Algorithm Theoretical Basis Document for MERIS Top of Canopy Land Products (TOC_VEG)

Version 3

Frédéric Baret, Cédric Bacour, David Béal
Marie Weiss, Beatrice Berthelot and Peter Regner

March 2006

Contract ESA AO/1-4233/02/I-LG

Executive summary

This ATBD (Algorithm Theoretical Based Document) describes the proposed algorithm for Level 2 biophysical land products derived from MERIS top of canopy reflectance data with justification of the choices made. The biophysical land products considered are the following set of biophysical variables: *LAI*, *fCover*, C_{ab} and $LAI.C_{ab}$. The algorithm accepts as inputs the top of canopy reflectances, i.e. atmospherically corrected top of atmosphere data as derived from MERIS L1b images. It applies both to full and reduced resolution MERIS images.

The proposed algorithm called here TOC_VEG is based on the training of neural networks over a data base simulated using radiative transfer models. The SAIL, PROSPECT models are coupled and used to simulate the reflectance in the 11 MERIS bands considered (490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681.25 nm, 708.75 nm, 753.75 nm, 778.75 nm, 865 nm, 885 nm). The shortest wavelength bands, the oxygen and water absorption bands have not been used because they would convey significant uncertainties associated while providing only marginal information on the surface. The background optical properties are simulated using a collection of soil, water and snow typical reflectance spectra. A brightness factor is used to provide additional flexibility of the background reflectance. Finally, to account for the medium resolution of MERIS observations, mixed pixels are simulated with variable fractions of pure background and pure vegetation.

The simulation of the top of atmosphere reflectance in the 11 MERIS bands requires 14 input variables. They were drawn randomly according to an experimental plan aiming at getting a more evenly populated space of canopy realization. To provide more robust performances of the network, the distributions of each input variable was close to the actual distributions and, when possible, realistic co-distributions were also used. This was achieved by considering a representative distribution of targets over the earth surface that constrains the observation geometry, as well as possible vegetation amount. A total number of 46533 cases were simulated. Half of this data set was used for training, one quarter to evaluate hyper-specialization, and the last quarter to quantify the theoretical performances.

Back-propagation neural networks were trained for each variable considered. The architecture was optimized, resulting in 2 hidden layers of tangent-sigmoid neurones corresponding to a total around. The four variables were estimated concurrently with the same network to provide more consistency between the variables.

The theoretical performances were evaluated over the test simulated data set. It allowed providing estimates of uncertainties. They are close to 0.06 (absolute value) for *fAPAR* and 0.08 for *fCover*, . For LAI, the rmse is close to 0.8 (absolute value) and to 53 for $LAI.C_{ab}$ that shows some loss of sensitivity for the larger values of *LAI* and $LAI.C_{ab}$ due to saturation effects.

Finally, quality assessment criterions are proposed, including the theoretical uncertainties on the product, the reflectance mismatch quantifying the agreement with the training data base, and flags indicating possible values out of range.

This algorithm is to be implemented within the BEAM toolbox.

1D	One directional radiative transfer model
3D	Three directional radiative transfer model
ALA	Average Leaf inclination Angle
albedo	fraction of reflected radiation integrated over view direction & wavelength
ANN	Artificial Neural Network
ATBD	Algorithm theoretical based Document
ASCAR	Algorithm Survey and Critical Analysis Report
BRDF	Bidirectional Reflectance Distribution Function
C	Biophysical variables covariance matrix of (used in the cost function)
C_{ab}	Leaf chlorophyll content
C_{bp}	Leaf brown pigment content
C_m	Leaf dry matter content
C_w	Leaf water content
C_i	Content of constituent *i* per unit leaf area
CYTTARES	Cyclopes Training and Testing Algorithm Reference Ensemble of Sites
DPM	Detailed Processing Methods
ENVISAT	Environment Satellite
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FR	Full resolution (300m)
H	Background moisture
HOT	Hot spot parameter (leaf size relative to canopy height)
fCover	Fraction of vegetation cover
IODD	Input Output Description Document
K	Absorption coefficient used in the leaf model
k_i	Specific absorbtion coefficient for constituent *i*
L2	Level 2 product
L3	Level 3 product
LAI	Leaf Area Index
LB	Lower Bound
LUT	Look Up Table
MERIS	Medium resolution imaging spectrometer
MGVI	MERIS Global Vegetation Index
MISR	Multi-angular Imaging Spectroradiometer
MODIS	Moderate Imaging Spectrometer
M^*_{input}	Matrix of normalised inputs of the Aritifical Neural Network
N	Leaf structure parameter
NDVI	Normalized Difference Vegetation Index
NNT	Neural Network Technique
PROSPECT	A leaf optical properties model
RMSE	Root Mean Square Error
R_n	TOC Reflectance measured in configuration *n*
R_b	Soil background reflectance
R_{snow}	Reflectance of snow
R_{TOC}^{sims}	Simulated top of canopy reflectance
R_{TOC}	Top of canopy reflectance
R_{TOC}^*	Normalized top of canopy reflectance
R_{veg}	Reflectance of the vegetation part of a simulated scene
RR	Reduced Resolution (1200m)
RTM	Radiative Transfer Model

SAIL	Scattering by Arbitrarily inclined Leaves (a 1D radiative transfer model)
SLA	Specific leaf area
SLW	Specific leaf weight
TOA	Top of Atmosphere
TOC	Top of Canopy
TOA_VEG	This algorithm
UB	Upper Bound
V^{max}	Maximum value of the variable *V*
V^{min}	Minimum value of the variable *V*
VI	Vegetation Index
z	Background roughness
ϕ	view azimuth angle relative to the illumination direction
σ	Standard deviation
θ_S	Sun zenith angle
θ_S^*	Normalized sun zenith angle
θ_v	View zenith angle
θ_V^*	Normalized view zenith angle
Ω	Illumination and view geometrical configuration
λ	wavelength

Table of contents

1. Introduction

The MERIS sensor launched in 2002 by the European Space Agency provides unique spectral spatial and temporal characteristics that will make it a very efficient tool for the global monitoring of land surfaces (Rast, Bézy et al., 1999). However, if the potential is high, there were currently no proper products corresponding to biophysical variables to characterize vegetation. The European Space Agency has therefore supported this study dedicated to the development of level 2 biophysical products from MERIS observations acquired both at full and reduced resolutions.

The objective of this document is to provide a detailed description and justification of the algorithm proposed to derive level 2 products from MERIS observations. These products correspond to the following set of biophysical variables: *LAI*, *fCover*, C_{ab} and $LAIxC_{ab}$.

The proposed algorithm follows the recommendations issued in the ASCAR document (Algorithm Survey and Critical Analysis Report) (Baret, Bacour et al., 2003) that was reviewing the current algorithms implemented for several sensors. A validation document is also available, that presents preliminary evaluation evidences, and draws few conclusions on the performances of the algorithm.

This algorithm is dedicated to the estimation of the biophysical variables considered from top of canopy MERIS products.

