S2ToolBox Level 2 products LAI, FAPAR, FCOVER

Marie Weiss, Frederic Baret, Sylvain Jay

▶ **To cite this version:**

Marie Weiss, Frederic Baret, Sylvain Jay. S2ToolBox Level 2 products LAI, FAPAR, FCOVER. [Research Report] EMMAH-CAPTE, INRAe Avignon. 2020. hal-03584016

HAL Id: hal-03584016
https://hal.inrae.fr/hal-03584016

Submitted on 22 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

S2ToolBox Level 2 products: LAI, FAPAR, FCOVER

Version 2.0

Book Captain: Marie Weiss (INRAE)

Contributing Author: Fred Baret (INRAE)

Sylvain Jay (INRAE)

Change Record

Issue/Rev	*Date*	*Page(s)*	*Description of Change*
1.0	30.04.2016	All	Release of Draft 1
1.1	03.05.2016	§ 2.1.4 § 2.1.5 § 3.3.4.1 § 3.4.4 Table 4 Table 10	Specifications and theoretical performances of neural networks trained for Canopy Chlorophyll Content and Canopy Water Content retrieval were added.
1.1	03.05.2016		Include a sentence in the conclusion concerning the possibility to use the neural networks at lower spatial resolution than 20m.
2.0	24.08.2016		Addition of neural networks for 10m resolution (using only the 3 10m bands). Addition of neural networks for SENTINEL-2B Include a new flag about pixel heterogeneity

TABLE OF CONTENTS

1 Background of the document 9
1.1 Executive Summary 9
1.2 Scope and Objectives 9
1.3 Content of the document 9
1.4 Symbols and Acronyms 10
2 Algorithm Justification and Overview 12
2.1 Products and their definitions 12
2.1.1 Leaf Area Index (LAI) 12
2.1.2 FAPAR 12
2.1.3 Cover fraction (FVC) 12
2.1.4 Canopy chlorophyll content (CCC) 13
2.1.5 Canopy water content (CWC) 13
2.2 Instrument characteristics 13
2.3 Justification for the algorithm selection and design 14
2.4 Algorithm outline 16
2.4.1 Training the neural network 17
2.4.2 Operational use of the neural network 17
3 Algorithm Description 18
3.1 Inputs and outputs 18
3.1.1 Inputs 18
3.1.2 Outputs 20
3.2 Reflectance models 20
3.2.1 Leaf optical properties 21
3.2.2 Canopy radiative transfer models 22
3.2.3 Background reflectance model 22
3.3 Generation of the training database 24
3.3.1 Radiative transfer model 24
3.3.2 Distribution of the vegetation input variables 25
3.3.3 Actual generation of the inputs for the training data base 31
3.3.4 Consequence on the distribution of the output variables 38
3.4 Training the neural network 39
3.4.1 Normalization of the input and output values 39

3.4.2 Network architecture 39
3.4.3 Learning process 41
3.4.4 Theoretical Performances 41
3.5 Field of application of the NNET Algorithm 47
3.5.1 Definition Domain 48
3.5.2 Characterization of the heterogeneity of the pixel environment 50
4 Practical considerations for the algorithm implementation 54
4.1 Requirements for easily updating the algorithm 54
4.2 Algorithm Implementation 54
5 Conclusion 56
6 References 57

List of Figures

Figure 1. SENTINEL2A and 2B MSI spectral from ESA SENTINEL online web site. 14

Figure 2. Flow chart showing how the products ($\hat{V}$) are generated operationally. ANN corresponds to an Artificial Neural Network characterized par its structure and its coefficients (synaptic weights and bias); R_{TOC} corresponds to the SENTINEL2 Top Of Canopy 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 SENTINEL2 TOC reflectance and geometry. 17

Figure 3. Comparison between the performances of a global neural network versus interpolated ones. Top: theoretical performances against a simulated test database, Bottom: scatterplot between interpolated and global neural network LAI estimates in the test database. Two interpolated neural networks are considered: 1°step or 5° step. 20

Figure 4: The coupled PROSPECT+SAIL model to generate the training database made of TOC reflectances and corresponding biophysical variables. 24

