



Algorithm Theoretical Basis Document for MERIS Top of Atmosphere Land Products (TOA_VEG)

Version 3

Frédéric Baret, K. Pavageau, David Béal
Marie Weiss, Beatrice Berthelot and Peter Regner

March 2006

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



This ATBD (Algorithm Theoretical Based Document) describes the proposed algorithm for Level 2 biophysical land products derived from MERIS top of atmosphere 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 atmosphere reflectances as directly derived from MERIS L1b images. It applies both to full and reduced resolution MERIS images.

The proposed algorithm called here TOA_VEG is based on the training of neural networks over a data base simulated using radiative transfer models. The SAIL, PROSPECT and SMAC models are coupled and used to simulate the reflectance in the 13 MERIS bands considered (412 nm, 442 nm, 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 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 13 MERIS bands requires 15 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 129600 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. The data base was further streamlined according to the expected relationship between LAI and fAPAR, as well as using an actual MERIS L1B observation data sets. This resulted in the elimination of about 24% of the cases simulated.

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 220 coefficients to adjust, and providing a good ratio (≈ 400) with the size of the training data base.

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.

Symbols and Acronyms

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 absorption 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 Artificial 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	7
2. ALGORITHM OVERVIEW	8
2.1. Instrument characteristics	8
2.2. The Products considered	9
2.2.1. fAPAR	9
2.2.2. Cover fraction (<i>fCover</i>)	9
2.2.3. Leaf Area Index (<i>LAI</i>)	9
2.2.4. The canopy chlorophyll content (<i>LAI.Cab</i>)	10
2.3. Full or reduced resolution	10
2.4. Requirements for the algorithm selection and design	11
2.5. Algorithm outline	13
2.5.1. Training the neural network	13
2.5.1.1. <i>Generation of the training data base</i>	13
2.5.1.2. Training the neural network	13
2.5.2. Operational use of the neural network	14
3. DESCRIPTION OF THE ALGORITHMIC ELEMENTS	15
3.1. Inputs and outputs	15
3.1.1. Inputs	15
3.1.1.1. MERIS top of atmosphere reflectance	15
3.1.1.2. MERIS geometry of observation	16
3.1.1.3. Additional atmosphere characteristics	16
3.1.1.4. Quality indicators	16
3.1.2. Outputs	16
3.1.2.1. Biophysical variables estimation	16
3.1.2.2. Quality indicators	16
3.2. Reflectance models	17
3.2.1. Leaf optical properties	17
3.2.2. Canopy radiative transfer models	18
3.2.3. Background reflectance model	19
3.2.3.1. The background brightness concept	19
3.2.3.2. Background spectral variation	20
3.2.4. Atmosphere model	21
3.3. Quality Assessment	22
4. ALGORITHM PROTOTYPING	23
4.1. Generation of the training database	23
4.1.1. Radiative transfer model	23
4.1.2. Generation of the vegetation input variables database	24
4.1.3. Simulation of the top of atmosphere reflectance for the 13 MERIS bands	25
4.1.4. Streamlining the training data base	25
4.1.4.1. Reflectance mis-match	25
4.1.4.2. LAI/fAPAR streamlining	26
4.1.5. Realism of the simulated reflectances	27
4.2. Training the neural network	28
4.2.1. Normalization of the input and output values	28
4.2.2. Network architecture	28
4.2.3. Learning process	28
4.3. Theoretical performances of the artificial neural network	30
5. CONCLUSION	34

6. REFERENCES.....35

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 $LAI \times C_{ab}$.

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 by nature very similar to the one proposed for exploiting top of canopy MERIS reflectances (Baret, Bacour et al., 2005). However, its main advantage is to be able to estimate the biophysical variables considered directly from L1b MERIS products within one single step.

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 SMAC model is then used to simulate the top of atmosphere reflectance from the surface reflectance.
- 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-synchronous 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 planned 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 than 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 than 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 ($LAI.C_{ab}$)

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. Full or reduced resolution

Although it would have been possible to present algorithm specifically tuned for the reduced or full spatial resolution, we will not make any difference between these two resolutions: The same algorithm is expected to work similarly both on FR and RR images. This is simply proven by the fact that any full resolution spectra could be found with a very good match in the aggregated reduced resolution image. To demonstrate this important finding a particular study was conducted.

Site name	Lat (°)	Long (°)	Date	surface type	size pixels)
North-Spain	43.11	-1.46	28/05/05	Forests-grassland	700x700
North France	46.62	1.39	28/05/05	Forests-crops	700x700
North Chile	-17.22	-66.45	28/05/05	Forests-grassland	500x500
Central Africa	13.58	-2.03	16/05/05	tropical forests	500x500
East Argentina	-36.90	-58.21	04/11/03	forests	300x300
East Finland	57.43	26.38	14/05/04	forests	700x700
South France	44.68	4.17	15/05/04	forests-crops	300x300
North Chile	-17.07	-68.34	02/05/05	forests	600x600
West Russia	46.18	39.17	14/09/04	crops	300x300
North Spain	44.12	0.63	25/05/05	crops	300x300

Table 3. The sites used to investigate the spatial resolution

Ten MERIS FR images were considered, representing a large range of variation of the surface conditions. The images were mostly clear, but few images were showing small cloud fraction (Table 3). Then, the corresponding RR images were simulated by aggregating the FR images by cells of 4x4 pixels. For each of the 10 sites, 200 pixels were randomly selected in the FR images. The pixel showing the closest reflectance spectra (quadratic distance) in the RR images was identified. The same process was repeated over the 10 scenes, representing a total of 2000 FR pixels and the residuals for each band between the FR and closest RR reflectance was computed. Figure 1 shows that for this representative set of pixels, it is possible to approximate any of the FR reflectance spectra by one of the one of the RR image. Therefore, it is concluded that if the algorithm applies on the RR images, it should also apply to the FR images in a very similar way.

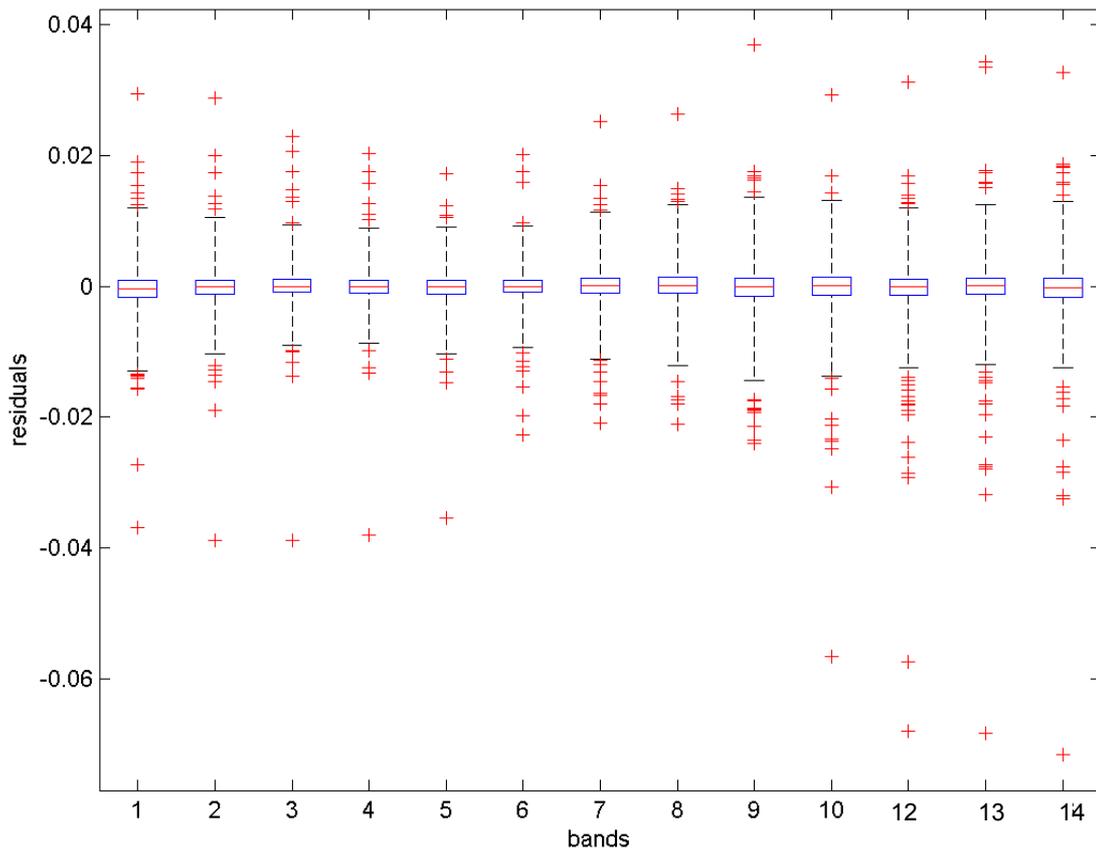


Figure 1. Residuals between reflectance spectra of the FR images and the closest one in the RF images. Box plot representation: the box corresponds to 50% of the cases, and the whiskers to 99%. Red ‘plus’ correspond to individual outliers.