This ATBD document is split in 3 main sections:

1. **Algorithm overview.** This section contains:
    - A brief description of MERIS main characteristics
    - A definition of the proposed products that could apply both to FR (full resolution) and RR (reduced resolution) MERIS L1b images.
    - The outline of the algorithm.
2. **Description of the algorithm.** This section contains:
    - The inputs required and outputs provided by the algorithm.
    - The reflectance models used. The SAIL canopy reflectance model is used along with the PROSPECT model for the leaf optical properties. The background is described using reference reflectance spectra of soil, modulated using a brightness parameter. In addition, the mixed nature of the medium resolution MERIS pixels is considered here by introducing a vegetation cover fraction.
    - The inversion technique used. Neural network techniques will constitute the core of the operational algorithm. Quality indicators are also provided.
3. **Algorithm prototyping**. In this section the training data base on which the networks are calibrated is described. It is made of reflectance simulations achieved with the radiative transfer models presented above. The theoretical performances are finally described.

2. Algorithm overview

2.1. Instrument characteristics

MERIS is a medium spatial resolution imaging spectrometer operating in the solar reflective spectral range. Fifteen spectral bands are routinely acquired in the 390 nm to 1040 nm spectral range (see Table 1). As compared to other medium resolution instruments, his spectral sampling is very unique and the algorithm developed hereafter will take full advantage of this design.

#	Centre (nm)	Width (nm)	Potential Applications
1	412.5	10	Yellow substance and detrital pigments
2	442.5	10	Chlorophyll absorption maximum
3	490	10	Chlorophyll and other pigments
4	510	10	Suspended sediment, red tides
5	560	10	Chlorophyll absorption minimum
6	620	10	Suspended sediment
7	665	10	Chlorophyll absorption and fluo. reference
8	681.25	7.5	Chlorophyll fluorescence peak
9	708.75	10	Fluo. Reference, atmospheric corrections
10	753.75	7.5	Vegetation, cloud
11	760.625	3.75	Oxygen absorption R-branch
12	778.75	15	Atmosphere corrections
13	865	20	Vegetation, water vapour reference
14	885	10	Atmosphere corrections
15	900	10	Water vapour, land

Table 1. MERIS spectral characteristics: band centre and width

The following figures defined the additional instrument characteristics:

- Band-to-band registration: Less than 0.1 pixel
- Band-centre knowledge accuracy: Less than 1 nm
- Polarisation sensitivity: Less than 0.3%
- Radiometric accuracy: Less than 2% of detected signal, relative to sun
- Band-to-band accuracy: Less than 0.1%
- Dynamic range: Up to albedo 1.0

MERIS is onboard the ENVISAT platform with an helio-synchronic near polar orbit (Table 2).

Orbit altitude (km)	799.8
Repeat cycle (days)	35
Period (min)	100.59
Inclination (°)	98.55
Equatorial descending node crossing time (hr)	10:00

Table 2. Characteristics of the ENVISAT orbit

MERIS scans the Earth's surface by the so called 'push broom' method. CCDs arrays provide spatial sampling in the across track direction, while the satellite's motion provides scanning in the along-track direction. The Earth is imaged with a spatial resolution of 300 m (at nadir) that provides the full resolution data (FR). This resolution is reduced to 1200 m (reduced resolution:

RR) by the on board combination of four adjacent samples across track over four successive lines. The instrument's 68.5° field of view around nadir covers a swath width of 1150 km.

2.2. The Products considered

The products considered correspond to actual vegetation biophysical variables that are defined below:

2.2.1. fAPAR

Although not initially planed because there is already a MERIS fAPAR product, the MGVI (Gobron, Pinty et al., 2000), we included this fAPAR product within the initial list to be able to compare with MGVI and evaluate the consistency. fAPAR corresponds to the fraction of photosynthetically active radiation absorbed by the canopy. The fAPAR value results directly from the radiative transfer model in the canopy which is computed instantaneously and depends both on the canopy structure and illumination conditions. Therefore, fAPAR depends on the sun position. fAPAR is very useful as input to a number of primary productivity models based on simple efficiency considerations (Prince, 1991). Most of the primary productivity models using this efficiency concept are running at the daily time step. Consequently, the product definition should correspond to the daily integrated fAPAR value that can be approached by computation of the clear sky daily integrated fAPAR values as well as the fAPAR value computed for diffuse conditions. To improve the consistency with other fAPAR products that are sometimes considering the instantaneous fAPAR value at the time of the satellite overpass under clear sky conditions (e.g. MODIS). A study was proposed to investigate the differences between these several fAPAR definitions (Baret, Leroy et al., 2003). Results show that the instantaneous fAPAR value at 10:00 (or 14:00) solar time is very close to the daily integrated value under clear sky conditions. To keep a higher consistency with the fAPAR definition used in the CYCLOPES project, we decided to use the instantaneous fAPAR value at 10:00 solar time under clear sky conditions. Note that fAPAR corresponds to the gap fraction in the sun direction assuming that the leaves are black, which is about the case in the PAR spectral domain.
The variable fAPAR is relatively linearly related to reflectance values, and will be little sensitive to scaling issues. Note also that the fAPAR refers only to the green parts (leaf chlorophyll content higher that 15µg.cm-2) of the canopy.

2.2.2. Cover fraction (*fCover*)

It corresponds to the gap fraction for nadir direction. fCover is used for decoupling vegetation and soil in energy balance processes, including temperature and evapo-transpiration. This is also a secondary variable governed by the leaf area index and other canopy structural variables. It is a canopy intrinsic variable that does not depend on variables such as the geometry of illumination as compared to *fAPAR*. For this reason, it is a very good candidate for the replacement of classical vegetation indices for the monitoring of green vegetation. Because of its quasi-linear relationship with reflectances, *fCover* will be only marginally scale dependent (Weiss, Baret et al., 2000). Note that similarly to *LAI and fAPAR*, only the green elements (leaf chlorophyll content higher that 15µg.cm^{-2}) will be considered.

2.2.3. Leaf Area Index (*LAI*)

It defines the size of the interface for exchange of energy (including radiation) and mass between the canopy and the atmosphere. This is an intrinsic canopy primary variable. It is defined as half the developed area of green (leaf chlorophyll content higher than 15 µg.cm^{-2}) vegetation elements per unit of horizontal soil (Privette, Morisette et al., 2001). *LAI* is strongly non linearly related to reflectance. Therefore, its estimation from remote sensing observations will be strongly scale dependent (Weiss, Baret et al., 2000), (Liang, 2000). Note that *LAI* of vegetation as estimated from remote sensing will include all the green contributors, i.e. including understory when existing under forest canopies.

2.2.4. The canopy chlorophyll content (*LAI.Cab*)

The chlorophyll content is a very good indicator of stresses including nitrogen deficiencies. It is strongly related to leaf nitrogen content (Houlès, Mary et al., 2001). This quantity can be calculated both at the leaf level and at the canopy level by multiplication of the leaf level chlorophyll content by the leaf area index. In this case it is obviously an intrinsic secondary variable. Recent studies tend to prove that this product could be of very high interest in primary production models because it partly determines the photosynthetic efficiency (Green, Erickson et al., 2003). In addition, studies have demonstrated that a direct estimation of $LAI.C_{ab}$ is more robust and accurate than an estimation based on the product of the individual estimates of *LAI* and C_{ab} (Weiss, Baret et al., 2000). In addition, the medium resolution scale considered here, generally associated with heterogeneous pixels makes the product $LAI.C_{ab}$ more sound than the leaf level chlorophyll content: what would be the chlorophyll content of a pixel with half of very sparse canopy with very high leaf chlorophyll content and half of very dense canopy with very low leaf chlorophyll content? Therefore, the estimation of $LAI.C_{ab}$ has been preferred to that of the leaf chlorophyll content.

