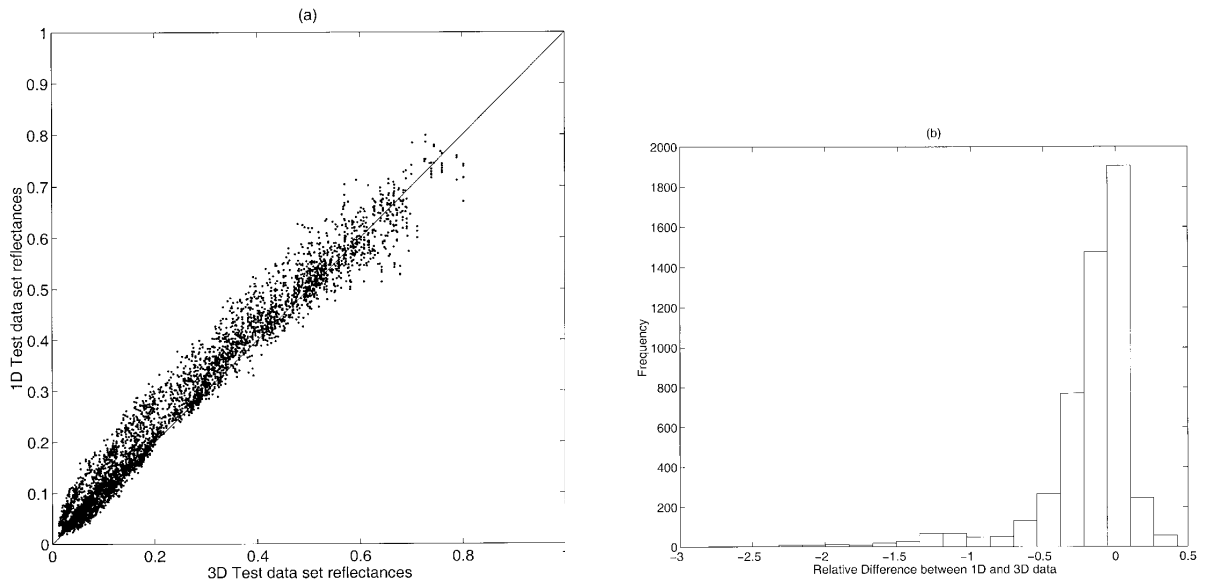


Figure 2. (a) Comparison between 1D and 3D data set reflectance values (18 cases, 9 wavelengths, 32 directions). (b) Distribution of relative difference between 1D and 3D data set reflectance values.

the 18 maize canopies. As can be noticed, there is no particular behaviour of the difference between the two models, depending on view directions and wavelengths. This is confirmed by the distribution of the relative difference between the 1D and 3D reflectance values, which can be assimilated to a Gaussian law.

3. Results

3.1. Optimal LUT size and number of cases selected for the solution

The LUT was defined by the distribution laws presented in the previous section and by the number of cases simulated defining the size of the LUT. Increasing the LUT size would result in a better sampling while requiring larger computer resources. Therefore, we have first to trade off by investigating the effect of the size of the LUT on the accuracy of the retrieval of canopy biophysical variables. The *LAI* was only considered here, used as an example for the four biophysical variables of interest. We also have to define how many cases will be selected from the LUT and considered to be the solution.

For sake of simplicity, this preliminary study was performed on a limited basis using the 1D test data set and considering only the nadir viewing direction in a red (630 nm) and a near-infrared (882 nm) bands. This corresponds to the minimum sampling necessary to provide estimates of *LAI*, such as the one provided by NOAA/AVHRR sensors [3, 30]. Four look-up tables with increasing size $s = 25\,000$, $50\,000$, $100\,000$ and $280\,000$ were derived from the $280\,000$ initial simulations.

For each of the 18 cases considered in the test data set, the q LUT elements having the closest reflectance values according to equation (1) were selected. The median *LAI* value was computed over the corresponding table of input variable and was then considered as the estimated *LAI* value. Then, the absolute RMSE value between the actual *LAI* values of the test data set and the estimated *LAI* value was computed. Finally, this process was

iterated for the number of cases, q , ranging from 1 to the total size, s , of the LUT.

Figure 3 shows that for the fraction of selected cases (q/s) lower than 0.5%, the effect of the size of the look up table was quite important for the accuracy of *LAI* estimation. In these conditions, the absolute *RMSE* for *LAI* was the highest for a LUT size smaller than $s = 50\,000$. The two highest LUT sizes gave similar low *RMSE* values. We thus decided to use LUT tables made of $s = 100\,000$ cases that provided a good compromise between the computer resources requirement and the accuracy of canopy variable estimates. It can also be noticed that, for q/s higher than 0.5% and whatever the LUT size was, the *RMSE* was increasing with the percentage of selected cases. This was due to the fact that the number of selected cases became too high, and thus the distribution of the solutions

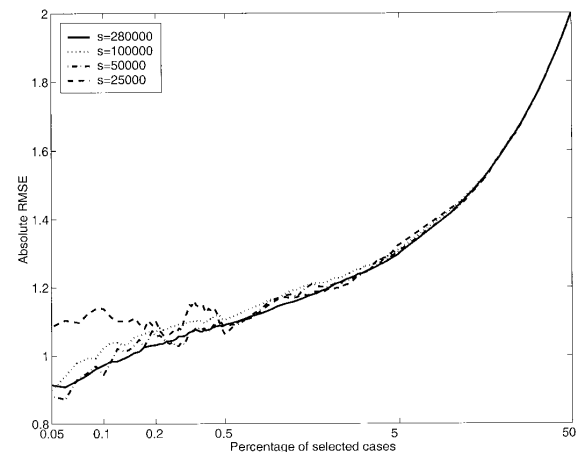


Figure 3. Absolute *RMSE* value between actual and estimated *LAI* for the 1D test data set as a function of the percentage of selected cases in the LUT ($q/s \cdot 100$). The estimation was performed for a range of LUT size: ---- $s=25\,000$, -.-. $s=50\,000$, ... $s=100\,000$, — $s=280\,000$ elements.