Figure 5: Distribution of leaf Chlorophyll content as described by Feret et al (2008). Values of Chlorophyll include chlorophylls a and b as well as carotenoids. Distribution in blue corresponds to data mainly from tropical forests (Hawai data set in Feret's paper). Distribution in red corresponds to other species. 26

Figure 6: On the left, distribution of dry matter contents (Cdm) derived from literature review: (Prado and De Moraes 1997; Poorter and Evans 1998; Reich, Ellsworth et al. 1998; Poorter and De Jong 1999). On the right, distributions reported by Feret et al (2008): distribution in blue corresponds to data mainly from tropical forests (Hawai data set in Feret's paper). Distribution in red corresponds to other species. 26

Figure 7. RMSE values associated to the reconstruction of all the soils investigated for SENTINEL2 bands (left) and distribution of the residuals when considering 7 soils. 27

Figure 8. The 7 standard soil reflectance spectra in SENTINEL2 spectral domain as measured by Liu et al (2002). 27

Figure 9. Distribution of brightness values as observed over Liu's data base (Liu et al., 2002) when using 7 reference spectra. 27

Figure 10. Distribution of leaf area index from the 1008 ground LAI measurements provided by Scurlock et al (2002) 28

Figure 11. Effective LAI distribution over the VALERI sites 29

Figure 12. Distribution of the average leaf angle (ALA) over the VALERI sites 30

Figure 13. Relative leaf size (hot spot parameter) as a function of plant densities from (Baret et al., 2009). Typical values for 3D simulated scenes (crosses) and for actual canopies (circles): Wheat (Wh), Maize (Mz), Sorghum (Sg), Vineyards (Vy), Sunflower (Sf), Soybean (Sy), Tomatoe (To), Olive tree (Ol), Peach (Pe), Lettuce (Lt) and Beet (Bt). 30

Figure 14. Scheme explaining the way the distribution of variable V is linked to that of LAI. 31

Figure 15: Distribution of values of observational characteristics used to simulate the learning data base. 33

Figure 16. Distribution of the soil brightness (Bs, left) and soil type (right, I_Soil) used. 34

Figure 17. Distribution used for canopy variables LAI, ALA and Hot (HsD). 34

Figure 18. Distribution of the leaf characteristics used. 35

Figure 19. Distribution and co-distributions of the input soil, canopy and leaf characteristics used to populate the training data base. 35

Figure 20: Distribution and co-distribution of the training database simulated reflectances, in the 8 SENTINEL2B bands (20m resolution). 37

Figure 21: Distribution and co-distribution of the training database simulated reflectances, in the 3 SENTINEL2A bands (20m resolution). 37

Figure 22. Distribution and co-distribution of the output variables. 38

Figure 23: On the left, relationship between LAI and FAPAR as established over the ensemble of VALERI sites. Symbols correspond to the sites. The solid line corresponds to the best fit model. On the right, symbols correspond to MODIS C4 relationship between LAI and FAPAR as simulated by the 3D radiative transfer model for the several vegetation types considered. The several lines correspond to the distribution within the learning data base. The line with triangles represents a threshold under which values are not expected. 39

Figure 24. Neural network architecture developed for the estimation of the biophysical variables considered from the eight SENTINEL2 bands (20m spatial resolution) and the 3 angles defining the geometry of observation. The network is made of 1 hidden layer of 5 neurons and 1 linear output neuron. The 'Norm' symbols correspond to the normalization process as described by Equation 10. Symbols 'S' and 'L' correspond respectively to the sigmoid (tansig) and linear transfer functions of the neurons. 41

Figure 25. Theoretical performances of the neural networks for LAI on the test database (S2B-20m) 43

Figure 26. Theoretical performances of the neural networks for FAPAR on the test database (S2B-20m) 43

Figure 27. Theoretical performances of the neural networks for FCOVER on the test database (S2B-20m) 44

Figure 28. Theoretical performances of the neural networks for CCC on the test database (S2B-20m) 44

Figure 29. Theoretical performances of the neural networks for CWC on the test database (S2B-20m) 45