2.3. Requirements for the algorithm selection and design

A review of current state of the art for the estimation of biophysical variables from remote sensing data (Baret, Bacour et al., 2003) allowed to drive requirements for the selection and design of the algorithm proposed in this study for MERIS level 2 products. The main issues required are presented below:

- ***Explicit use of all the MERIS pertinent spectral information***. The spectral sampling of MERIS provides potentially a higher level of information on canopy structure and optical properties of its elements as compared to the simple use of the classical red and near infrared bands implemented in most other retrieval approaches. The exploitation of the whole MERIS spectral information should hopefully allow to restrain the solution space and lead to a more robust and accurate retrieval.
- ***Accuracy of the retrieval and computational efficiency***. Among the several retrieval algorithms, those that are based on the minimisation of the distance in the space of canopy variables appeared to be optimal from the accuracy of the retrievals while being very efficient computationally wise. Therefore, techniques based on neural networks will be selected in this study. In addition, their limitation mainly driven by the necessity to have fixed number of input variables would not constitute any problem to process MERIS data up to level 2, if the geometrical configuration is input explicitly. Note that such techniques have already been implemented and lead to good retrieval performances (Weiss, Baret et al., 2002); (Baret, Weiss et al., 1997); (Combal, Baret et al., 2002); (Kimes, Gastellu-Etchegorry et al., 2002).
- ***Generation of the training data base***. The training data base should sample all the vegetation types and conditions that can be observed from MERIS over land surfaces. In addition it should reflect the uncertainties in the reflectance values as observed by MERIS. Ideally, the training data base should therefore be made of MERIS observations that are paired with accurate ground measurements of the considered biophysical variables. However, because of the uncertainties attached to the ground measurements and the difficulty associated to the collection of such measurements over 300×300m² areas taken within a large range of vegetation types and conditions, this simple 'experimental' approach is not feasible. Therefore, the use of simulations by radiative transfer models would be preferable. The radiative transfer model should simulate within a good accuracy the atmosphere reflectance as observed within MERIS bands and geometry over most vegetation types and conditions that can be observed over the Earth. A particular attention should be brought on:
    - the leaf optical properties, particularly regarding the effect of the chlorophyll content on reflectance and transmittance,
    - the background reflectance that should include in addition to a large variety of soils.

- ***Quality assessment***. Quantitative and qualitative indicators should be attached to the product so that the user could properly 'weigh' the data within his application according to the confidence he puts on. This could be achieved within several ways:
    - Quality of the L1B TOA reflectance used as input to the algorithm. This would simply correspond to the replication of indicators produced previously such as cloud occurrence and sensor problem.
    - Additional indicators based on:
        - The reflectance mismatch. This corresponds to the distance between the MERIS measured reflectance and that simulated by the radiative transfer models. If the distance is too large, then the reliability of the derived product will be questionable.
        - Product uncertainty. The algorithm provides a quantitative estimation of the uncertainty associated to the product.
        - Flags raised when the product appears to be out of the nominal range of variation.

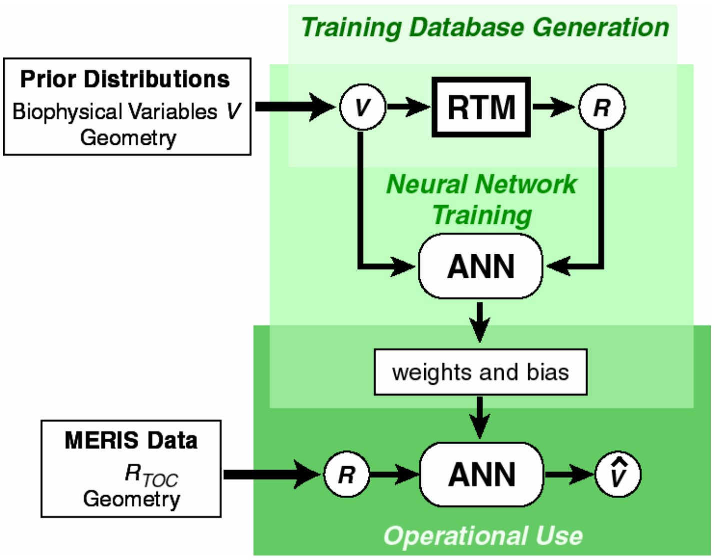

Figure 1. Flow chart showing how the products ($\hat{V}$) are generated operationally. ANN corresponds to Artificial Neural Network characterized par its structure and its coefficients (corresponding to the synaptic weights and bias); R_{TOC} corresponds to the MERIS Top Of Atmosphere reflectance used in the operational mode and V correspond to the biophysical variable in the training data base and estimated by running the ANN over the simulated MERIS TOC reflectance and geometry.

2.4. Algorithm outline

From the arguments previously developed in the ASCAR (Baret, Bacour et al., 2003), we propose to use neural networks to generate the MERIS level 2 products considered. For each product, one particular network will be calibrated. Two main steps are foreseen (Figure 1):

- **Training the neural network**.
- **Operational use of the neural network.**

Note that in addition to the biophysical variables derived by the proposed algorithm, quality indicators will also be computed. This will be described in more details later.

2.4.1. Training the neural network

This process consists mainly in two steps:

- Generation of the training data base
- Defining the neural network architecture and adjusting the corresponding synaptic weight and biases.

2.4.1.1. *Generation of the training data base.*

The generation of the training data base corresponds to the most critical issue to be solved. As stated earlier, it should be based on accurate and representative simulations of the top of atmosphere reflectance and incorporate prior information on the distribution of the input variables. The same training data base will be used for all the products as well as the quality assessment criterions when applicable. The generation of the data base is mainly made within three steps:

- **Generation of the distribution of the input biophysical variables** of the radiative transfer models. The distribution of the other input variables is derived from prior knowledge of their distribution. The geometrical observational conditions are defined by MERIS swath and the ENVISAT orbitography that depends on location and date. Locations and dates are randomly drawn to represent most of the conditions.
- **Simulating the MERIS TOC reflectance**. A radiative transfer model is used to simulate the top of atmosphere reflectance in MERIS bands and observation conditions.
- **Computation of fCover, fAPAR and LAI·C_{ab}**. These secondary variables are computed by the radiative transfer model, as a function of canopy structure and its optical properties.

Once these three steps are completed, the neural network will be actually trained.

2.4.1.2. Training the neural network

The training of the neural network consists in defining the optimal structure (typically the number of layers and the number of neurons per layers) and the corresponding coefficients (i.e. the synaptic weights and biases) that provide the best estimates of the biophysical variables. Dedicated tools are available to achieve this training, and this issue will be detailed later on.