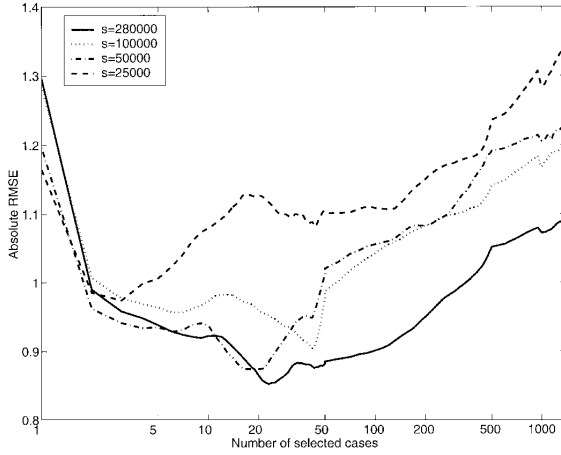


Figure 4. Absolute *RMSE* value (averaged over a 10 cases width moving window) between actual and estimated *LAI* values for the 1D test data set as a function of the number of selected cases in the look-up-tables (q). The estimation was performed for a range of LUT size:

---- $s=25\,000$, -.-. $s=50\,000$, ... $s=100\,000$, — $s=280\,000$.

in terms of *LAI* was too wide to obtain good estimates (median value).

By focusing on the lowest part of Figure 3 ($q < 1000$), i.e. the closest to the solution, results show (Fig. 4) that the *RMSE* values on *LAI* gets a minimum value for $20 < q < 50$, except for the smallest LUT size ($s = 25\,000$). We observed that for the lowest values of q , the *RMSE* values increased significantly, indicating that the solution for *LAI* estimation was not the best one although corresponding to a closer match in the reflectance domain. This illustrated some possible degree of hyper-specialisation of the process. We therefore decided that the 50 best cases in the reflectance LUT ($q = 50$) would be considered as the solution. This solution was thus characterised by a distribution of canopy variables values.

3.2. Retrieval performances from nadir reflectance in the red and near infrared for *LAI*, *fAPAR*, *fCover* and C_{ab}

Before considering a larger number of bands and directions to address the spectral and directional sampling problem, we first analysed the results obtained in the simple case of nadir viewing in the red and near infrared bands using the 1D test data set. For each of the primary biophysical variables and the two associated secondary ones (*fAPAR*, *fCover*), we computed the distribution of the estimated values over the 50 selected cases corresponding to the 50 cases in the reflectance table ($s = 100\,000$) that were the closest to the reflectance values in the test data set.

The results were illustrated mainly using 2 cases in the test data set corresponding to low ($LAI = 0.25$) and high *LAI* values ($LAI = 6.25$) for $C_{ab} = 50\,\mu\text{g}\cdot\text{cm}^{-2}$ and dry soil conditions. However, similar trends were observed for the other test data set cases not shown.

Results showed that for low *LAI* values, the *LAI* was accurately estimated, with a distribution well centred on the actual *LAI* value and a small standard deviation (Tab. III, Fig. 5). When *LAI* was increasing, the standard deviation of the distribution of the estimated *LAI* values increased. Further, for large *LAI* values, the distribution was shifted towards lower *LAI* values because of the saturation and the distribution law used for generating the input variable table that over-sampled the lower *LAI* values. The average leaf inclination angle was generally poorly estimated (high standard deviation), with however a median value close to the actual value. The leaf chlorophyll content was also poorly estimated with median values often quite different from the actual value, and large standard deviation. This is explained by the insufficient information provided by the two bands and the nadir direction to accurately retrieve leaf chlorophyll content. This was also illustrated by the co-distributions of the variables: high correlations were observed between C_{ab} and respectively *LAI*, *ALA* and *N* (Fig. 6). It is interesting to notice that the correlation coefficient between C_{ab} and *LAI* is negative, i.e. an increase of *LAI* is compensated by

Table III. Primary and secondary SAIL variable estimates on the 1D test data: actual, estimated value and standard deviation computed on the 50 cases selected in the LUT. Three treatments are presented, with three LAI values for a leaf chlorophyll content of $50 \mu\text{g}\cdot\text{cm}^{-2}$ and a dry soil.

		LAI = 0.25	LAI = 1.64	LAI = 6.25
LAI	Actual.	0.25	1.6	6.2
	Median	0.29	1.7	3.5
	Std	0.09	1	1.1
ALA	Actual.	56	56	56
	Median	55	51	45
	Std	13	15	13
h	Actual.	0.1	0.1	0.1
	Median	0.3	0.22	0.27
	Std	0.21	0.19	0.26
C_{ab}	Actual.	50	50	50
	Median	43	52	78
	Std	22	21	14
C_w	Actual.	0.014	0.014	0.014
	Median	0.015	0.013	0.013
	Std	0.006	0.006	0.005
N	Actual.	1.4	1.4	1.4
	Median	1.9	1.5	1.7
	Std	0.45	0.39	0.39
b_s	Actual.	1.4	1.4	1.4
	Median	1.4	1.1	0.95
	Std	0.08	0.26	0.3
$fAPAR$	Actual.	0.21	0.76	0.99
	Median	0.24	0.76	0.95
	Std	0.06	0.11	0.04
$fCover$	Actual.	0.13	0.58	0.96
	Median	0.14	0.6	0.89
	Std	0.029	0.11	0.06
RMSE	Actual	0.016	0.01	0.01
	std	0.003	0.003	0.004

a decrease of C_{ab} to achieve the same radiometric response. Conversely, the correlation coefficient between LAI and ALA is positive, i.e. an increase

of LAI is compensated by an increase of ALA to achieve the same radiometric response. Retrieval performances for the soil brightness were quite good for low LAI values, with however some correlation with both ALA and LAI . Conversely, it appeared impossible to retrieve the soil optical properties when saturation occurs for high LAI values, which can be easily understood. As only the nadir direction was considered, estimation of the hot spot parameter was unsatisfactory. In the same way, the leaf water content, and thus the leaf dry matter content, were not well estimated since the middle infrared domain where both leaf constituents absorb strongly was not used here. The retrieval performances for the leaf structure parameter N , were also unsatisfactory because the effect of this variable spans very similarly over the whole spectral domain and there are thus little clues to get accurate estimates using only the spectral variation.

The secondary variables, $fAPAR$ and $fCover$, which are particular combinations of LAI and ALA , were better evaluated since they partly account for the compensations between LAI and ALA (Fig. 6). Further, these secondary variables represent the canopy gap fraction in particular directions, which is one of the main driver of canopy radiative transfer.