Figure 30. Intercomparison between the LAI estimates over the simulated validation dataset. Top Left: SENTINEL-2A 8 bands versus SENTINEL-2A 3 bands, Top Right: SENTINEL-2B 8 bands versus SENTINEL-2B 3 bands, Bottom left: SENTINEL 2A 3 bands versus SENTINEL-2B 3 bands, and bottom right: SENTINEL 2A 8 bands versus SENTINEL-2B 8 bands 46

Figure 31. Distribution of the residuals between LAI estimates as a function of the true LAI value over the simulated validation database. Top Left: SENTINEL-2A 3 bands - SENTINEL-2A 8 bands, Top Right: SENTINEL-2B 3 bands - SENTINEL-2B 8 bands, Bottom left: SENTINEL 2A 3 bands - SENTINEL-2B 3 bands, and bottom right: SENTINEL 2A 8 bands - SENTINEL-2B 8 bands. Bars correspond to 25%, dots to individual values below 1% or over 99%. The solid red line corresponds to the median value. 47

Figure 32. Schematic representation of the convex hull in the case of 2 dimensional inputs of the network. The convex hull is approximated by a regular grid. The cells with ones (the gray cells)

correspond to cases of possible input range (i.e. inside the convex hull) while cells with zeros correspond to inputs out of range (outside the convex hull). 48

Figure 33. Distribution of the LAI standard deviation over 20x20m² (left), 30x30m², 40x40m², and 50x50m² (right) over the BELMANIP2.1 sites during year 2018. Each line is dedicated to a given biome type (top to bottom: shrub savanas and bare soils, crops and grasslands, deciduous broadleaf forests, Evergreen broadleaf forests and needleleaf forests. N is the number of pixels considered over the 2018 period and the BELMANIP2.1 sites. 51

Figure 34. Distribution of the slope between LAI standard deviation computed over 20x20m², 30x30m², 40x40m², and 50x50m²) over the BELMANIP2.1 sites during year 2018 for given biome types N is the number of pixels considered over the 2018 period and the BELMANIP2.1 sites. 52

Figure 35. Left: Sentinel2 images extract over a mountainous area (South West of France, Pyrenees, top) and farming area (South East of France – Camargue, bottom). Right standard deviation slope computed over the area. 53

List of Tables

Table 1. SENTINEL2 orbit characteristics and maximum scan angle 13

Table 2. SENTINEL2 spectral characteristics: band centre and width, spatial resolution and use. 14

Table 3. SENTINEL2 spectral characteristics: band centre and width, spatial resolution. Crosses indicate the bands selected to train the neural network for the 20m resolution product (NNET 20m) and the 10m product (NNET 10m). 18

Table 4: Minimum, maximum values and associated resolution for the five products and the two spatial resolutions. 20

Table 5. Distribution of the input variables of the radiative transfer model used to generate the training data base. Truncated Gaussian, log-normal or uniform distribution laws are used, characterized by the mode, the standard deviation (s), and minimum and maximum values. The number of classes for each variable is presented (Nb_Class). 32

Table 6. Values used for the co-distributions of the variables with LAI (see Figure 14). 32

Table 7. Characteristics of the uncertainties model used. 36

Table 8. Theoretical performances of the neural networks trained for each variables and using the spectral bands available at 10m resolution, and the spectral bands at 20m resolution. 42

Table 9. Minimum and maximum input values (bounding box) of the definition domain for inputs from SENTINEL2A TOC reflectance (20m). Note that these values vary from one dataset to another and the user has to referred to the excel datasheet delivered with the neural network design description. 49

Table 10. Tolerance, and Minimum (P_{min}) and maximum (P_{max}) values admitted for the products. 50

1 BACKGROUND OF THE DOCUMENT

1.1 EXECUTIVE SUMMARY

This ATBD (Algorithm Theoretical Based Document) describes the proposed algorithm for Level 2 products that will be implemented in the SENTINEL2 Toolbox. The level2 products are derived from SENTINEL2 top of canopy normalized reflectance data and correspond to the following set of biophysical variables: *LAI,* (Leaf Area Index) FAPAR (fraction of absorbed photosynthetically active Radiation) and FVC that are essential climate variables (ECVs) as recognized by international organizations such as GCOS and GTOS.