2.4.2. Operational use of the neural network

The neural network once trained will be run in operational mode. Four networks will produce in parallel estimates of the four biophysical variables considered. A complementary step will provide estimates of the associated uncertainties. Additionally, quality assessment indicators will also be generated:

- **Theoretical uncertainties:** This represents the expected error expressed in RMSE between the estimated and the actual biophysical values. As a first approximation, this can be derived from the theoretical performances of the algorithm as evaluated over an independently simulated data set.
- **Quality indicators:** These are a replication of the previously computed quality indicators, including those related to the cloud filtering and sensor possible problems.

3. Prototyping the algorithm

In this section, the algorithmic elements used are described, including:

- The definition of the inputs and outputs,
- The radiative transfer models used
- The inversion technique
- The quality assessment

3.1. Inputs and outputs

This section lists the inputs required and the outputs provided by the algorithm.

3.1.1. Inputs

All these inputs are required for each pixel considered, the image being either full or reduced resolution.

3.1.1.1. MERIS top of canopy reflectance.

Because some wavebands are strongly affected by atmospheric processes while providing only marginal additional information on the canopy, they will be discarded from our analysis. Table 3 lists the 13 bands that are used in the algorithm.

#	Centre (nm)	Width (nm)	Potential Applications
1	412.5	10	Yellow substance and detrital pigments
2	442.5	10	Chlorophyll absorption maximum
3	490	10	Chlorophyll and other pigments
4	510	10	Suspended sediment, red tides
5	560	10	Chlorophyll absorption minimum
6	620	10	Suspended sediment
7	665	10	Chlorophyll absorption and fluo. reference
8	681.25	7.5	Chlorophyll fluorescence peak
9	708.75	10	Fluo. Reference, atmospheric corrections
10	753.75	7.5	Vegetation, cloud
11	760.625	3.75	Oxygen absorption R-branch
12	778.75	15	Atmosphere corrections
13	865	20	Vegetation, water vapour reference
14	885	10	Atmosphere corrections
15	900	10	Water vapour, land

Table 3. The 13 MERIS bands used in the algorithm. The bands appearing in grey are not used.

Bands 1, 2, 11 and 15 were not used for the following reasons:

- **Bands 1 and 2** correspond to the sorter wavelengths were atmospheric correction accuracy is minimum. They would thus convey large uncertainties as well as little additional information on canopy characteristics.
- **Band 11**. This very narrow band is just located in the oxygen absorption band at the end of chlorophyll absorption. It would bring only marginal additional information on leaf and

background optical properties while conveying errors due to uncertainties in oxygen pressure values.

- **Band 15**. This water absorption band will not bring very significant information on canopy characteristics as compared to bands 12 to 13 while also conveying errors due to uncertainties in water vapour values.

3.1.1.2. MERIS geometry of observation.

The following angles are required:

- **View zenith angle (θ_v),**
- **Sun zenith angle (θ_s)**
- **Relative azimuth angle** between sun and view directions (ϕ).The **cosine** of this angle was used as an input to the NNET in order to keep its circular character.

These angles derive from the ENVISAT orbitography and MERIS swath, as a function of the date of observation, expressed in day of the year (from 1 to 366), and of the location of the pixel, expressed in latitude and longitude.

3.1.1.3. Quality indicators

These indicators will come from the previous products. They mainly correspond to:

- **MERIS radiometric quality**, including cloud snow and water flags. These flags will be used to turn on or off the algorithm in case of very poor radiometric quality, cloud contamination or water pixels.

3.1.2. Outputs

The outputs will be provided by application of the algorithm over each pixel and will include the following:

3.1.2.1. Biophysical variables estimation

It corresponds to the neural network derived *fAPAR fCover*, *LAI*, and $LAI.C_{ab}$ values as described in §.2.2. The range of variation and resolution steps proposed are presented in Table 4.

Product	Unit	Minimum	Maximum	resolution
fAPAR	-	0	1.0	0.01
fCover	-	0	1.0	0.01
LAI	$m^2.m^{-2}$	0	6.0	0.01
$LAIxC_{ab}$	$g.m^{-2}$	0	500	1

Table 4. Minimum, maximum values and associated resolution for all the products considered.

3.1.2.2. Quality indicators

These indicators will provide information on the quality of:

- **The inputs used to compute the products.** This includes
  - replication of previously computed indicators (clouds, type of surface, flags for MERIS radiometric quality, …),
  - information on aerosol optical thickness derived and associated uncertainties,
- **Product uncertainties**, i.e. expected standard deviation of the estimates,
- **Out of range flags**. In the case where the ANN provides biophysical variable estimates outside their definition range as defined in Table 4, a flag will be delivered and the

corresponding product value will be set to the closest bound of the range, i.e. either the minimum or the maximum accepted values. The uncertainty value will be set to 999.

3.2. The training database

The training data base consists in an ensemble of canopy reflectance spectra in the MERIS configurations of observation, together with the corresponding biophysical products. Ideally, the database must be representative of actual biomes and conditions as observed by MERIS. Nevertheless, the difficulty to conduct accurate ground measurements at such spatial resolution over a large range of surface types implies that the training database generally reduces to radiative transfer model simulations (Atzberger, 2004; Danson et al., 2003) with inherent empirical assumptions on the distributions of the variables.

3.2.1. Radiative transfer model

Radiative transfer model allows simulating both the top of canopy reflectances and the secondary canopy variables that are *fAPAR* and *fCover*. Although very realistic, complex 3D models appear difficult for simulating a very large range of canopy situations, because they require extensive parameterizations and are very computationally demanding. One dimensional canopy radiative transfer models that run very fast are more managable for the generation of the training database. They are based on a simple description of the canopy architecture that do not account for horizontal variations of the leaf area density corresponding to heterogeneous canopies.

The widespread SAIL (Verhoef, 1984; Verhoef, 1985) and PROSPECT (Jacquemoud and Baret, 1990) models are used to simulate reflectances in 11 MERIS narrow spectral bands centered at 490, 510, 560, 620, 665, 681, 709, 754, 779, 865, and 885 nm, as well as *fAPAR* and *fCover*. The four additional wavebands of the MERIS instrument are not used here because they are too strongly affected by atmosphere effects. The top of canopy reflectance are simulated as a function of the configuration of observation and a limited number of variables describing its architecture: *LAI*, the mean leaf inclination angle (*ALA*), assuming an ellipsoidal distribution of foliage elements (Campbell, 1990), and the hot spot parameter (*HotS*) as implemented by Kuusk (1991). The optical properties of the leaves are simulated by PROSPECT. The version used here (Fourty and Baret, 1997) depends on the leaf structure parameter (*N*), the chlorophyll a and b content (*Cab* in $\mu g.cm^{-2}$), the equivalent water thickness (C_w in cm), the dry matter content (C_m in $g.cm^{-2}$), and the brown pigment concentration (C_{bp} in relative units). Background reflectance is described by one – among ten – standard soil reflectance spectrum (Liu et al., 2003), potentially combined with a typical snow or water spectrum, the presence of water and snow being exclusive. Such composite backgrounds are expected to occur frequently because of the medium spatial resolution of MERIS. Possible variations of the background reflectance are accounted for by a multiplicative brightness parameter (β). The latter is assumed wavelength independent and allows considering changes in the background optical properties, confounding the effects of geometrical conditions, roughness and moisture. An additional variable describing the fraction of ground covered by 'pure' vegetation (*vCover*) is introduced to represent some spatial heterogeneity within a pixel. A simulated scene is thus composed of a fraction of vegetation covering the background (reflectance R_{veg}) and a lower fraction of bare background (reflectance R_{soil} similar to the one under the pure canopy); the reflectance *R* of the composite scene then expresses as: $R = R_{veg} \times vCover + R_{soil} \times (1\text{-}vCover)$. In total, the PROSPECT+SAIL model requires ten input variables to simulate the spectro-directional variation of the reflectance. Because no assumptions are made on the spatial resolution of the scene, the algorithm should therefore performs on MERIS data acquired both at full and reduced resolutions.