We therefore concluded that the amount of information corresponding to this measurement configuration was too limited to provide accurate estimates of some key primary biophysical variables such as LAI and C_{ab} . One way to improve the estimation would be to impose some constraints on the cost function (Eq. (1)) using a priori information on the distribution of the probable input variable values that would depend on canopy type. The other way that was investigated in the next section consists in increasing the amount of radiometric information by optimising the spectral and directional sampling. For sake of simplicity, we will first address the optimal spectral sampling. We will then address the optimal directional sampling based on the optimal set of bands determined in the previous step.

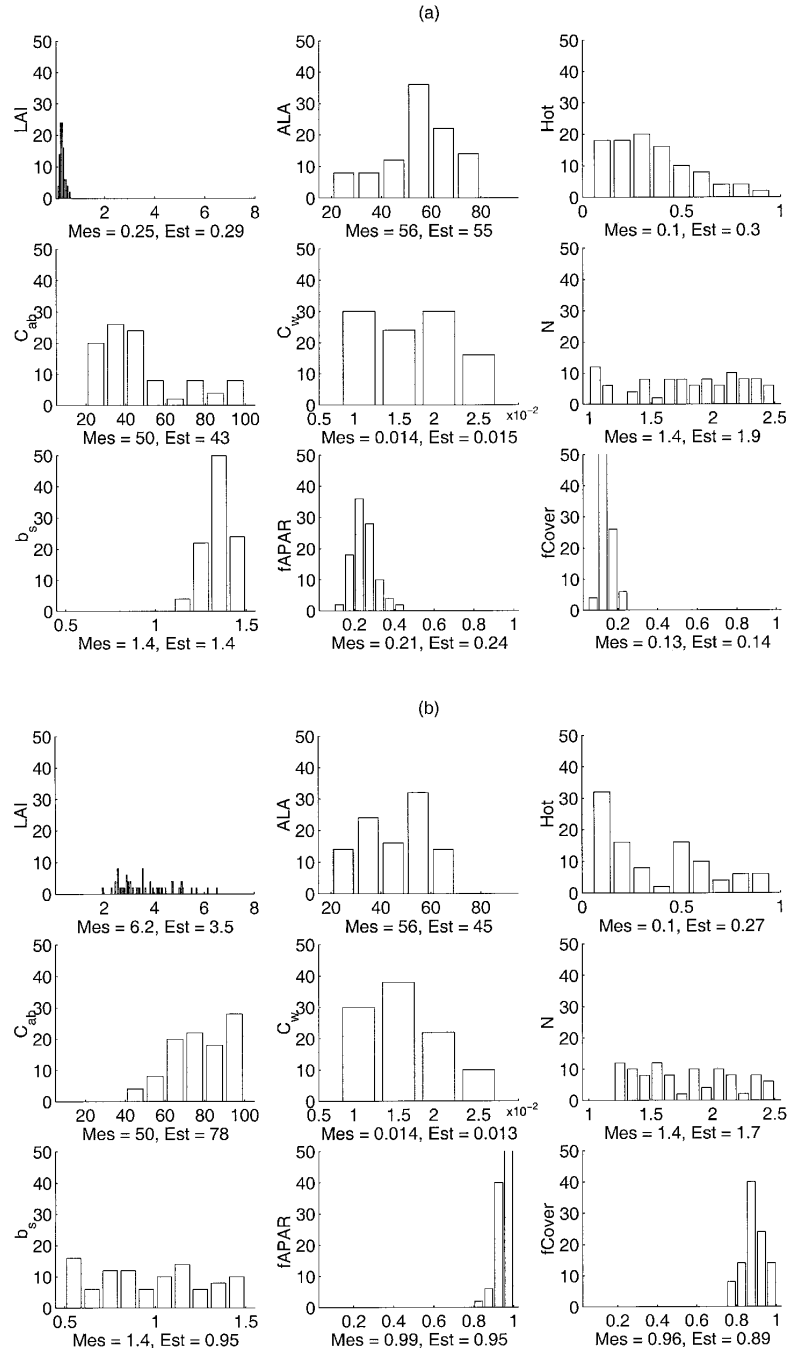


Figure 5. Distribution of the primary SAIL input variable estimates on the 1D test data set. Two examples are presented: $LAI = 0.25$ and $LAI = 6.25$, leaf chlorophyll content $C_{ab} = 50 \mu\text{g}\cdot\text{cm}^{-2}$ and dry soil, $b_s = 1.4$.

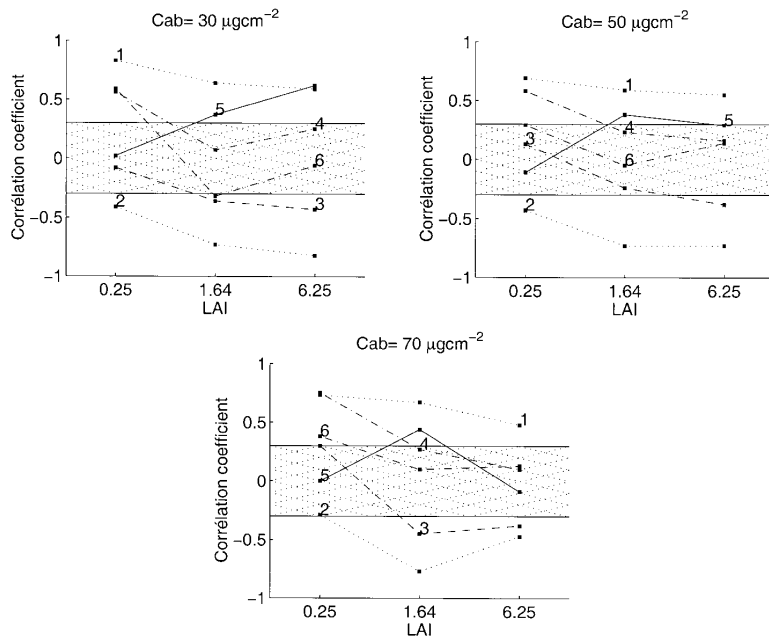


Figure 6. Correlation coefficients between estimates of SAIL input variables (1: LAI/ALA , 2: LAI/C_{ab} , 3: C_{ab}/ALA , 4: ALA/b_s , 5: C_{ab}/N , 6: LAI/b_s) for different LAI and configuration. These correlations were computed over the 50 selected cases of the LUT. Values were thus significant when the correlation coefficient was higher than 0.3 ($\alpha = 0.03$).