The proposed algorithm is based on methods that have already been proven to be efficient. They have been implemented to generate biophysical products from VEGETATION, MERIS, SPOT, and LANDSAT sensors. It mainly consists in generating a comprehensive data base of vegetation characteristics and the associated SENTINEL2 top of canopy (TOC) reflectances. Neural networks are then trained to estimate the canopy characteristics from the TOC reflectances along with the corresponding angles defining the observational configuration.

This ATBD is derived from the ones proposed in the frame of previous ESA projects:

- S2SPAD, Contract ESRIN #21450/08/I-EC: first ATBD delivered, theoretical performances only.
- VALSE2 (VALidation of SEntinel 2 – ESTEC A0/1-6958/11/NL/BJ): validation of the S2SPAD algorithm over experiment ground campaigns (airborne acquisitions to simulate S2-like data). Improvements of the algorithm (inclusion of acquisition geometry and illumination conditions)
- SL2P (Simplified L2 Product Prototype Processor – ESTEC AO/1-7455/13/NL/BJ): refinement of the optimal set of band inputs with regards to atmospheric noise, set up of a definition domain flag, and uncertainty estimates.

The modification with regards to SL2P latest version concern the fact that the actual SENTINEL2 spectral sensitivity (in replacement to the waveband centre values) were taken into account when generating the training database. We also compared two ways of taking into account the geometry of acquisition: either as inputs of the neural nets, or by interpolating values between multiple neural nets trained on specific geometries of acquisition. We concluded that the two methods led to similar results.

1.2 SCOPE AND OBJECTIVES

SENTINEL2 is part of the GMES space segment. Users include the scientific community as well as other stakeholders including policy makers, the proper information required for several applications, as detailed in the Mission Requirement Document (Gascon and Berger, 2007). For the exploitation of MSI (Multi Spectral Instrument) data, ESA develops the Sentinel-2 Toolbox which consists of a rich set of visualisation, analysis and processing tools products. The objective of this document is to provide a detailed description and justification of the algorithm proposed for the SENTINEL2 Toolbox level2 biophysical variables algorithm.

1.3 CONTENT OF THE DOCUMENT

This ATBD document is split in 3 main sections:

1. **Algorithm Justification and overview.** This section contains:
    - A definition of the proposed products.
    - A brief description of SENTINEL2 sensor from which the products will be derived
    - A justification of the algorithm selected
    - The outline of the algorithm.
2. **Description of the algorithm.** This section contains:
    - The required inputs and outputs provided by the algorithm.
    - The retrieval technique used: neural network techniques constitute the core of the operational algorithm. Quality indicators are also provided.
3. **Recommendations for the algorithm implementation.** This section contains details on how the algorithm could be implemented in the processing chain.

1.4 SYMBOLS AND ACRONYMS

µg	Microgram
ALA	Average Leaf Angle
ANN	Artificial Neural Network
ATBD	Algorithm theoretical based Document
BHR	Bi-Hemispherical Reflectance
BRF	Bidirectional reflectance factor
BRDF	Bidirectional Reflectance Distribution Function
Bs	Soil Brightness
Cab	Chlorophyll content in the leaf ($\mu g.cm^{-2}$)
Cbp	Content of brown pigments in the leaf (no units)
CCC	Canopy Chlorophyll Content
Cdm	Content of dry matter in the leaf ($g.cm^{-2}$)
CEOS	Committee for Earth Observation Satellite
ECV	Essential Climate Variable
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
CWC	Canopy Water Content
FVC	Fraction of vegetation cover
GCOS	Global Climate Observation System
GMES	Global Monitoring of Environment and Security
GTOS	Global Terrestrial Observation System
L2	Level 2 product

L3	Level 3 product
LAI	Leaf Area Index
MODIS	Moderate Imaging Spectrometer
N	Structural parameter of the leaf (unitless)
NIR	Near Infrared
NNT	Neural Network Technique
RMSE	Root Mean Square Error
RTM	Radiative Transfer Model
SPOT	Satellite Pour l'Observation de la Terre
SWIR	Short Wave Infra red
TOA	Top of Atmosphere
TOC	Top of Canopy
VI	Vegetation Index

2 ALGORITHM JUSTIFICATION AND OVERVIEW

2.1 PRODUCTS AND THEIR DEFINITIONS

The considered products correspond to the actual vegetation biophysical variables defined below.