2.4. 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.

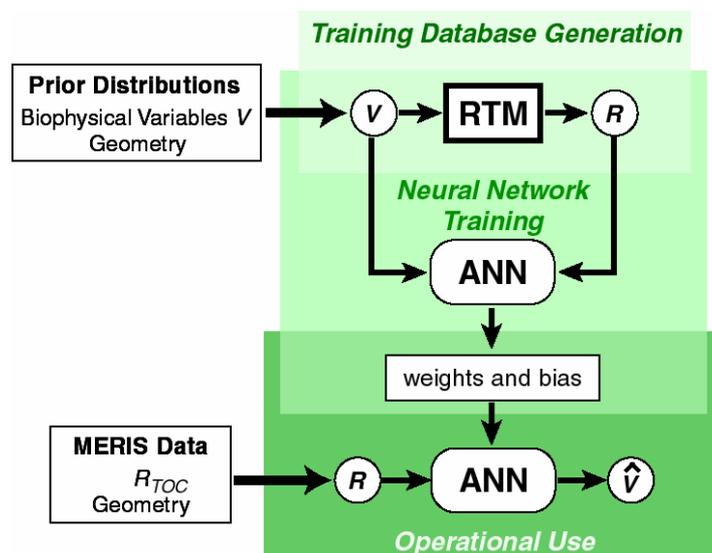


Figure 2. Flow chart showing how the products (\hat{V}) are generated operationally. ANN corresponds to Artificial Neural Network characterized par its structure and its coefficients (corresponding to the synaptic weights and bias); R_{TOC} corresponds to the MERIS Top Of Atmosphere reflectance used 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 TOA reflectance and geometry.

2.5. 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 2):

- **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.

2.5.1. 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.

2.5.1.1. 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 TOA 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.

2.5.1.2. 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.

2.5.2. 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:

- **Reflectance mis-match:** This represents the consistency of the measured MERIS reflectance with that simulated in the training data base.
- **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. Description of the algorithmic elements

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 atmosphere 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 4. The 13 MERIS bands used in the algorithm. The bands appearing in grey are not used.

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

- **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. Additional atmosphere characteristics

Tests have been conducted to evaluate the interest of using additional atmosphere characteristics as inputs to the network (pressure, water vapor or ozone content). Results show that the performances are generally not improved as compared to the implicit use of these variables (Baret, Pavageau et al., 2004). It was therefore decided not to use explicitly these additional atmospheric characteristics as inputs. Additionally this eases considerably the implementation of the algorithm while inducing only marginal decrease of the performances as evaluated theoretically.

3.1.1.4. 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 $LAI \times C_{ab}$ values as described in §.2.2. The range of variation and resolution steps proposed are presented in Table 5.

Product	Unit	Minimum	Maximum	resolution
$fAPAR$	-	0	1.0	0.01
$fCover$	-	0	1.0	0.01
LAI	$m^2 \cdot m^{-2}$	0	6.0	0.01
$LAI \times C_{ab}$	$g \cdot m^{-2}$	0	500	1

Table 5. 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,
- **reflectance mis-match** indicating the consistency of the MERIS measured TOC reflectance with that of the data base used to train the operational neural networks.
- **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. Reflectance models

Physically based radiative transfer models are considering 3 main components that will be described separately in the following:

- The leaf optical properties
- The canopy structure
- The background reflectance

3.2.1. Leaf optical properties

To estimate the chlorophyll content from canopy reflectance, chlorophyll content has to be explicitly introduced into the radiative transfer model to be used. Because of its versatility and performances, the PROSPECT model (Jacquemoud and Baret, 1990) with the updated absorption coefficients proposed by (Fourty and Baret, 1997) appears therefore to be a good candidate.

Note that the PROSPECT model considers the leaf as a lambertian surface. (Sanz, Espana et al., 1997) showed that leaves were mainly characterized by a specular behaviour in addition to an important diffuse scattering process that takes place within the leaf. These authors demonstrated that, except in the specular direction, the lambertian approximation was valid in all other view directions. In addition, PROSPECT assumes that the optical properties of both leaf faces are equal.

Several authors (Fourty and Baret, 1997; Jacquemoud and Baret, 1990; Newnham and Burt, 2001) have successfully validated the model over broadleaf types. In addition, the PROSPECT model provides a reasonable description of the optical properties of the needles, even though the basic assumptions associated to the plate model are obviously violated (Zarco-Tejada, Miller et al., 2001). The following variables are required as input to the PROSPECT model:

- N leaf mesophyll structure index. It varies between 1.0 for the most compact leaves (such as young cereal leaves) up to 3.5 for thick leaves with well developed spongy mesophyll or even senescent leaves having disorganized mesophyll with large amount of air spaces.
- C_{ab} Leaf Chlorophyll content ($\mu\text{g}\cdot\text{cm}^{-2}$). It actually corresponds to the content of chlorophyll a, chlorophyll b and carotenoids (Fourty and Baret, 1997). Note that chlorophyll a and b are generally strongly correlated. The same is observed between chlorophyll a and b and carotenoids, particularly for medium to large chlorophyll content values. It basically varies between 0 to $100\mu\text{g}\cdot\text{cm}^{-2}$, although a threshold value of $15\mu\text{g}\cdot\text{cm}^{-2}$ has been proposed to consider a leaf as 'green'.
- C_{dm} Leaf dry matter content ($\text{g}\cdot\text{cm}^{-2}$). Dry matter absorbs over the whole spectral domain, and its effect will be maximal in the near infrared region. The leaf dry matter content is also called the specific leaf weight (*SLW*) which is also the inverse of the specific leaf area (*SLA*) used by physiologists. C_{dm} typically varies from 0.002 up to $0.02\text{g}\cdot\text{cm}^{-2}$.
- C_w Leaf water content ($\text{g}\cdot\text{cm}^{-2}$). Several studies showed that the relative water content could be approximated to a value close to 80% for the green leaves, and to around 20% for the senescent leaves. This allows tying the water (C_w) and the dry matter (C_{dm}) contents together.

- C_s Leaf brown pigment content (relative units). Green leaves will be considered as having both possible brown pigments. C_s typically varies from 0 for green leaves, up to 3.5 for the senescent dark brown leaves.

(Bacour, Jacquemoud et al., 2002) and (Le Maire, 2002) have analysed the sensitivity of the radiometric response both at the leaf and canopy levels. They showed that the chlorophyll content, the dry matter and the structure index are the main drivers of the optical properties in the visible to near infrared spectral domain.

3.2.2. Canopy radiative transfer models

The use of pure 3D models such as DART (Gastellu-Etchegorry, Demarez et al., 1996) or DISORD (Myneni, Asrar et al., 1992) for simulating a very large range of situations appears very appealing. Even though, the use of detailed 3D models that mimics actual canopy architecture and combined with ray tracing (Govaerts and Verstraete, 1998), (España, Baret et al., 1999) or radiosity (Gerstl and Borel, 1992), (Borel, Gerstl et al., 1991), (Chelle, Andrieu et al., 1997), (Soler, F. et al., 2001) radiative transfer description and applied to a representative sample of biomes and conditions would be ideal. However, it might be difficult to implement for two practical reasons:

- The necessity to describe a very large range of realistic canopy architectures. This requires a huge effort in canopy architecture and optical properties measurements at the ground level.
- The computer time required to run these models is generally important, and would therefore strongly limit the number of possible cases to be simulated.

We thus propose to use a reflectance model that is computer efficient and exploits a small number of input variables. The SAIL radiative transfer model (Verhoef, 1984; Verhoef, 1985) is widespread in the remote sensing community for the estimation of vegetation biophysical variables. The canopy is described as an homogeneous medium where leaves are randomly distributed. The SAIL model uses a limited number of structural variables in addition to leaf reflectance and transmittance and soil back ground reflectance.