3.3. Optimal spectral sampling

Many studies demonstrated the interest of multi-spectral and even hyper-spectral data [2, 12, 20]. However, Price showed that only a limited number of bands were necessary to describe the whole reflectance spectrum of soil [34] or to discriminate water, snow, fire, vegetation and soil [35]. In a same way, Weiss *et al.* [44] could estimate canopy albedo using 5 bands with a relative accuracy better than 2%. Our objective was thus to determine how many, and which bands, among the 9 available, were necessary for accurately estimate each biophysical variable of interest when just using nadir observations. The directional sampling was investigated in the next section.

Starting from the initial combination of red and near-infrared bands, we selected the band, among the 7 remaining bands that provided the best estimation performances evaluated by the absolute RMSE computed for the canopy variable considered over the 1D test data set. The same process was repeated until all the nine bands were selected (Fig. 7).

Results showed that for all the four variables considered, when the number of bands used increased, the absolute RMSE value was decreasing until reaching a minimum, and then started to increase with however a moderate slope. This demonstrated that only a limited number of wavebands were required for canopy biophysical variable estimation. Once the optimal number of bands was reached, extra bands added some noise without adding significant information on canopy characteristics. The estimation of the secondary variables required less bands (4 bands for $fAPAR$ and 5 for $fCover$) than for the primary variables (6 bands for LAI and C_{ab}). The RMSE value difference computed over the biophysical variables of interest between the best and the poorest band selection demonstrated that the choice of the bands was obviously important, particularly for the first bands selected (Fig. 7). Further, the third selected band was always the green one, confirming that it is highly recommended for vegetation applications. The other bands were generally selected in the red-edge domain (680-740 nm) where the dynamics of the vegetation spectrum is the most important. For the estimation of LAI , $fCover$ and $fAPAR$, only one

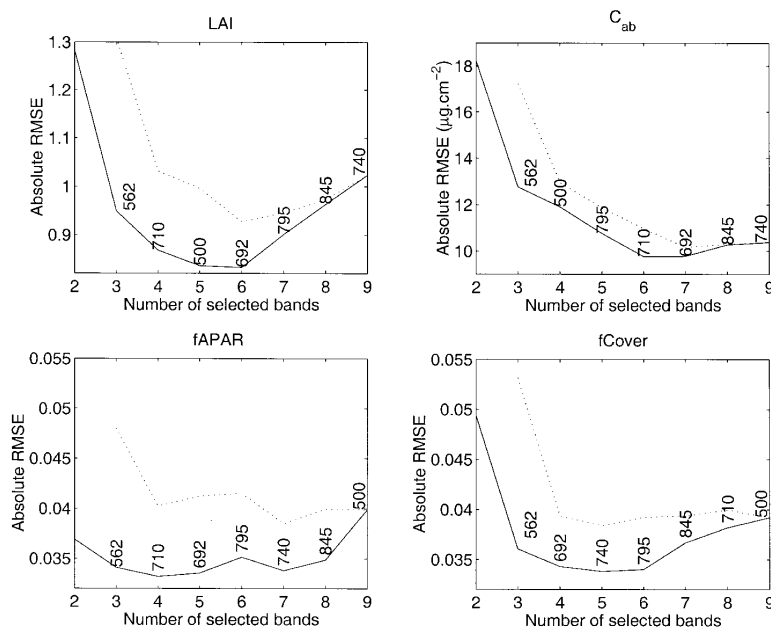


Figure 7. Absolute *RMSE* value between actual and estimated canopy variables considering the 18 cases of the 1D test data set as a function of the number of bands selected for each canopy biophysical variable of interest. The solid line corresponds to the best band combination and the dotted line to the worst combination.

band in the near infrared plateau was required. Surprisingly, the band at the middle of the red-edge (710 nm) was not one of the first bands needed for the estimation of leaf chlorophyll content. However, when considering the *RMSE* computed during the band selection (Fig. 8), the *RMSE* obtained by adding the 710 nm band was always very close from the actual selected band.

As a conclusion of this step, we decided to keep the band combination that gave the lowest *RMSE* for each variable: *LAI* (630, 882, 562, 710, 500, 692 nm), C_{ab} (630, 882, 562, 500, 795, 710 nm), *fAPAR* (630, 882, 562, 710), *fCover* (630, 882, 562, 692, 740). For these optimal band sets, we will investigate the optimal directional sampling.

3.4. Optimal directional sampling

Considering the optimal set of bands for each variable derived previously for nadir viewing, the same procedure as the one used previously for the spectral sampling issue was applied to select the optimal set of view directions.

Results showed that for each of the four biophysical variables considered, less than six directions were required (Fig. 9). The other directions contained redundant information and thus, their use as additional bands induced noise in the retrieval process. The first selected direction was able to reduce significantly the uncertainty on the retrieval of the four biophysical variables. For this direction, the difference between the optimal and the worst view angle was large. The first directions were always located in the principal plane, either close to the hot spot direction, or in the forward direction at large zenith angles (Fig. 10).

For *LAI* estimation, only one direction, the hot spot, was required to obtain the lowest *RMSE* value, confirming the importance of that direction for canopy structure estimation. Moreover, the following selected angles were very close to the hot spot, except the second one which was located in the forward direction with high zenith angle. The leaf chlorophyll content estimation required more directions (6), mainly located in the backward and forward directions for large zenith angles, and

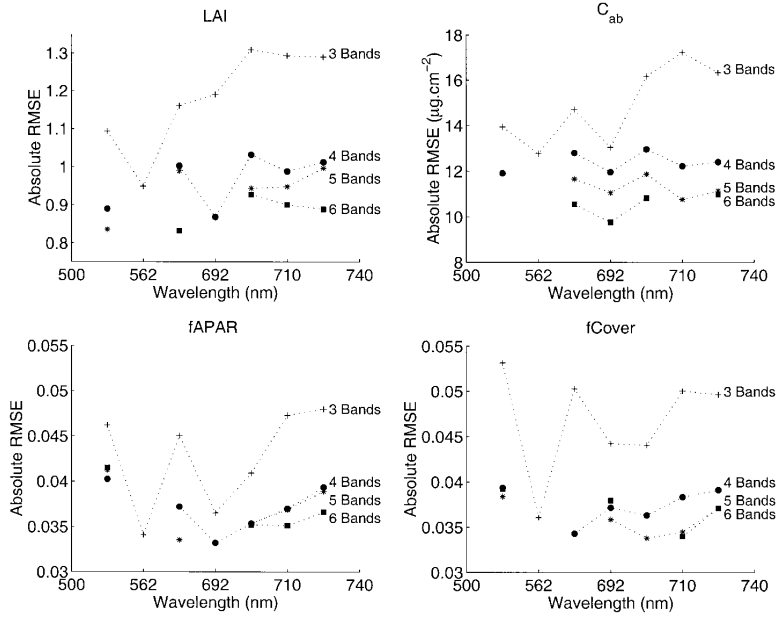


Figure 8. Selection of the best band combination (+ 3 bands, o 4 bands, * 5 bands, 6 bands): Absolute *RMSE* value between actual and estimated variables considering the set 18 cases of the 1D data, as a function of the waveband. The numbers correspond to the rank of the band selected, in addition to the two first red and near-infrared bands.