3.2.2. Simulations

The input variables of the PROSPECT+SAIL model are set to follow Gaussian distributions within their respective definition interval (Table 5). The distributions derive from empirical knowledge and are assumed independent. In the case of *LAI*, this original distribution is modified in order to give more weight to low values: the distribution is set uniform for *LAI* values lower than the prescribed mean as it will be justified later. To reduce the degrees of freedom (and consequently the number of simulations of the training database), the leaf water content is tied to the dry matter content assuming that green leaves have a relative water content close to 80%. This is justified as leaf water has only marginal effect on TOC reflectances in the 11 MERIS.

The geometry of observation (view and sun zenith angles; relative azimuth angle) is driven by the ENVISAT orbit characteristics and MERIS swath, then depending on location and date. The date is drawn uniformly within four periods of 42 days centred over the solstices and equinoxes. The determination of location relies on a network of 350 sites (Derive et al., 2003), chosen so as to sample all possible types of vegetation structure that can be encountered over the Earth's surface. Finally, the fraction of surface covered by water or snow is determined according to empirical distribution laws that favour the soil as the dominant background type.

		mean	σ	LB	UB	# classes
Leaves	***N***	1.5	1	1	4.5	3
	Cab (μg.cm^{-2})	50	50	15	100	4
	C_m (g.cm^{-2})	0.0075	0.0075	0.002	0.02	3
	C_{bp}	0	0.6	0	1.5	3
Soil	β	0.8	0.3	0.3	1.3	3
Canopy	***LAI***	2.5	3.5	0	8	4
	ALA (°)	60	20	30	85	3
	HOT	0.1	0.3	0.0001	1	3
	vCover	uniform		0.85	1	1

Table 5: Distribution characteristics of the input variables along with the corresponding number of classes of the experimental plan used to simulate the training dataset with the PROSPECT+SAIL model. Truncated Gaussian distributions are used; they are characterized by their mean, standard deviation (σ), and lower (LB) and upper bounds (UB) of their variation range.

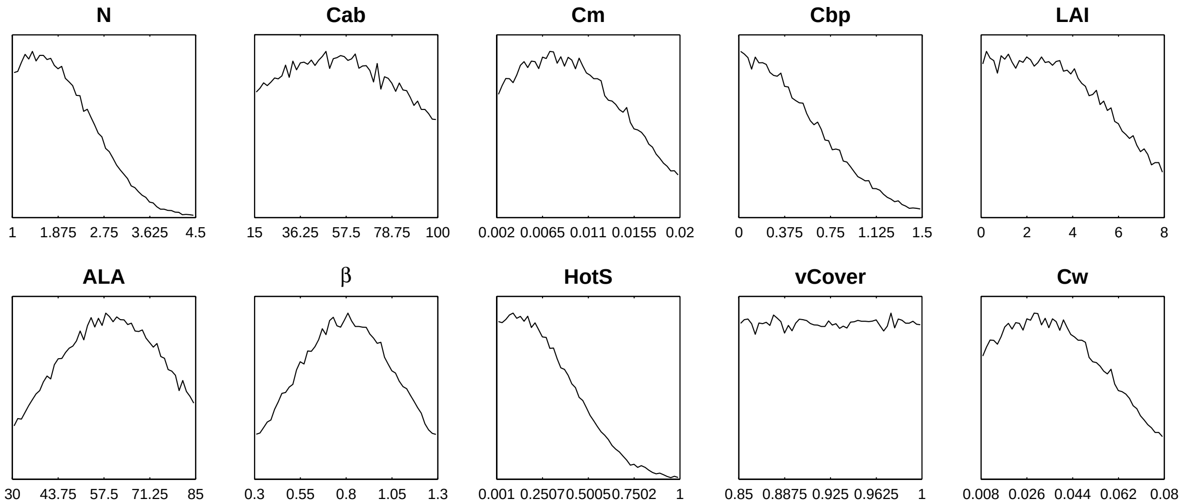

Figure 2: Probability distributions of the PROSPECT+SAIL input variables of the training database.

The simulations correspond to several combinations of the above described input model variables and of the observation geometries. The corresponding various canopy situations are set according to a sampling scheme based on a full orthogonal experimental plan (Bacour et al., 2002). For this purpose, the definition interval of each variable is split in a given number of classes. All combinations of classes are sampled once. This allows accounting for all the interactions between variables, while having their range of variation densely and near randomly populated. The generation of the database for the biophysical variables is performed by drawing randomly their actual values, with respect to the distribution laws specified above, within the pre-identified classes (Table 5**Erreur ! Source du renvoi introuvable.**). The distributions of the PROSPECT+SAIL input variables, and the corresponding simulated *fAPAR*, *fCover* are presented in Figure 2 and Figure 3. Despite the emphasis put on low *LAI* values, *fAPAR* and *fCover* distributions have more frequent higher values (Figure 3 a and b). Note that, in this study, *fAPAR* is defined as that observed at 10:00 local solar time when the canopy is only illuminated by the sun (black sky). It is thus rather close to the black sky *fAPAR* values at the time of satellite overpass used by other products (MODIS, *MGVI*). It corresponds also to the value the closest to the daily integrated black sky value (results not shown here) required by the majority of users. A 4% Gaussian noise with no bias was added to the reflectances. It corresponds roughly to MERIS performances as evaluated over vicarious calibration exercises (Zurita-Milla et al., 2006). The distribution of reflectances in the 11 MERIS bands show the contrast between the visible (lower values) and near infrared bands (higher values) with the red-edge bands being intermediate (Figure 3 c).