- Leaf area index (LAI),
- the average leaf angle (ALA), characterizing the leaf angle distribution that will be described by an ellipsoidal distribution (Campbell, 1986). Note that a spherical distribution corresponds to an average angle close to 57° ,
- the hot spot parameter (HOT) ((Kuusk, 1985)),

To better account for the fact that MERIS is generally observing over heterogeneous areas because of its medium spatial resolution, mixed pixels made of pure vegetation and background fractions are considered. The vegetation cover fraction ($vCover$) variable is introduced to describe the fraction of ground covered by pure vegetation. A pixel is therefore considered to be composed of a fraction of bare soil (of the reflectance R_b) and of a fraction $vCover$ of vegetation (reflectance R_{veg}) over a background of reflectance R_b (Figure 3). The reflectance of the composite scene expresses as:

$$R_{TOC} = R_{veg} \times vCover + R_b \times (1-vCover) \quad \text{Equation 1}$$

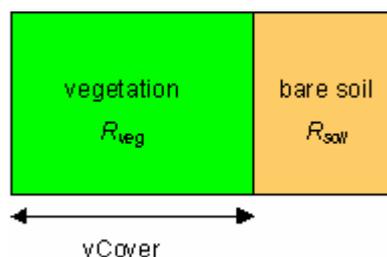


Figure 3: Compositing of the simulated scenes

3.2.3. Background reflectance model

The background reflectance corresponds to all the non green materials that constitute the last bottom layer in the canopy. Following the definition of the LAI , $fAPAR$, $fCover$ and $LAIxCab$ variables, all the green vegetation layers have to be accounted in the computation of these variables. Therefore, the understory, if green (including lichens and moss), will not be considered as the background here. It will be included within the green vegetation layer. The background reflectance may thus correspond to soil or litter. Water and snow cases will not be considered here, since the corresponding pixels would be flagged as water bodies. Note that the reflectance mismatch criterion might also help discarding these cases.

3.2.3.1. The background brightness concept

The background reflectance, for a given wavelength, will depend on the background type (soil type, litter), geometrical illumination and view conditions (Ω), roughness (z) or moisture (H). The approach used here to describe the background reflectance properties is based on the brightness concept allowing confounding the effect of geometrical conditions, roughness and moisture within a single parameter that will be assumed independent from wavelength.

The background reflectance $\rho_b(\lambda, \Omega_i, H_j, z_k)$ for any wavelength λ , observation geometrical configuration Ω_i , moisture H_j and roughness z_k is assumed proportional to the reflectance background for the same wavelength λ but different observation geometrical configuration Ω_l , moisture H_m and roughness z_n :

$$\rho_b(\lambda, \Omega_i, H_j, z_k) = Bs \cdot \rho_b(\lambda, \Omega_l, H_m, z_n) \quad \text{Equation 2}$$

where Bs is a brightness parameter that does not depend on wavelength λ , but depends on all the other factors (Ω , H , z). This convenient property is a consequence of the well known soil line concept (Baret, Jacquemoud et al., 1993) stating that a linear relationship exists between the reflectance of soils (and litter) in two wavelengths λ_1 and λ_2 when either the roughness, moisture or illumination or view directions vary:

$$\rho_b(\lambda_1, \Omega_i, H_j, z_k) = a(\lambda_1, \lambda_2) \cdot \rho_b(\lambda_2, \Omega_i, H_j, z_k) + b(\lambda_1, \lambda_2) \quad \text{Equation 3}$$

This property could be written for an other set of sun and view directions:

$$\rho_b(\lambda_1, \Omega_l, H_m, z_n) = a(\lambda_1, \lambda_2) \cdot \rho_b(\lambda_2, \Omega_l, H_m, z_n) + b(\lambda_1, \lambda_2) \quad \text{Equation 4}$$

Replacing in Equation 3 $\rho_b(\lambda_1, \Omega_i, H_j, z_k)$ and $\rho_b(\lambda_2, \Omega_i, H_j, z_k)$ by their expression derived from Equation 2:

$$Bs \cdot \rho_b(\lambda_1, \Omega_l, H_m, z_n) = a(\lambda_1, \lambda_2) \cdot Bs \cdot \rho_b(\lambda_2, \Omega_l, H_m, z_n) + b(\lambda_1, \lambda_2) \quad \text{Equation 5}$$

Identifying Equation 5 to Equation 4 provides the condition under which Equation 2 is valid:

$$Bs \cdot b(\lambda_1, \lambda_2) \approx b(\lambda_1, \lambda_2) \quad \text{Equation 6}$$

which is true either for $Bs \approx 1$ or $b(\lambda_1, \lambda_2) \approx 0$. Experimental and theoretical results (Baret, Jacquemoud et al., 1993) show that the soil line intercept, $b(\lambda_1, \lambda_2)$, is generally very small in comparison to the background reflectance value. For example, in the red and near infrared bands, $0 < b(\lambda_{red}, \lambda_{nir}) < 0.1$. Similarly, experimental evidences (Liu, 2003, results to be published), when referring to a standard situation (dry soil, medium roughness, no hot-spot configuration), $0.3 < Bs < 1.3$. Therefore, the brightness concept is generally valid and has already been used extensively in past studies (Weiss, Baret et al., 2002); Bacour, Jacquemoud et al. 2002).

The brightness concept allows describing the spectral variation of a given background when the geometrical configuration, moisture or roughness varies with two inputs:

- The brightness parameter (Bs) that is independent on wavelength

- Reference reflectance spectra for any given other geometrical configuration, moisture and roughness. These reference spectra should represent the variability encountered over the Earth surface.

3.2.3.2. Background spectral variation

The reference soil spectra will be derived from a soil reflectance data base available at INRA Avignon representing a large variation of soil types, moisture, roughness and geometrical configurations (Jacquemoud, Baret et al., 1992; Liu, Baret et al., 2002b). Considering the brightness concept will allow increasing the diversity in actual soil properties. The measurements were performed using an ASD Fieldspec Pro spectrophotometer providing a 1nm spectral sampling close to the spectral resolution in the visible and near infrared domains. 460 soil reflectance spectra are available.

To reduce the number of reference soil used while keeping the variability observed between soil properties, an iterative selection process was developed to evaluate the reconstruction performances of any soil reflectance spectra of the data base using a subset of N selected reference soil spectra. For each set of N reference soils investigated, $1 < N < 40$, an automatic classification with N classes was made to identify the corresponding N reference spectra. This was achieved over the 460 reflectance spectra data base that were normalized so that each individual soil spectra has the same average reflectance value. This averaging process was applied in order to retain mainly reference soil spectra that differ by their spectral features rather by their brightness. The classification process will ensure that the N selected reference spectra are representing most of the variability in terms of spectral features. Then, for each of the $(460-N)$ remaining soil spectra, a reference spectra among the N ones is selected, and the Bs coefficients adjusted to get the best match. This process is replicated over the N set of reference spectra to select the one that approximates the best the remaining soil considered with the adjusted Bs parameter. Finally, this is applied for all the remaining soils and then by investigating the cases for N varying from 1 to 40.

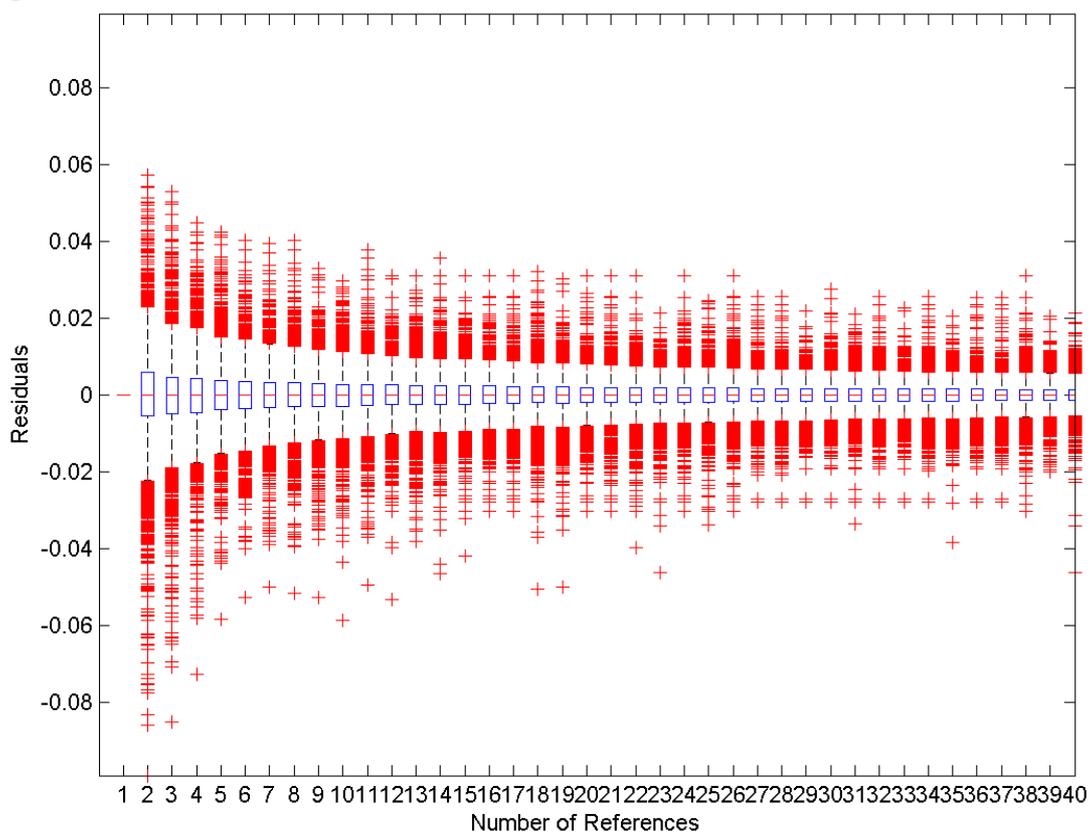


Figure 4. Distribution of the residuals when reconstructing all the soils (460 spectra) with a subset of N ($1 < N < 40$) reference soils and a brightness coefficient. The box contains 50% of the cases, the whiskers 99%. The red 'plus' are outliers.