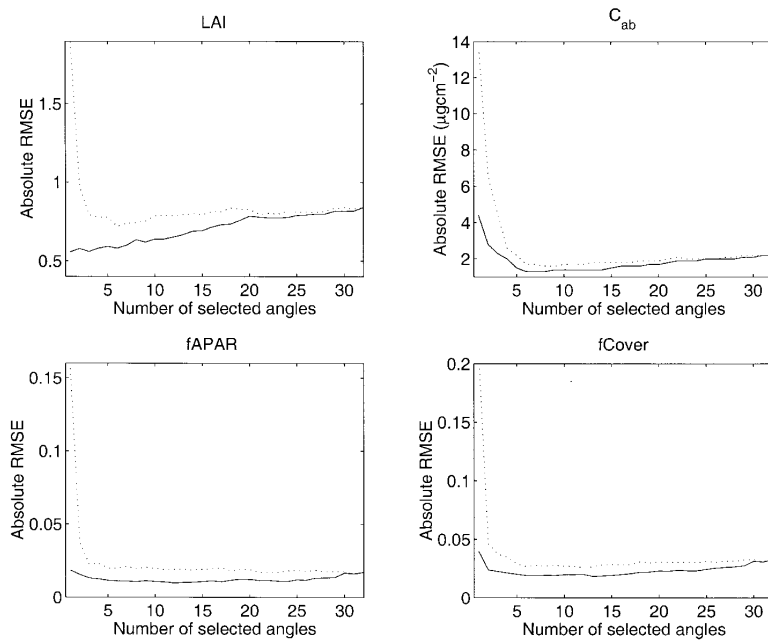


Figure 9. Absolute *RMSE* values associated to the estimation of canopy biophysical variables considering the 18 cases of the 1D test data set as a function of the number of selected directions. The solid line and the dashed line corresponds respectively to the best and worse sets of directions.

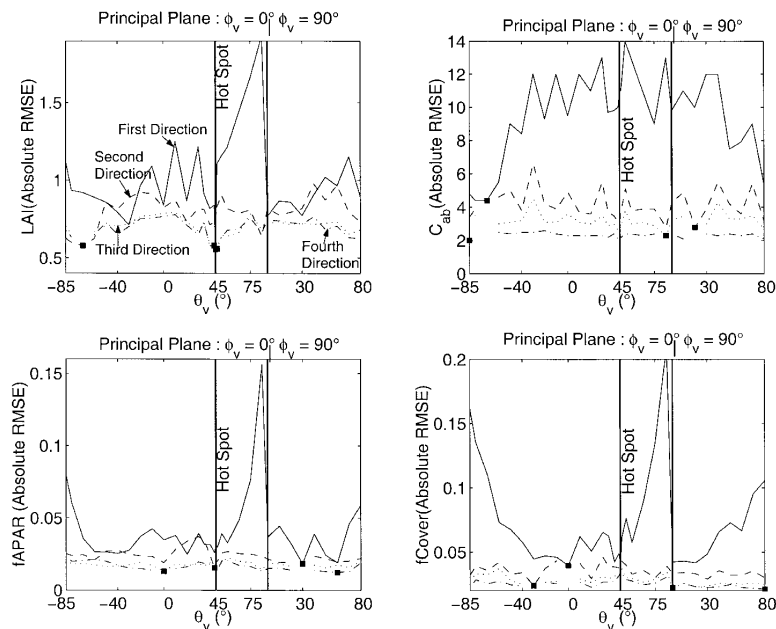


Figure 10. Selection of the five first optimal directions (■) for biophysical variable estimation: absolute *RMSE* value between actual and estimated canopy variable considering the 18 cases of the 1D test data set as a function of the view directions. Each line corresponds to an increasing number of view direction selected (solid line: one direction, dashed line: 2, dotted line: 3, dashed-dotted line: 4 and dotted line: 5). Taking the chlorophyll content as example, the first selected direction is in the principal plane, at a view zenith angle at -70° (lowest point of the solid line). The chlorophyll content is then estimated considering this direction and a second one (dashed line). The second selected direction is then located in the perpendicular plane, at a view zenith angle of 20° , etc.

around the hot spot peak. One direction in the perpendicular plane was also needed. However, for the selection of the first direction, when considering the *RMSE* computed for each of the 32 directions, some directions were performing almost as well as the selected one, especially for C_{ab} which showed low dynamics in terms of *RMSE* values (Fig. 10). These directions were always placed at large zenith angles. Considering $fAPAR$ and $fCover$, about six directions were required and only the first one was actually discriminating. Moreover, for $fCover$ estimation, the first selected direction was nadir, which was fully consistent with the fact that $fCover$ corresponds to the gap fraction in the nadir direction. For $fAPAR$, the first selected direction was located in the perpendicular plane (zenith angle of 30°) and the second one close to the hot spot direction. This was again consistent with the fact that $fAPAR$ could be approximated by the gap fraction in the solar direction.

As a conclusion of this step, directional sampling in the principal plane near the hot spot and for forward scattering directions appeared to be of

great interest for canopy biophysical variable estimation. We should notice that the nadir direction was generally not optimal for the estimation of canopy biophysical variables, except for $fCover$. We should however note that these results were limited to the 45° sun zenith angle and top of canopy reflectance data considered in this study.

3.5. Impact of spatial heterogeneity on the retrieval performances

Considering that the spatial resolution of Earth observing sensors range from about 20 m (SPOT4) to about 7 km (POLDER), the pixel size may be large enough to observe heterogeneous pixels composed of different canopies. The aim of this section was to evaluate the sensitivity of the retrieval performances on the degree of heterogeneity of the pixel.