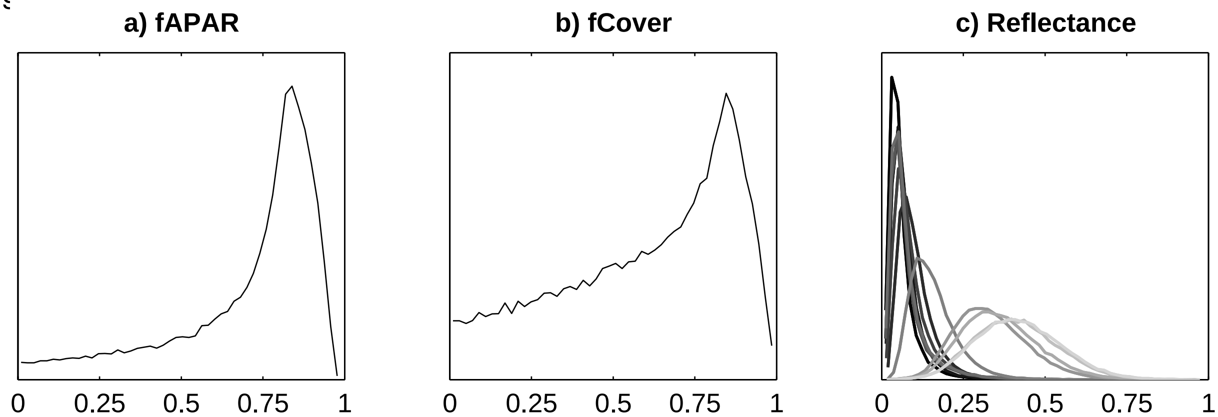

Figure 3: Probability distributions of the simulated **a)** *fAPAR*, **b)** *fCover*, **c)** reflectances in the 11 MERIS spectral bands (lines darken from the highest to the lowest waveband).

In total, 46533 simulations were performed. Half of these were randomly selected to compose the training database, the remaining half being split into the hyper-specialization and testing datasets. The hyper-specialization base is used during the learning process of the network to avoid over fitting the training dataset which would result in a lack of genericity. The testing base serves as an internal evaluation of the NN estimation performances.

3.2.3. Neural network design

3.2.3.1. Architecture and training

Multilayer perceptrons are commonly used to interpret remote sensing measurements owing to their ability to approximate complete inverse functions (Abuelgasim et al., 1998; Gross et al., 2000; Kimes et al., 1998). A multilayer perceptron is here designed to jointly estimate *fAPAR*, *fCover*, *LAI*, and *LAIx Cab*. Using a single NN for estimating four biophysical variables at the same time rather than four different NNs, each specialized for a given variable, allows imposing an additional physical constraints to the inverse problem as the variables are not independent. This way, the NN is trained to learn the intrinsic relationships that link them together.

The inputs of the network correspond to the reflectance in the 11 MERIS bands and the 3 angles defining the observation geometry. The relative azimuth angle is expressed as its cosine function to preserve continuity with respect to the principal plane. The values x of the 14 inputs and 4 outputs are standardized according to $x^* = 0.66.(x - \bar{x}) / \sigma(x)$ so that 80% of the standardized values (x^*) fall within [-1; 1], $\bar{x}$ being the mean of the variable determined on the training database, and $\sigma(x)$ its standard deviation. For the reflectance data, $\bar{x}$ and $\sigma(x)$ are determined over all wavebands so as to maintain the relative proportions between them. The standardization aims at preventing any scaling factor problem between data of different physical nature, while increasing at the same time the convergence performances of the training algorithm.

The training process (calibration of the NN coefficients) relies on the minimization of a misfit function by a back-propagation algorithm (Atkinson and Tatnall, 1997; Gross et al., 2000). The misfit function is here defined as the mean square error between the targeted variables (from the database) and the NN outputs. The optimal network architecture was determined after several trials, to provide the best estimation performances with the minimum number of layers and neurons per layers. The final network is made of one input layer with 14 linear neurons (one per input), two hidden layers of respectively 13 and 7 neurons with tangent sigmoid transfer functions, and one output layer with 4 linear neurons (one per variable to estimate). The combination of tangent sigmoid and linear transfer functions is used because it is recognized as capable of fitting any type of function.

Posterior quality tests on the estimates are determined. First, if the retrieved value falls outside the definition range of the corresponding variable, it is strained to take the value of the closest bound. The second test is based on the assumption that *LAI*, *fAPAR*, and *fCover*, are physically interconnected: the distribution of the values taken by two of them are contained in a bounded bidimensional space. Therefore, whenever the estimates of the variables taken two by two will fall outside the corresponding bidimensional convex hull they will be discarded. The hulls are defined by the co-distribution of the variables from the training database.

3.3. Theoretical estimation performances

The NN performances are evaluated on the testing dataset. As compared to the original co-variations of the biophysical variables, these of the estimates are contained inside the central region of the convex hulls (Figure 4a), corresponding to the most probable values. Figure 4b presents the comparison of the true values of the biophysical variables with the NN estimates. The corresponding root mean square errors and coefficients of correlation (Table 6) appraise for the theoretical estimation uncertainties related to the inversion process.

	fAPAR	***fCover***	***LAI***	***LAIxCab***
RMSE	0.06	0.09	1.15	77.44
R	0.95	0.93	0.80	0.84

Table 6: Root mean square errors and coefficients of correlation (R) between the NN estimates of *fAPAR*, *fCover*, *LAI*, and *LAIxCab*, and the true values from the testing database.

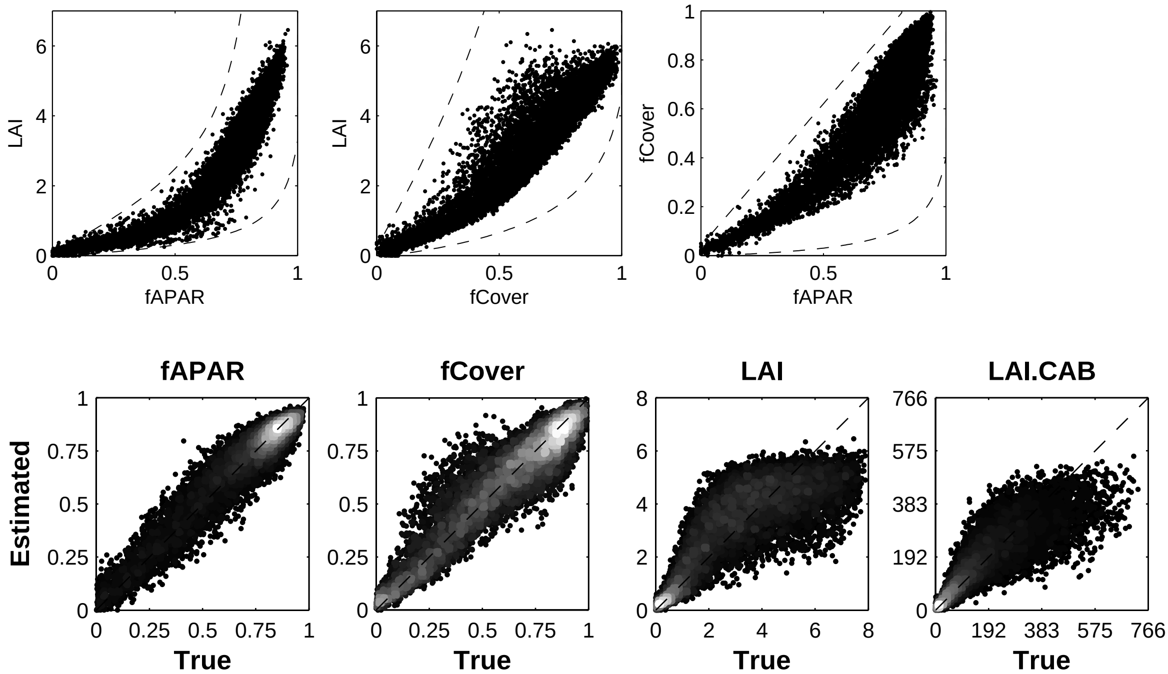

Figure 4: a - top) Scatter plots between the variables estimated on the testing base and corresponding convex hulls from the training database. **b - bottom)** Comparison between the values of the variables of the testing database and the corresponding NN estimates for *fAPAR*, *fCover*, *LAI*, and *LAIxCab* (µg.cm-2). Each inset is a bidimensional histogram where the gray level intensity increases with the density of points.