Results (Figure 4) show that the accuracy of the reconstruction as measured by the residuals decreases continuously with the number of reference spectra used as expected. We then selected 5 reference soils for our simulations. This number provides a reasonable compromise between a small number of soils required for the simulations, and a reasonable accuracy in the representation of the variability of the soil reflectance spectra as illustrated by the 460 soils available in our data base. Figure 5 shows the 5 soil reflectance spectra finally selected.

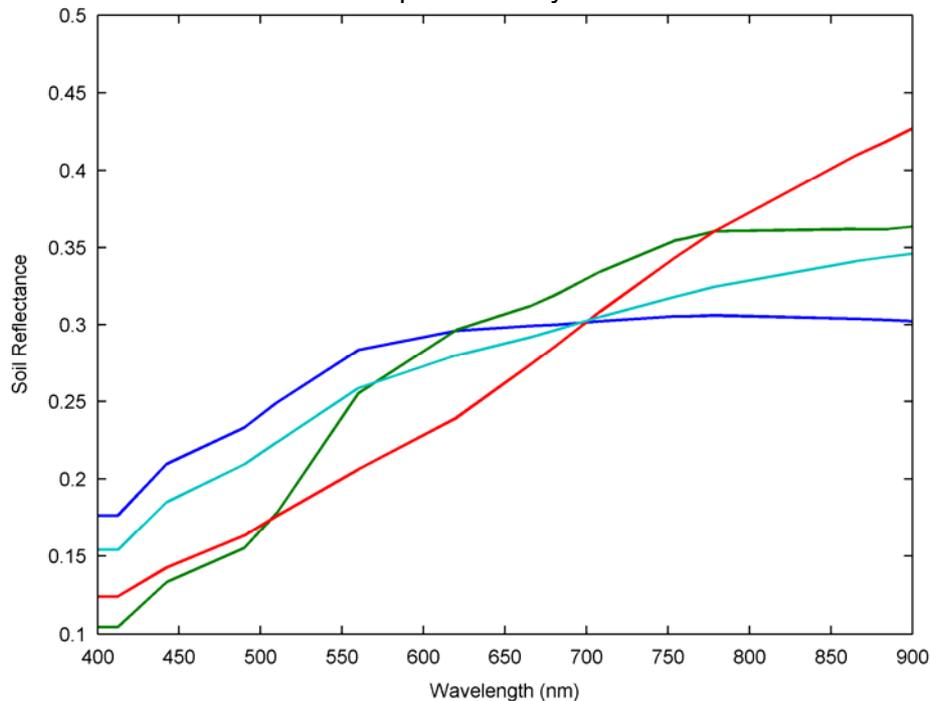


Figure 5. The 5 standard soil reflectance spectra in MERIS wavebands, as measured by (Liu, Baret et al., 2002a). Note that here the soils were normalized to the average soil reflectance values.

Litter and vegetation residues. The litter corresponds to an important background, particularly over forest areas. The spectral signature of litter is very close to that of the soil as noticed by several studies (Asner, Wessman et al., 1998). Crop residues and natural vegetation residues may have also important contribution to the reflected signal during specific seasons. Similarly to litter, the reflectance of vegetation residues is also very similar to that of soil background (Gausman, Gerbermann et al., 1975), (Biard and Baret, 1997), (Chen and McKyes, 1993). Because of the similarity between litter, residues and soil reflectance, these will finally be aggregated within the 'soil' background category.

3.2.4. Atmosphere model

Among the several atmosphere irradiative transfer models, the SMAC code (Rahman and Dedieu, 1994) was selected for the good compromise it provides between the realism of the simulations, the relatively small number of inputs and the computation requirements. SMAC is actually a parametric version of the 6S code (Vermote, Tanré et al., 1997). The coefficients for the SMAC model have been adjusted specifically for the spectral characteristics of MERIS. The inputs of the SMAC model are the atmospheric pressure, the aerosol type (only continental and desertic aerosols are currently available), the aerosol optical thickness at 550nm (AOT), the ozone and water vapour contents and the sun and view directions. The coupling between the surface reflectance model and SMAC is achieved according to the following equation:

$$\rho_{TOA} = t_g \cdot (\rho_{atmo}(X) + \frac{T_s(X)T_v(X)\rho_{TOC}}{1 - S(X)\rho_{TOC}})$$

Where ρ_{TOA} and ρ_{TOC} represent respectively the top of atmosphere and top of canopy reflectance, ρ_{atmo} is the contribution due to atmospheric scattering, T_s and T_v are the transmittance in the sun and view directions, and S is the spherical albedo of the atmosphere. The vector X represents the

vector of input variables of the SMAC model: the atmospheric pressure at the surface, the ozone and water vapour contents, the aerosol type (desertic or continental) and the AOT at 550 nm. The coupling scheme between the atmosphere and the surface was approximated by assuming a lambertian surface with a reflectance value equal to that of the bidirectional reflectance computed for the actual sun and view directions. This approximation is known to introduce only second order errors.

3.3. 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 5. 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. Algorithm prototyping

The prototyping corresponds mainly to the calibration of the algorithm over the training data base. The requirements for generating the training data base are the following:

- Having a representative sampling of all the vegetation types and conditions that can be observed over the Earth's surface
- Each case should be properly weighed by its frequency of occurrence to avoid putting too much emphasis on very scarce cases.
- Using as much as possible prior information while letting enough flexibility so that the radiometric information is still worth!
- The training data base should be large enough to get a reasonably well populated space of canopy realization. This should yield robust training results.

4.1. Generation of the training database

The training data base is generated in three steps:

- Generation of the data base containing the input radiative transfer model variables
- Generation of the corresponding top of atmosphere reflectance for the 13 MERIS bands considered
- Addition of uncertainties to the simulated top of atmosphere reflectance values previously simulated.

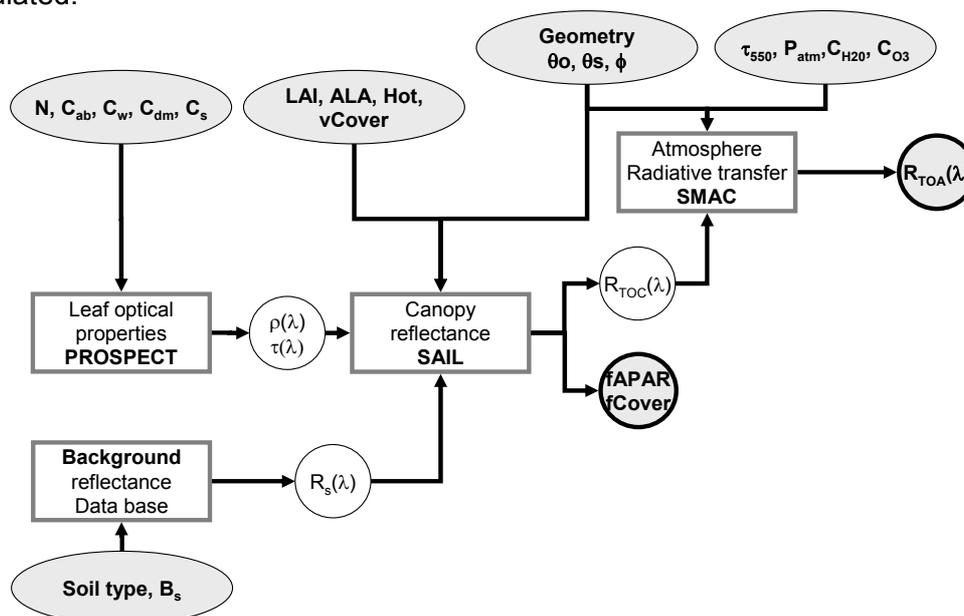


Figure 6: The coupled PROSPECT+SAIL+SMAC models used to generate the training database made of TOA reflectances and corresponding biophysical variables.

4.1.1. Radiative transfer model

The top of atmosphere reflectances of the training database is made of are simulated with the PROSPECT+SAIL+SMAC model described earlier Figure 6. The coupled model allows as well deriving the secondary biophysical variables $fCover$ and $fAPAR$ (from the computation of the absorptance by green elements). The inputs of the model are:

- the geometrical configuration of illumination and observation (i.e. the solar and view zenith angles, θ_s and θ_v , and the relative azimuth), that derives from MERIS orbit mechanics and swath, for a given date and location,
- the background reflectance spectrum, as described earlier,
- the primary biophysical variables related to leaf optical properties (N , C_{ab} , C_w , C_m , and C_{bp}) and to the canopy structure (LAI , ALA , HOT , and Bs),
- The atmosphere characteristics (P_{atm} , τ_{550} , C_{H_2O} , C_{O_3}).