A mixed pixel database was derived from the 18 homogeneous canopies of the 1D data set. This was achieved by dividing the 1D data sets in three

subsets corresponding to low, medium and high values of the biophysical variable considered. The reflectance and corresponding biophysical variables of these three subsets were linearly combined to simulate a larger range of heterogeneity for the mixed pixels:

$$\bar{X} = w \cdot X_i + (1 - w) \cdot X_j; w \leq 1 \text{ and } i \neq j \quad (2)$$

where w corresponded to the fraction of the mixed pixel composed of the canopy type i . A pixel heterogeneity index, H , was computed for each mixed pixels with $H = 1$ for the most heterogeneous pixel, and $H = 0$ for the homogeneous pixel:

$$H = \frac{|X_i - X_j|}{\max_X} (1 - |2w - 1|), i \neq j \quad (3)$$

where \max_X corresponds to the maximum value of the considered X variable in the test data set (Tab. II).

The LUT technique generated previously with pure pixels was applied to the mixed pixels using the optimal sets of bands and directions selected previously. Comparison between the estimated biophysical variables and the actual ones of the mixed pixels was achieved using the residual error between the actual and the estimated variables, divided by the amplitude of variation of the variable considered on the 1D test data set (Fig. 11). It allowed to compare the effect of the spatial heterogeneity between the four different variables.

Results show that for LAI, the relative residual error was increasing drastically with the pixel heterogeneity up to 45%, whatever the pixel LAI level. High residual error values for LAI retrieval were also observed for low levels of heterogeneity combined with high values of LAI. In these conditions, LAI was generally underestimated due to the strong non-linearity between canopy reflectance and LAI values. The same tendency was observed for the estimation of C_{ab} , with higher residual value level for low C_{ab} values. The estimation of

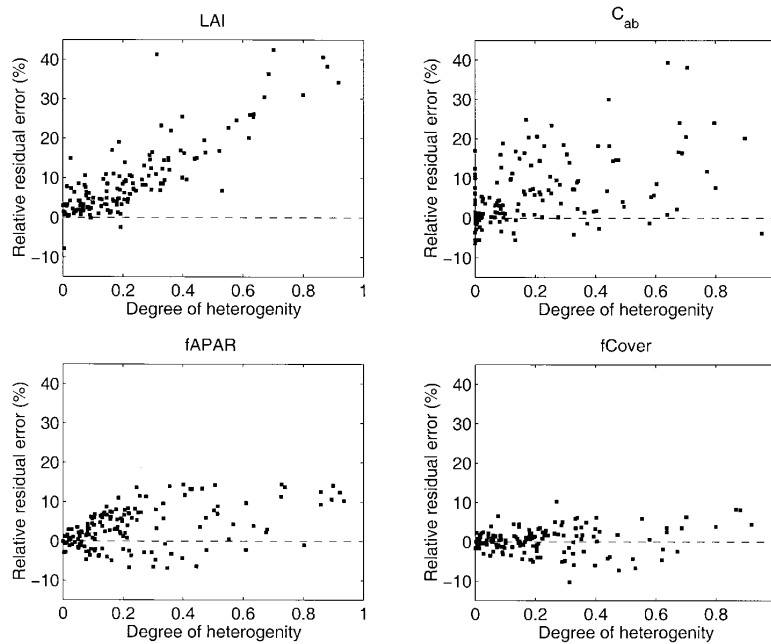


Figure 11. Relative residual error (%) between actual and estimated variable as a function of the heterogeneity index (H) of the mixed pixel.

the secondary variables ($fCover$ and $fAPAR$) was not very sensitive to the pixel heterogeneity level. Further, the relative residual error remained low ($< 15\%$) and better distributed around the x-coordinate axis, as compared to LAI and C_{ab} . This demonstrated that a pixel average value of $fCover$ and $fAPAR$ is more susceptible of being accurately retrieved from low spatial resolution sensors. These two variables can be therefore considered almost scale independent.

3.6. Impact of noise and SAIL model assumptions on the retrieval performances

The optimal spectral and directional sampling as well as the effect of the spatial resolution were determined using the 1D test data set with only little radiometric noise added. In these conditions, the test data set was fully consistent with the LUT because both data sets were generated with the same 1D SAIL model. However, both the SAIL assumptions and the inaccuracies induced by atmospheric, geometric and radiometric corrections could be limiting in case of actual remote

sensing observations. The performances of the LUT technique were thus compared in four situations: 1D and 3D models with 2.5% and 10% noise. Figure 12 showed that the effect of the noise was definitely less important than the accuracy of the model used to generate the look-up-table, whatever the biophysical variable of interest considered: the increase of the relative $RMSE$ ($RMSE$ divided by the amplitude of the variable over the test data set) was lower than 5% when considering both (1D, 2.5%) and (1D, 10%), and (3D, 2.5%) and (3D, 10%). For the same radiometric noise level, the estimates of the variables were highly sensitive to the radiative transfer model. This was observed especially for the leaf chlorophyll content, which was already poorly estimated with the 1D test data set (Tab. III).

4. Conclusion

The performances of model inversion were first evaluated in the case where only limited radiometric information was available. We showed that

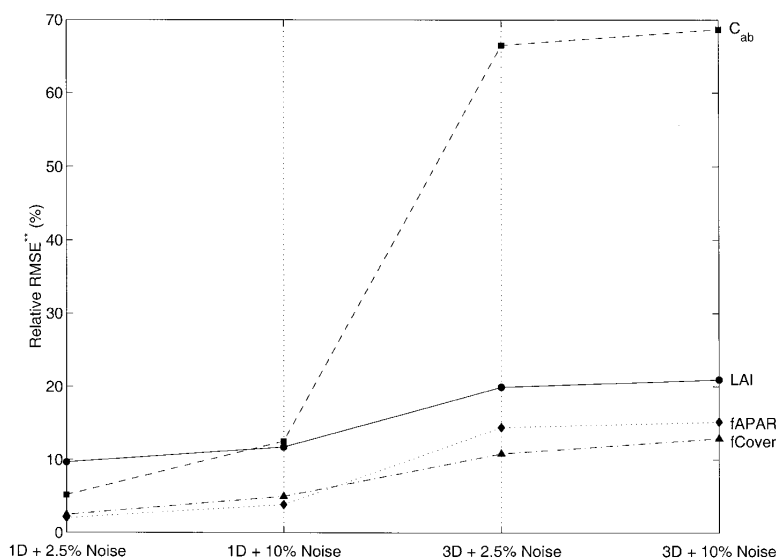


Figure 12. Relative $RMSE$ value associated to canopy variable estimation (- LAI, -- , $fAPAR$, -- $fCover$) as a function of the radiometric noise level (2.5% or 10%) and the radiative transfer model (1D or 3D) used.

many canopy variables could not be accurately estimated when using only the red and near infrared bands observed from nadir direction. This was particularly the case of LAI and more over leaf chlorophyll content. However, we observed that secondary variables such as the cover fraction and $fAPAR$ were already relatively well retrieved. This was explained by the fact that these variables were not sensitive to the saturation problem as LAI or C_{ab} were. Further the strong covariance observed between estimates of LAI and ALA making the retrieval inaccurate for these individual variables, were partly cancelled when combining LAI and ALA into the secondary variables $fCover$ and $fAPAR$.