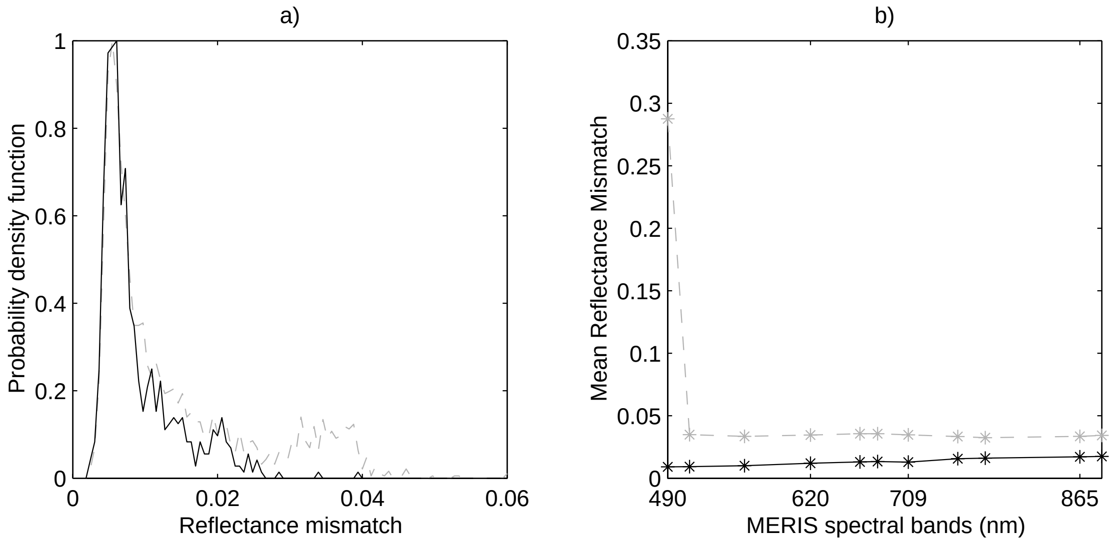

Figure 5: a) Distribution of the reflectance mismatch integrated over the 11 wavebands between the actual MERIS data and the simulated reflectance spectra of the training database, before (dashed grey line) and after cloud screening (black solid line). **b)** Reflectance mismatch averaged over all scenes for each MERIS channel.

Regarding to the complexity of the inverse problem to solve and the high diversity of canopy realizations and observation geometries of the training database, the results show the ability of the neural inversion to retrieve *fAPAR* and *fCover* with a fairly good accuracy, thus confirming previous results (Weiss et al., 1999). The performances degrade for *LAI* and *LAIxCab*: the scattering around the 1:1 line increases with *LAI* or *LAIxCab* values and some underestimation is observed in these conditions. This feature is due to the non-linearity in the radiative transfer with respect to the canopy structural and optical characteristics. The radiometric signal saturates for high *LAI* values. In these cases, the NN logically underestimates *LAI* values (typically for *LAI* higher than 5). On another hand, the neural network is trained to globally estimate a variable without bias. This explains why intermediate *LAI* (values between 2-5) are generally slightly overestimated to compensate for the underestimation that occurs for the larger *LAI* values. The modification of the initial Gaussian *LAI* distribution in the training database is here justified *a posteriori* : it allows giving more weights to the low and intermediate values in the training process.

3.4. Quality Assessment

A brief list of quality assessment criterions was presented in §.2.3. In this section, more details are provided except for the product uncertainties and reflectance mis-match that will be described along with the algorithm prototyping section §.4.2.

- **Quality indicators.** The same quality indicators to those presented as inputs will be replicated as outputs.
- **Out of range flag**. In the case where the ANN provides biophysical variable estimates outside their definition range a flag will be triggered. The corresponding product value will be set to the closest bound of the range, i.e. either the minimum or the maximum accepted values. The product uncertainty value will be set to 999.
- **Product uncertainties** The uncertainties associated to each biophysical variable are also coded with the same resolution as that used for the biophysical variables presented in Table 4. The way they will be derived will be presented in §.4.3.
- **Spectra out of the training domain.** When the L1B MERIS reflectance spectra appears to be out of the definition domain of the training data base, a flag is raised..

4. Conclusion

This ATBD provides a description of the TOC_VEG algorithm used to compute *fAPAR*, *fCover*, *LAI*, and $LAI.C_{ab}$ products from MERIS top of canopy reflectance data both at full and reduced spatial resolution. The performances of this TOC_VEG algorithm were evaluated over an independently simulated data set. They show accurate estimates for *fAPAR* and *fCover*, independent from the value of the variable. However, *LAI* and $LAI.C_{ab}$ show less accurate estimates, particularly for the larger *LAI* or $LAI.C_{ab}$ values. This is obviously due to the physics of the radiative transfer, although improvements could be foreseen by adaptation of the training data base, with probably more cases with larger *LAI* values.

This algorithm was validated using actual MERIS observations. The corresponding results are reported in a separate document (Baret, Weiss et al., 2006).

5. References

Asner, G.P., Wessman, C.A., Schimel, D.S. and Archer, S., 1998. Variability in leaf and litter optical properties: implications for BRDF model inversions using AVHRR, MODIS, and MISR. Remote Sensing of Environement, 62:243-257.

Bacour, C., Jacquemoud, S., Tourbier, Y., Dechambre, M. and 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. et al., 2005. Algorithm Theoretical Basis Document for MERIS Top of canopy Land Products (TOC_VEG), INRA-CSE, Avignon.

Baret, F., Bacour, C. and Weiss, M., 2003. Algorithm Survey and Critical Analysis Report for MERIS L2 Land products, INRA & Noveltis, Avignon.

Baret, F., Jacquemoud, S. and Hanocq, J.F., 1993. The soil line concept in remote sensing. Remote Sensing Reviews, 7: 65-82.

Baret, F. et al., 2003. CYCLOPES User Requirement Document, INRA-CSE, Avignon.

Baret, F., Pavageau, K., Béal, D. and Bacour, C., 2004. Development and evaluation of a fully coupled approach for biophysical products estimation from MERIS data., INRA-CSE, Avignon.

Baret, F. et al., 1997. Impact of surface anisotropies on the observation of optical imaging sensors, INRA Bioclimatologie.

Baret, F. et al., 2006. Report on the validation of vegetation products derived from MERIS using the TOA_VEG_V3 algorithm, INRA-CSE, Avignon.

Biard, F. and Baret, F., 1997. Crop residue estimates using multiband reflectance data. Remote Sensing of Environment, 59: 530-536.

Borel, C.C., Gerstl, S.A.W. and Powers, B.J., 1991. The radiosity method in optical remote sensing of structured 3-D surfaces. Remote Sens. Environ., 36: 13-44.