4.1.2. Generation of the vegetation input variables database

This is the most delicate step in the generation of the training data base. As a matter of fact, the training data base has to reflect the actual distribution of the vegetation types over the Earth's surface. The distribution of the variables will be constrained by the knowledge of date and location of the observation. The distribution of date and location will be described in the following section. Table 5 presents the range of variation and the actual distribution used for the input variables of the vegetation and background.

The water content was tied to the dry matter content assuming that the green leaves have a relative water content close to 80%: $C_w=4.C_m$. Moreover, leaf water should have only marginal effect on TOC reflectance in the 13 MERIS wavebands. The distribution of the $vCover$ variable is set uniform. Leaf area index values are inferred from the ECOCLIMAP LAI climatology (Masson, Champeaux et al., 2003).

The training data base has to be sufficiently large to allow a robust calibration of the network, and also get a sub-set of the data base for hyper-specialization and test. The optimal size of the training data base depends on the complexity of the problem to solve. Previous studies (Combal, Baret et al., 2002) have shown that for a medium complexity problem, a training table close to 10 000 cases was satisfactory. In the current case that corresponds to a more complex algorithm. the size of the training data base should increase.

Variable	Min	Max	Mode	Std	Nb. Class	Law
Day of the year Year 2003	1	182	-	-	1	uniform
Latitude (°)	-60	60	-	-	1	uniform
Longitude (°)	0	360	-	-	1	uniform
Leaf-Area-Index (m ² /m ²)	0	6	-1	4	5	gauss
Average Leaf inclination Angle (°)	30	85	60	20	3	gauss
Hot-Spot parameter (HotS)	0.001	1	0.1	0.3	1	gauss
Mixed Pixels (vCover)	0	1	1	0.3	3	gauss
Chlorophyll Content (µg/cm ²)	30	100	50	30	5	gauss
Dry matter content (g/cm ²)	0.002	0.02	0.0075	0.0075	2	gauss
Relative water content	0.65	0.9	0.8	0.05	2	gauss
Brown Pigment index	0	1.5	0.1	0.3	3	gauss
Mesophyll structure (N)	1	2.5	1.5	1	3	gauss
Brightness Parameter (Bs)	0.5	2.5	1	0.5	4	gauss
Pressure (mg/cm ²)	600	1080	918	200	1	gauss
Aerosol optical thickness (g/cm ²)	0	0.8	0.35	0.3	4	gauss
Water vapour content	1	6.5	3.8254	1.5	1	gauss
Ozone	0.2	0.5	0.35	0.1	1	gauss

Table 6. Distribution of the input variables of the radiative transfer model used to generate the training data base. Truncated Gaussian distributions are used, characterized by the mode, the standard deviation (Std), and the lower (Min) and upper bounds (Max).

The sampling scheme is based on a full orthogonal experimental plan (Bacour, Jacquemoud et al., 2002). This consists to identify classes of values for each variable. Then all the combinations of classes are sampled once. Finally the actual values of each variable are randomly drawn within the range of variation defined by the corresponding class, according to the distribution law specified for the variable considered. This process allows accounting for all the interactions, while having the range of variation for each variable densely and near randomly populated. The number of classes (equally spaced) for each variable is shown in Table 6.

1. **Geometry.** The geometry of observations is defined by the location (latitude and longitude) and day of the year. All these input variables were drawn randomly from the whole range of possible variation. Note that only half of the year is considered, the geometry of observation for the other half being symmetrical. Note also that the maximum latitude is 60°. Beyond these latitudes, there are actually very little vegetation and illumination conditions are generally poor.
2. **Canopy variables.** The following distributions were used (Table 5 and **Erreur ! Source du renvoi introuvable.**):
 - **LAI:** The LAI values are randomly drawn from a truncated Gaussian distribution that enhances cases with low leaf area index. The 'local' LAI value corresponding to the vegetated fraction of the mixed pixel is computed as: $LAI_{pure} = LAI_{pixel} / vCover$. Cases with $LAI_{pure} > 8.0$ are eliminated.
 - **ALA, HOT and vCover:** The average leaf inclination angle is assumed to follow a truncated Gaussian distribution.
3. **Leaf optical properties.** Here also, very little knowledge is available on the actual leaf characteristics. Truncated Gaussian distributions were used for all these variables (Table 6).
4. **Atmosphere characteristics.** Because very little information is available on the co-distribution of the variables, they were supposed independent and following the truncated Gaussian laws as described in Table 6

129 600 sets of input variables were available (Table 6).

4.1.3. Simulation of the top of atmosphere reflectance for the 13 MERIS bands

The previously derived table of input variables is used to simulate the corresponding MERIS top of atmosphere reflectance in the 13 MERIS bands using the PROSPECT+SAIL+SMAC model. A relative uncertainty corresponding to a 3% Gaussian noise with no bias was added to account for the sensor overall radiometric performances. It corresponds roughly to MERIS performances as evaluated over vicarious calibration exercises. Note that the spectral calibration uncertainties will not be accounted for because they are small and applies on relatively smooth reflectance spectra. Additionally, the $fAPAR$ for sun position at 10:00 solar time, $fCover$ and $LAI.C_{ab}$ values were finally computed.

The uncertainties attached to the radiative transfer model mainly derive from the representation of canopy architecture, leaf and soil background optical properties which is difficult to evaluate. A posterior estimation will be issued using the reflectance mis-match criterion as computed over actual MERIS data.

4.1.4. Streamlining the training data base

The training data base should represent realistic cases. Two tests were made to ensure the consistency of the reflectance simulations with actual observations:

4.1.4.1. Reflectance mis-match.

The reflectance mismatch expresses the difference between a given MERIS L1B reflectance spectra and the one the closest as simulated in the data base. For this purpose, a compilation of

actual MERIS L1B reflectance spectra was achieved from using a large set of contrasted images. Then each reflectance spectra in the training data base was compared to the closest one in the actual MERIS L1B data base. The RMSE value is computed, and if the RMSE value is larger than the mismatch threshold, the reflectance spectra in the data base is rejected. Figure 7 shows that most simulated reflectance spectra is very close to a reflectance spectra actually measured by MERIS. The mismatch threshold was set to 0.02, which leads to eliminate about 5% cases in the data base.

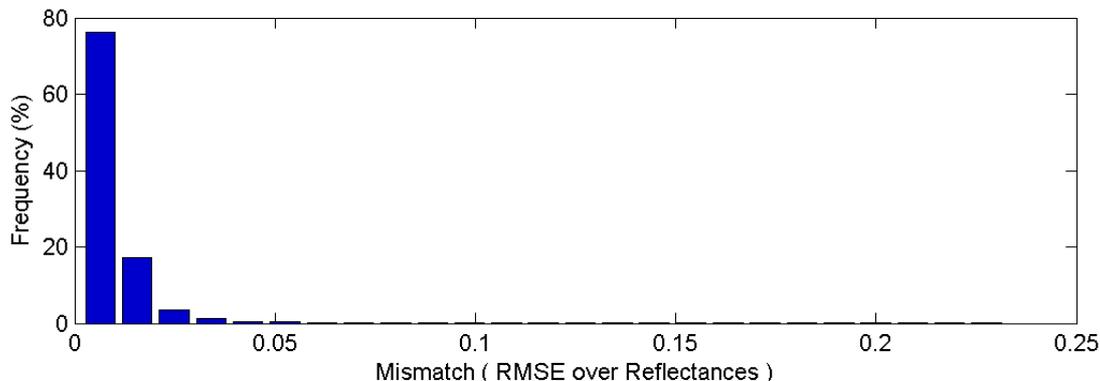


Figure 7. Distribution of the reflectance mismatch values (RMSE between a reflectance spectra in the training data base and the closest one in the MERIS L1B actual reflectance database).

4.1.4.2. LAI/fAPAR streamlining

A relatively well defined *LAI/fAPAR* relationship is expected. However, the use of the *vCover* factor, although necessary because of the mixed nature of the medium resolution pixels introduces large scattering the *LAI/fAPAR* relationship generally not observed. This also induces difficulties in the learning of the algorithm for the higher *LAI* values. For this reason, data showing too much scatter will be discarded. In our case, after several trials, we decided to discard all the data for which the *fAPAR* value was below a threshold value that depends on the *LAI* value (Figure 8). This resulted in the elimination of 24% additional cases.