The possible improvement in retrieval accuracy was then evaluated when the radiometric information available increases, by better sampling the spectral and directional domains. To simplify the optimisation process both aspects were considered independent, starting from the spectral issue. When increasing the number of bands used still for nadir viewing, the accuracy in the estimation of LAI and C_{ab} improved significantly, almost by a factor of 2.0. The optimal number of bands was around 6 bands. For the secondary variables $fAPAR$ and $fCover$, the accuracy of their estimation did not improve much when more than three to four bands were used. However, and particularly for LAI , the performances of the estimation degraded when all the nine bands available were used, probably because the noise induced by the bands added to the optimal bands induced more noise than useful information. Starting from the optimal sets of bands, we then defined similarly the best set of view directions. The results showed that four to seven directions generally allowed to get the best accuracy. However, the optimal view directions were located mainly in the principal plane either close to the hot spot direction, or in the forward direction for large zenith angles.

New sensors, such as POLDER are now able to provide well distributed sampling in the principal plane [9] and thus seem well adapted for canopy biophysical estimation. However, sampling for large zenith angle values will probably be associat-

ed to larger uncertainties due to atmospheric correction.

The spatial heterogeneity problem was finally investigated. It appears that the two secondary variables $fAPAR$ and $fCover$ were little sensitive to the spatial heterogeneity. We should note that these two quantities are actually particular flux physical quantities, which are known to be generally additive. Conversely, this was not the case for LAI and C_{ab} , which will be thus very sensitive to the scaling issue. The only way to improve their estimation should be to use a priori information on the heterogeneity that could be derived from higher spatial resolution satellites. This important issue should be one the main avenue of research for the use of low spatial resolution satellite data.

We should however note that these results were limited to the particular conditions of the simulations. Only nine particular wavebands in the visible to near infrared spectral domain were used, the sun zenith direction was always set to 45° , and the atmosphere was not accounted for.

The retrieval performances were evaluated on two different data sets: one generated with the same model as the one used to simulate the LUT, and one with a 3D ray-tracing model applied on a realistic 3D canopy architecture. We showed that the accuracy of the model used to generate the LUT was more important than the effect of the radiometric noise. The radiometric noise level considered was quite small, with no particular structure. The effect of the radiometric uncertainties taking into account the structure of the noise should be further evaluated.

This study focused on top of canopy reflectance data. However, the atmosphere and the canopy radiative transfer are strongly coupled. It should therefore be quite pertinent to apply the methodology developed in this study to top of atmosphere reflectance data in order to deal with the complete remote sensing problem. There is no doubt that the sampling strategy in the spectral and directional domains will be changed when the atmosphere is to be accounted for.

This study clearly demonstrated the potentials of using LUT to investigate the model inversion

problem in remote sensing. It mainly presented a methodology that has to be refined as stated earlier, but already drove original results in terms of the number of bands, directions, scaling problem, which appear to differ as a function of the biophysical variables considered.

Acknowledgements: Thanks are conveyed to the French space agency (CNES) and European space agency (ESA) for funding of this study.

References

- [1] Andrieu B., Baret F., Jacquemoud S., Malthus T., Steven M., Evaluation of an improved version of SAIL model to simulate bi-directional reflectance of sugar beet canopies, *Remote Sens. Environ.* 60 (1997) 247–257.
- [2] Asner G.P., Biophysical and biochemical sources of variability in canopy reflectance, *Remote Sens. Environ.* 64 (1998) 234–253.
- [3] Asrar G., Fuchs M., Kanemasu E.T., Hatfield J.L., Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat, *Agron. J.* 76 (1984) 300–306.
- [4] Badwahr G., Bunnik N., Verhoef W., Comparative study of Suits and SAIL canopy reflectance models, *Remote Sens. Environ.* 17 (1985) 179–195.
- [5] Baret F., Guyot G., Potentials and limits of vegetation indices for LAI and APAR assessment, *Remote Sens. Environ.* 35 (1991) 161–173.
- [6] Baret F., Clevers J., Steven M., The robustness of gap fraction estimates from red and near-infrared reflectances: a comparison of approaches, *Remote Sens. Environ.* 54 (1995) 141–151.
- [7] Baret F., Knyazikhin Y., Weiss M., Pragnère A., Myneni R.B., Overview of retrieval techniques for LAI and fAPAR, in: *Proceedings of ALPS99 workshop*, Meribel, France, 1999, pp. WK3o05.
- [8] Bicheron P., Leroy M., A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model, *Remote Sens. Environ.* 67 (1999) 251–266.
- [9] Bréon F.-M., Vanderbilt V., Leroy M., Bicheron P., Walthall C.L., Kalshoven J.E., Evidence of hot spot directional signature from airborne POLDER measurements, *IEEE Trans. Geosci. and Remote Sens.* 35 (1997) 2, pp. 479–484.
- [10] Campbell G.S., Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution, *Agric. For. Meteorol.* 36 (1985) 317–321.
- [11] Chelle M., Andrieu B., The nested radiosity model for the distribution of light within plant canopies, *Ecol. Model.*, 111(1998) 75–91.
- [12] Danson F.M., Rollin E.M., Information content of high spectral resolution reflectance data for forest canopies, in: *Proceedings of the seventh symposium on Physical Measurements and Signatures in Remote Sens.*, Courchevel, France, 1997, pp. 457–462.
- [13] España M.L., Baret F., Chelle M., Aries F., Andrieu B., A dynamic model of 3D architecture application to the parameterisation of the clumpiness of the canopy, *Agronomie* 18 (1998) 609–626.
- [14] Fourty T., Baret F., Jacquemoud S., Schmuck G., Verdebout J., Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems, *Remote Sens. Environ.* 56 (1996) 104–117.
- [15] Gobron N., Pinty B., Verstraete M.M., Govaerts Y., Presentation and application of an advanced model for the scattering of light by vegetation in the solar domain, in: *Proceedings of the seventh symposium on Physical Measurements and Signatures in Remote Sens.*, Courchevel, France, 1997, pp. 267–273.
- [16] Goel N.S., Strebel D.E., Thompson R.L., Inversion of vegetation canopy reflectance models for estimating agronomic variables. II. Use of angle transforms and error analysis as illustrated by the Suit's model, *Remote Sens. Environ.* 14 (1984) 77–111.
- [17] Goel N.S., Thompson R.L., Inversion of vegetation canopy reflectance models for estimating agronomic variables. V. Estimation of leaf area index and average leaf inclination angle using measured canopy reflectance, *Remote Sens. Environ.* 15 (1984) 69–85.
- [18] Goel N.S., Deering D., Evaluation of a canopy reflectance model for LAI estimation through its inversion, *IEEE Trans. Geosci. and Remote Sens.* GE23 (1989) 674–684.
- [19] Jacquemoud S., Baret F., PROSPECT: a model of leaf optical properties spectra, *Remote Sens. Environ.* 34 (1990) 75–91.
- [20] Jacquemoud S., Inversion of the PROSPECT+SAIL canopy reflectance model from AVIRIS equivalent spectra: theoretical study, *Remote Sens. Environ.* 44 (1993) 281–292.