Campbell, G.S., 1986. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution. Agric. For. Meteorol., 36: 317-321.

Chelle, M., Andrieu, B. and Bouatouch, K., 1997. Nested radiosity for plant canopies. The Visual Computer, Nov1997-Mar1998: 1-24.

Chen, Y. and McKyes, E., 1993. Reflectance of light from the soil surface in relation to tillage practices, crop residues and the growth of corn. Soil & Tillage Research, 26: 99-114.

Combal, B. et al., 2002. Retrieval of canopy biophysical variables from bi-directional reflectance data. Using prior information to solve the ill-posed inverse problem. Remote Sensing of Environment, 84: 1-15.

Demuth, H. and Beale, M., 1998. Neural network toolbox user's guide.

España, M. et al., 1999. Modeling maize canopy 3D architecture. Application to reflectance simulation. Ecological Modeling, 122: 25-43.

Fourty, T. and Baret, F., 1997. Amélioration de la précision des coefficients d'absorption spécifique de la matière sèche et des pigments photosynthétiques, INRA Bioclimatologie, Avignon.

Gastellu-Etchegorry, J.P., Demarez, V., Pinel, V. and Zagolski, F., 1996. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. Remote Sensing of Environment, 58: 131-156.

Gausman, H.W. et al., 1975. Reflectance differences between crop residues and bare soils. Soil science society of America proceedings., 39(4): 752-755.

Gerstl, S.A.W. and Borel, C.C., 1992. Principles of the radiosity method versus radiative transfer for canopy reflectance modeling. IEEE transactions on geoscience and remote sensing, 30(2): 271-275.

Gobron, N., Pinty, B., Verstraete, M. and Widlowski, J.L., 2000. Advanced vegetation indices optimized for up-coming sensors: design, performances and applications. IEEE Transactions on Geoscience and Remote Sensing, 38(6): 2489-2505.

Govaerts, Y.M. and Verstraete, M.M., 1998. Raytran: a monte carlo ray tracing model to compute light scattering in three dimensional heterogeneous media. IEEE Transactions on Geoscience and Remote Sensing, 36(2): 493-505.

Green, D.S., Erickson, J.E. and Kruger, E., 2003. Foliar morphology and canopy nitrogen as predictors of light use efficiency in terrestrial vegetation. Agricultural and Forest Meteorology, 115: 163-171.
Harrel, F.E., 2001. Regression modeling strategies. Springer, New-York.
Houlès, V., Mary, B., Machet, J.M., Guérif, M. and Moulin, S., 2001. Do crop characteristics available from remote sensing allow to determine crop nitrogen status? In: G. Grenier and S. Blackmore (Editors), 3rd European Conferance on Precision Agriculture. Agro Montpellier, Montpellier, pp. 917-922.
Jacquemoud, S. and Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. Remote Sensing of Environment, 34: 75-91.
Jacquemoud, S., Baret, F. and Hanocq, J.F., 1992. Modeling spectral and directional soil reflectance. Remote sensing of the Environment, 41: 123-132.
Kimes, D.S., Gastellu-Etchegorry, J.P. and Esteve, P., 2002. Recovery of forest canopy characteristics through inversion of complex 3D model. Remote Sensing of Environment, 79: 320-328.
Knyazikhin, Y. et al., 1999. MODIS Leaf area index (LAi) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document, http://eospso.gsfc.nasa.gov/atbd/modistables.html.
Kuusk, A., 1985. The Hot Spot effect of a uniform vegetative cover. Sov. J. Remote Sensing, 3(4): 645-658.
Le Maire, G., 2002. Utilisation de la télédétection hyperspectrale pour la détermination des caractéristiques biophysiques et biochimiques des couverts végétaux: de l'échelle de la feuille à l'échelle du couvert., Laboratoire d'Ecologie Végétale, université Paris-Sud, Orsay (France).
Liang, S., 2000. Numerical experiments on the spatial scaling of land surfaces albedo and leaf area index. Remote Sens. Rev., 19: 225-242.
Liu, W. et al., 2002a. Relating soil surface moisture to reflectance. Remote Sensing of Environment, 81: 238-246.
Liu, W. et al., 2002b. Evaluation of methods for soil surface moisture estimation from reflectance data. International Journal of Remote Sensing, in press.
Masson, V., Champeaux, J.L., Chauvin, F., Meriguer, C. and Lacaze, R., 2003. A global database of land surface parameters at 1km resolution in meteorological and climate models. Journal of Climate, 16(9): 1261-1282.
Myneni, R.B., Asrar, G. and Hall, F.G., 1992. A Three-dimensional radiative transfer method for optical remote sensing of vegetated land surfaces. Remote Sensing of Environment, 41: 105-121.
Newnham, G.J. and Burt, T., 2001. Validation of leaf reflectance and transmittance model for three agricultural crop species. IEEE Transactions on Geoscience and Remote Sensing: 2976-2978.
Prince, S.D., 1991. A model of regional primary production for use with coarse resolution satellite data. International Journal of Remote Sensing.
Privette, J.L., Morisette, J., Baret, F., Gower, S.T. and Myneni, R.B., 2001. Summary of the International Workshop on LAI Product Validation. Earth Observer, 13(3): 18-22.
Rahman, H. and Dedieu, G., 1994. SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. International Journal of Remote Sensing, 15(1): 123-143.
Rast, M., Bézy, J.-L. and Bruzzi, S., 1999. The ESA Medium Resolution Imaging Spectrometer MERIS -a review of the instrument and its mission. International Journal of Remote Sensing, 20: 1682-1701.
Rummelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning internal representations by error propagation. In: D. Rummelhart and J. Mc Clelland (Editors), Parallel data processing. M.I.T. press, Cambridge, MA (USA), pp. 318-362.
Sanz, C. et al., 1997. Bi-directional characteristics of leaf reflectance and transmittance: measurement and influence on canopy bi-directional reflectance. In: G. Guyot and T. Phulpin (Editors), 7th International Symposium on physical measurements and signatures in remote sensing. Balkema, Courchevel (France), pp. 583-590.

Soler, C., F., S., Blaise, F. and de Reffye, P., 2001. A physiological plant growth simulation engine based on accurate radiant energy transfer. 4116, INRIA, Montbonnot-Saint-Martin (France).
Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote Sensing of Environment, 16: 123.
Verhoef, W., 1985. Earth Observation modeling based on layer scattering matrices. Remote Sensing of Environment, 17: 165-178.
Vermote, E.F., Tanré, D., Deuzé, J.L., Herman, M. and Mocrette, J.J., 1997. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. IEEE Transactions on Geoscience and Remote Sensing, 35(3): 675-686.
Weiss, M. et al., 2002. Validation of neural net techniques to estimate canopy biophysical variables from remote sensing data. Agronomie, 22: 547-554.
Weiss, M., Baret, F., Myneni, R., Pragnère, A. and Knyazikhin, Y., 2000. Investigation of a model inversion technique for the estimation of crop charcteristics from spectral and directional reflectance data. Agronomie, 20: 3-22.
Zarco-Tejada, P.J. et al., 2001. Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies. Remote Sensing of Environment, submitted.