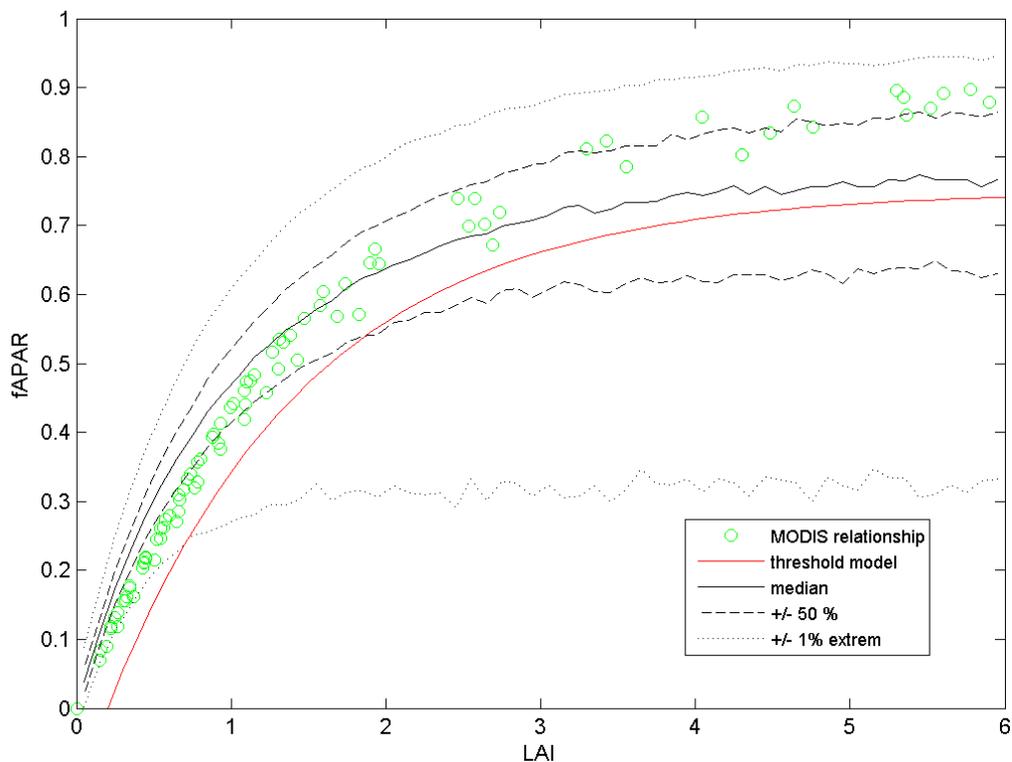


Figure 8. Relationship between LAI and fAPAR. The green dots correspond to the MODIS simulated values according to the ATBD (Knyazikhin, Glassy et al., 1999). The black lines

correspond to the median (solid line) and 50% and 98% of the data. The red line corresponds to the threshold value below which the data are eliminated.

This streamlined data set was finally split in three parts with a random selection process:

- **Training:** $\frac{1}{2}$ of the simulations are affected randomly to the training of the neural network
- **Hyper-specialization:** $\frac{1}{4}$ of the simulations are used for the hyper-specialization control
- **Testing:** $\frac{1}{4}$ are used for the theoretical evaluation of the algorithm

4.1.5. Realism of the simulated reflectances.

Apart from the reflectance mismatch that was showing that the reflectance simulations were individually well matching actual MERIS observations, inspection of the distribution laws of the reflectance in each band will also provide confidence on the simulated values. Figure 9 shows that the range of variation as well as the distribution of the simulated reflectance are roughly in agreement with those observed over actual MERIS observations.

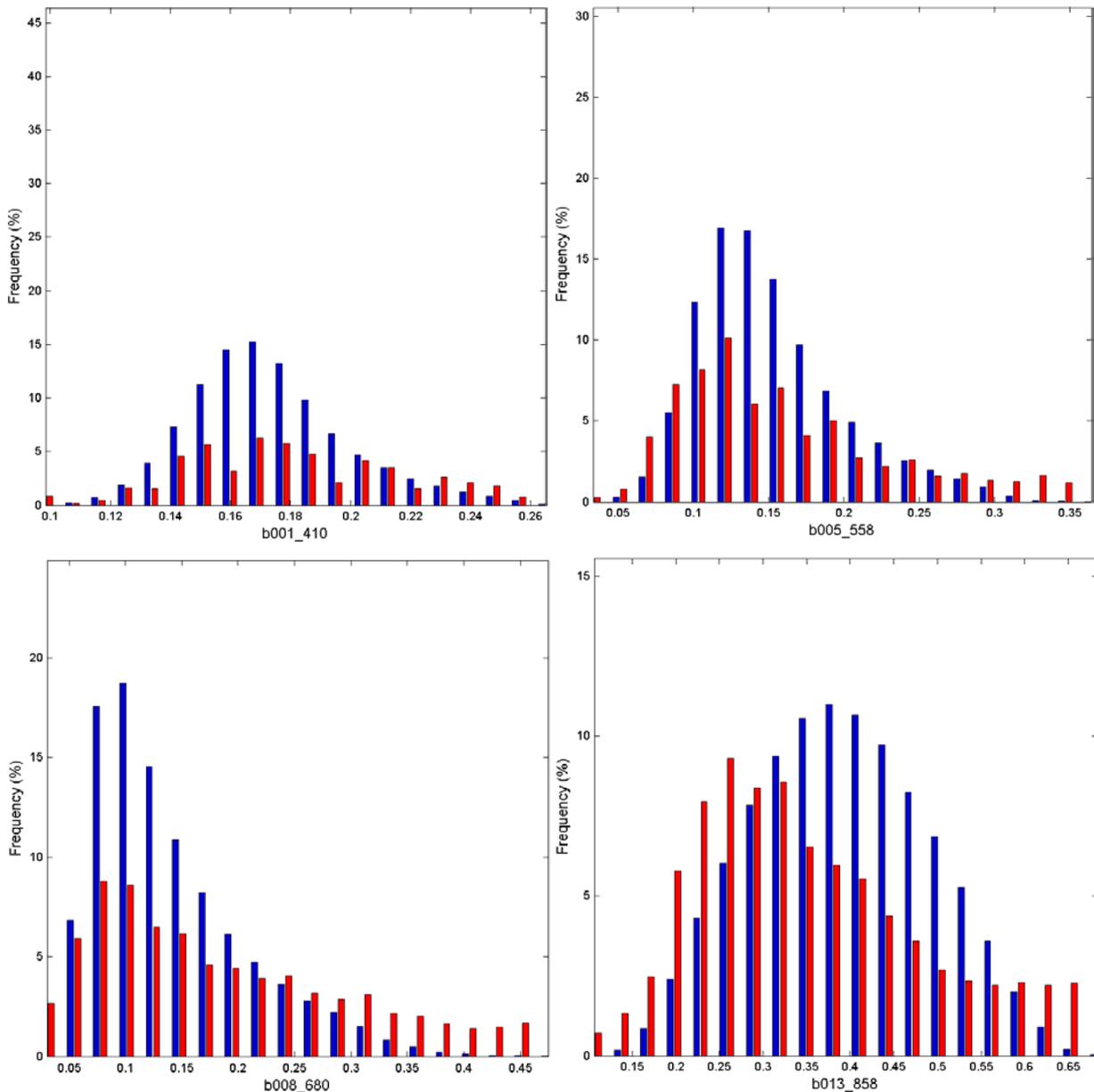


Figure 9. Distribution of the reflectance values as simulated from the training data base (blue) and from the actual MERIS images (red).

4.2. Training the neural network

Neural networks are defined mainly by the type of neurons used (the transfer function), the way they are organized and connected (the network architecture) and the learning rule. In addition, the input and output values need to be properly normalized to prevent any scaling factor or numerical problem. Back-propagation artificial neural network (Rummelhart, Hinton et al., 1986) is one of the most common neural networks used to solve our radiative transfer model inversion problem.

4.2.1. Normalization of the input and output values

The input (MERIS TOA reflectance in 13 bands and geometry) and output (the biophysical variable considered) values are first standardized according to Equation 10 so that 80% of the values fall within [-1;1]. Such data transformation is performed mainly to increase the performances of convergence of the training algorithm.

$$x^* = 2 \cdot \frac{x - \min(x)}{\max(x) - \min(x)} \quad \text{Equation 7}$$

4.2.2. Network architecture

The connections between neurons are associated to a “synaptic” weight. Each neuron transforms the sum of the weighted signal from the previous neurons according to a given transfer function and a bias. The combination of sigmoid and linear functions is recognized as capable of fitting any type of function (Demuth and Beale, 1998).

For our more complex problem, an optimal architecture had to be determined for each biophysical variable. Several network structures have been tested. For each possible structure, three neural networks, differing by the initialization of their coefficients, have been trained. The selection of the "optimal" network architecture is then based on the RMSE between the outputs and the "true" biophysical variables as well as on the number of coefficients to be adjusted. Lower numbers are preferred because they allow faster runs of the neural networks in operational mode while precluding hyper-specialization.