- [21] Jacquemoud S., Baret F., in: Varlet-Grancher C., Bonhomme R., Sinoquet H. (Eds.), *Crop structure and microclimate: characterizations and applications*, Estimating vegetation biophysical parameters by inversion of a reflectance model on high spectral resolution data, INRA Edition, Paris, France, 1993, pp. 339–350.
- [22] Kimes D.S., Nelson R.F., Manry M.T., Fung A.K., Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements, *Int. J. Remote Sens.* 19 (1998) 2639–2663.
- [23] Knyazikhin Y., Martonchik J.V., Diner D.J., Myneni R.B., Verstraete M.M., Pinty B., Gobron N., Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data, *J. Geophys. Res.* 103 (1998) 32 239–32 256.
- [24] Knyazikhin Y., Martonchik J.V., Diner D.J., Myneni R.B., Running S.W., Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data, *J. Geophys. Res.* 103 (1998) 32 257–32 275.
- [25] Kuusk A., The inversion of the Nilson-Kuusk canopy reflectance model, a test case, in: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'91)*, Helsinki, Finland, 1991, pp.1547–1550.
- [26] Kuusk A., The hot spot effect in plant canopy reflectance, in: Myneni R.B., Ross J. (Eds.), *Photon-Vegetation interactions*, Springer-Verlag, 1991, pp. 139–159.
- [27] Lucht W., Expected retrieval accuracies of bi-directional reflectance and albedo from EOS-MODIS and MISR angular sampling, *J. Geophys. Res.* 103 (1998) 8762–8778.
- [28] Major D., Schaalje G., Wiegand C., Blad B., Accuracy and sensitivity analysis of SAIL model-predicted reflectance of maize, *Remote Sens. Environ.* 41 (1992) 61–70.
- [29] Myneni R.B., Williams D., On the relationship between FAPAR and NDVI, *Remote Sens. Environ.* 49 (1994) 200–211.
- [30] Myneni R.B., Nemani R.R., Running S.W., Estimation of global leaf area index and absorbed PAR using radiative transfer models, *IEEE Trans. Geosci. and Remote Sens.* 35 (1997) 1380–1392.
- [31] Pinty B., Verstraete M., On the design and validation of surface bi-directional reflectance and albedo model, *Remote Sens. Environ.* 41 (1992) 155–167.
- [32] Pragnère A., Baret F., Weiss M., Myneni R.B., Knyazikhin Y., Wang L.B., Comparison of three radiative transfer model inversion techniques to estimate canopy biophysical variables from remote sensing data, in: *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'99)*, Hamburg, Germany, 1999.
- [33] Prévot L., Champion I., Guyot G., Estimating surface soil moisture and leaf area index of a wheat canopy using a dual-frequency (C and X-bands) scatterometer, *Remote Sens. Environ.* 46 (1993) 331–339.
- [34] Price J.C., On the information content of soil reflectance spectra, *Remote Sens. Environ.* 33 (1990) 113–119.
- [35] Price J.C., An approach for analysis of reflectance spectra, *Remote Sens. Environ.* 64 (1998) 316–330.
- [36] Privette J.L., Myneni R.B., Tucker C.J., Emery W.J., Invertibility of a 1D discrete ordinates canopy reflectance model, *Remote Sens. Environ.* 48 (1994) 89–105.
- [37] Qin W., Goel N.S., An evaluation of hot spot models for vegetation canopies, *Remote Sens. Rev.* 13 (1995) 121–159.
- [38] Ruimy A., Saugier B., Dedieu G., Methodology for the estimation of terrestrial net primary production from remotely sensed data, *J. Geophys. Res.* 99 (1994) 5263–5283.
- [39] Roujean J-L., Bréon F-M., Estimating PAR absorbed by vegetation from bidirectional reflectance measurements, *Remote Sens. Environ.* 51 (1995) 375–384.
- [40] Seguin B., Lagouarde J-P., Savané M., The assessment of regional crop water conditions from meteorological satellite thermal infrared data, *Remote Sens. Environ.* 35 (1991) 141–148.
- [41] Turner F.T., Fund M.F., Chlorophyll meter to predict nitrogen topdress requirement for semidwarf rice, *Agron. J.* 83 (1991) 926–928.
- [42] Verhoef W., Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model, *Remote Sens. Environ.* 16 (1984) 125–141.
- [43] Verhoef W., Earth observation modeling based on layer scattering matrices, *Remote Sens. Environ.* 17 (1985) 165–178.
- [44] Weiss M., Baret F., Leroy M., Bégué A., Hauteceur O., Santer R., Hemispherical reflectance and albedo estimates from the accumulation of across track

sun synchronous satellite data, *J. Geophys. Res.* (1999), in press.

[45] Weiss M., Baret F., Evaluation of canopy biophysical variables retrieval performances from the accumulation of large swath satellite data, *Remote Sens. Environ.* (1999), accepted.

[46] Wigneron J-P., Chanzy A., Calvet J-C., Bruguier N., A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields, *Remote Sens. Environ.* 51 (1995) 331–341.