The neural networks investigated that way are thus composed of:

- one input layer made of the 16 normalized input data (θ_s , θ_v , $\cos(\phi)$, and the TOC reflectances in the 13 MERIS wavebands).
- two hidden layers with tangent sigmoid transfer functions.
- one output layer with a linear transfer function.

Results show that 2 hidden layers perform generally slightly better than just a single one. The optimal architecture of the networks did not vary much between variables as well as if one or two neurons are added or subtracted from the first or the second layers. The first layer has 10 neurones, the second one between 5. This makes a total of 231 coefficients to tune, corresponding the ratio with the number of cases used in the training data base around 4000, which is much larger than the 100 value proposed by (Harrel, 2001).

4.2.3. Learning process

The learning process is mainly made of two elements: the training dataset that was described earlier and the learning rule that is now described. The Levenberg–Marquardt optimization algorithm is used to adjust the synaptic weights and neuron bias to get the best agreement between the output simulated by the network and the corresponding value of canopy biophysical variable simulated in the training data base. The initial values of the weights and biases were set to a random value between -1.0 and +1.0. To prevent from hyper-specialization, a sub-set of the training data base is used to control whether the network starts to hyper-specialize, i.e. represents the particular features of the training data set and therefore loosing its capacity to describe the general features targeted. When this starts to happen, the optimization process is stopped.

Five networks were trained in parallel to retrieve the canopy biophysical variables, each corresponding to independent random drawing of the initial values of the synaptic weights and bias. The network finally selected is the one providing the best performances over the test data set.

4.3. Theoretical performances of the artificial neural network

The theoretical performances of the networks were evaluated over the test data set (Figure 10, Figure 11, Figure 12 and Figure 13) for the four biophysical variables considered. Results show that the ANN performances for *fAPAR* and *fCover* are better than those observed for *LAI* and *LAI.C_{ab}*: In the case of *fAPAR* and *fCover*, the maximum scattering is observed for the medium values of the variables as expected for these type of variables. On the contrary, a saturation effect is noticeable for the estimation of for *LAI* above 4 and *LAI.C_{ab}* above 300 $\mu\text{g.cm}^{-2}$. The estimation residuals show unbiased behavior for these two variables when compared with the estimated values, which is not the case when looking at the 'true' value.

It is interesting to investigate the effect of the learning process on the relationship between LAI and *fAPAR*. **Figure 14** Shows that the variability in LAI/*fAPAR* relationship is much reduced after the training process. It shows also that above LAI of 3, the 'learned' relationship depart significantly from that of the training data set. This is obviously due to the reduced sensitivity for the larger LAI values.

To provide a first estimate of the theoretical uncertainties, the errors, approximated as the rmse value, was analyzed as a function of the product value. A second degree polynomial was adjusted to best describe the uncertainties as a function of the estimated variable value. Results (Figure 10, Figure 11, Figure 12 and Figure 13) confirm that the uncertainty varies only slightly with the variable value for *fAPAR* and *fCover*, with RMSE values around 0.06-0.08. However, for *LAI* and *LAI.C_{ab}*, the uncertainties increase rapidly with the value of the variable, justifying the use of a relative uncertainty that is around 25-30%. This is mainly due to the physics of the radiative transfer, where the sensitivity of reflectance to *LAI* or *LAI.C_{ab}* decreases when *LAI* or *LAI.C_{ab}* increase.

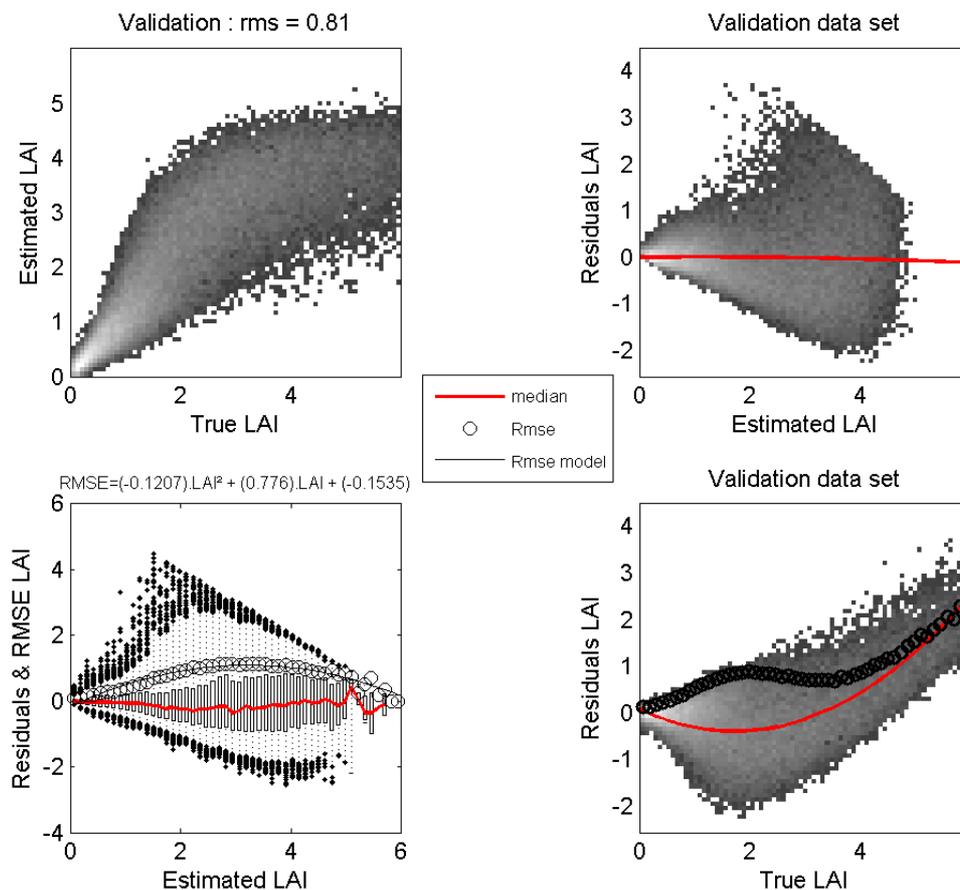


Figure 10: Theoretical performances for *LAI* estimation from the neural network.

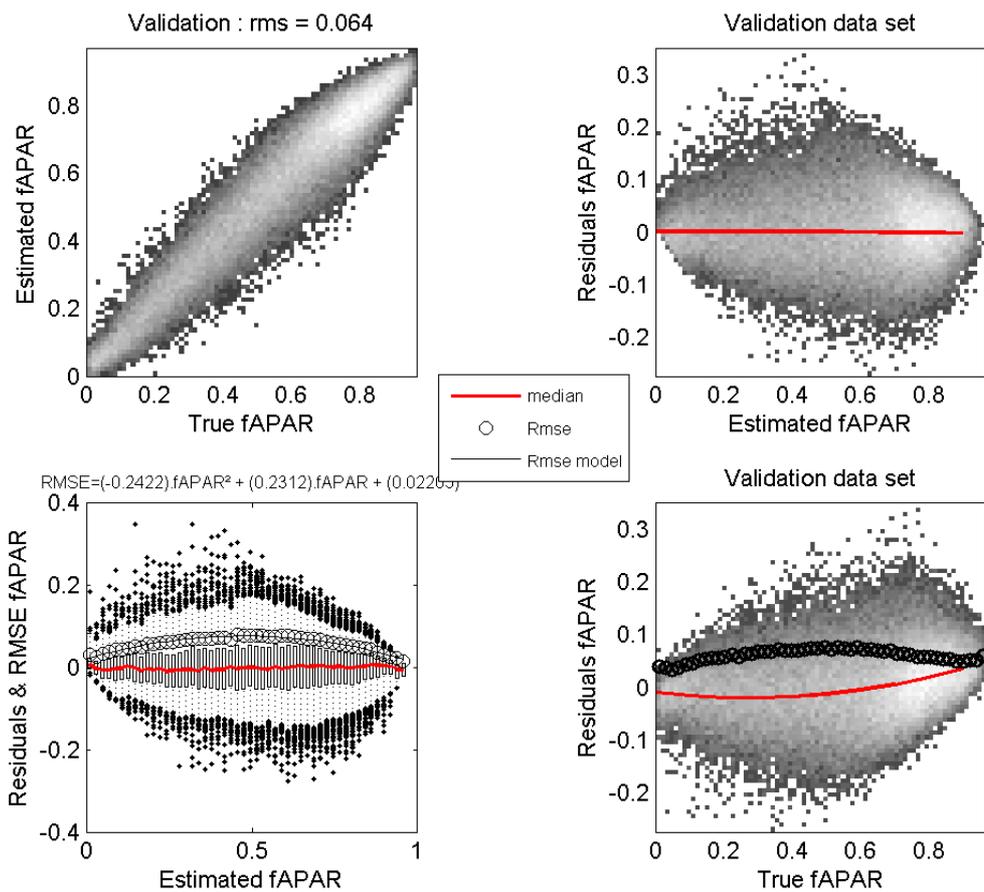


Figure 11: Theoretical performances for *fAPAR* estimation from the neural network.

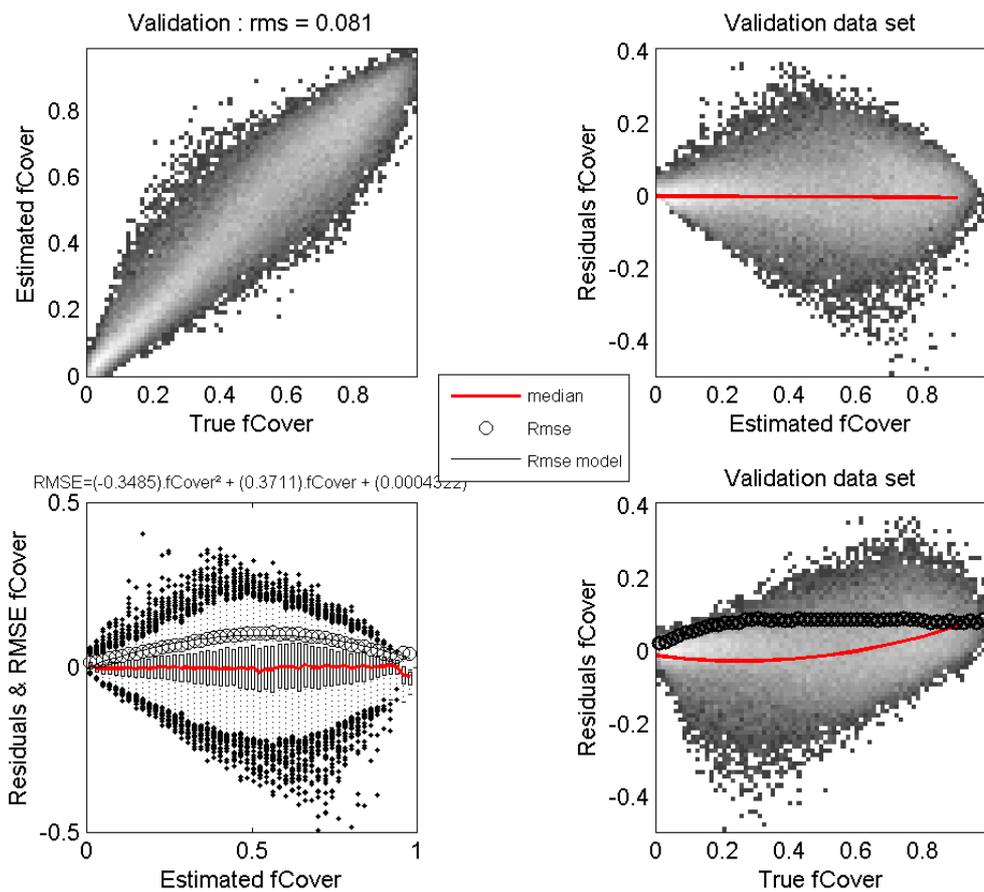


Figure 12: Theoretical performances for *fCover* estimation from the neural network.

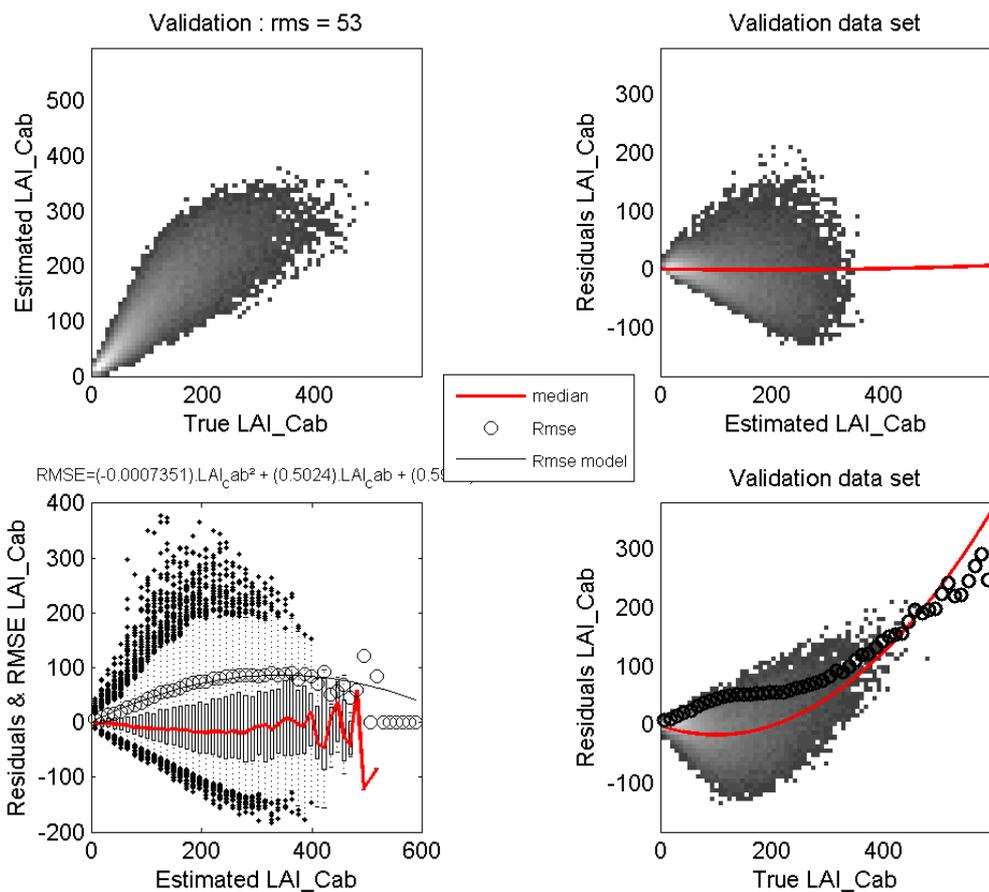


Figure 13: Theoretical performances for $LAI.C_{ab}$ estimation from the neural network.

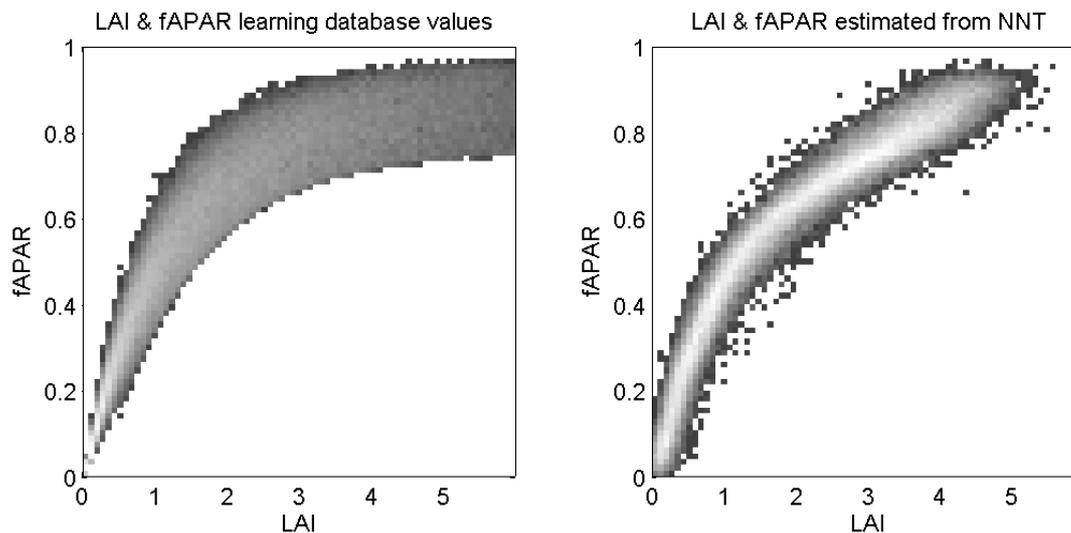


Figure 14: Relationship between LAI and fAPAR. Left: training dataset; right: results after the learning process.

5. Conclusion

This ATBD provides a description of the TOA_VEG algorithm used to compute $fAPAR$, $fCover$, LAI , and $LAI.C_{ab}$ products directly from MERIS top of atmosphere reflectance data both at full and reduced spatial resolution. Evaluation shows that the theoretical performances are slightly improved as compared to the classical approach based on the top of canopy reflectances as used as inputs to the network (see the corresponding ATBD (Baret, Bacour et al., 2005)). Therefore, because of possible errors in the atmospheric correction step necessary to get the TOC reflectances, the proposed TOA_VEG would have even better performances.

The performances of this TOA_VEG algorithm were evaluated on 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).

6. 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 Environment*, 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 (TOA_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 Conference 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://eosps0.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 Mockette, 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 characteristics 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.