2.1.1 Leaf Area Index (LAI)

LAI is defined as half the developed area of photosynthetically active elements of the vegetation per unit horizontal ground area. It determines 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 that should not depend on observation conditions. *LAI* is strongly non linearly related to reflectance. Therefore, its estimation from remote sensing observations will be strongly scale dependent (Weiss et al., 2000; Garrigues et al., 2006a). Note that vegetation *LAI* as estimated from remote sensing will include all the green contributors, i.e. including understory when existing under forests canopies. However, except when using directional observations (Chen et al., 2005), LAI is not directly accessible from remote sensing observations due to the possible heterogeneity in leaf distribution within the canopy volume. Therefore, remote sensing observations are rather sensitive to the 'effective' leaf area index, i.e. the value that would produce the same remote sensing signal as that actually recorded, while assuming a random distribution of leaves. The difference between the actual LAI and effective LAI may be quantified by the clumping index (Chen et al., 2005) that roughly varies between 0.5 (very clumped canopies) and 1.0 (randomly distributed leaves).

2.1.2 FAPAR

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. It depends on canopy structure, vegetation element optical properties and illumination conditions. 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 alternative FAPAR definitions (Baret 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.

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 (considered as leaf chlorophyll content higher that 15µg.cm-2) of the canopy.

2.1.3 Cover fraction (FVC)

It corresponds to the gap fraction for nadir direction. FVC is used to separate vegetation and soil in energy balance processes, including temperature and evapotranspiration. It is computed from the leaf area index and other canopy structural variables and 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, *FVC* will be only marginally scale dependent (Weiss 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.

2.1.4 Canopy chlorophyll content (CCC)

The chlorophyll content is a very good indicator of stresses including nitrogen deficiencies. It is strongly related to leaf nitrogen content (Houlès 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 et al., 2003). In addition, studies have demonstrated that a direct estimation of *CCC* is more robust and accurate than an estimation based on the product of the individual estimation of *LAI* and C_{ab} (Weiss et al., 2000). Therefore, the estimation of *CCC* has been preferred to that of the leaf chlorophyll content.

2.1.5 Canopy water content (CWC)

Since radiation is absorbed significantly by water in the near and middle infrared, the spectral configuration of SENTINEL2 allows accessing this variable. Water represents between 60 % and 80% of the living plant mass. The variable that is the best related to the remote sensing signal is defined as the mass of water per unit ground area (g.m^{-2}). One of the difficulties in retrieving this variable is the possible confusion with soil moisture effects. Canopy water content (CWC) is proposed here as a possible candidate in the list of SENTINEL2 products.

2.2 INSTRUMENT CHARACTERISTICS

SENTINEL2 is an optical sensor aboard a polar platform providing helio-synchronous observations. The characteristics of the orbit are presented in Table 1.

Acronym	Values
Orbit altitude (km)	786
Repeat cycle (days)	10
Period (min)	100.7
Inclination (°)	98.62°
Equatorial descending node crossing time (hr)	10:30
Maximum scan angle	20.6°

Table 1. SENTINEL2 orbit characteristics and maximum scan angle

The 13 characteristics of the 13 bands are presented in Table 2. Bands B1, B2, B9 and B10 are more dedicated to atmosphere or cloud. The 9 remaining bands may be used for vegetation characterization. They are all with 20 m spatial resolution, except B4 and B8 that have a 10 m resolution. This will thus need binning these bands to provide the 20 m spatial resolution. The detailed spectral characteristics are presented in Figure 1.

Acronym	Central (nm)	Width (nm)	Spatial resolution (m)	Potential Applications
B1	443	20	60	Atmosphere
B2	490	65	10	Atmosphere
B3	560	35	10	Vegetation

B4	665	30	10	Vegetation
B5	705	15	20	Vegetation
B6	740	15	20	Vegetation
B7	783	20	20	Vegetation
B8	842	115	10	Vegetation
B8a	865	20	20	Vegetation
B9	945	20	60	Atmosphere
B10	1375	30	60	Atmosphere
B11	1610	90	20	Vegetation
B12	2190	180	20	Vegetation

Table 2. SENTINEL2 spectral characteristics: band centre and width, spatial resolution and use.

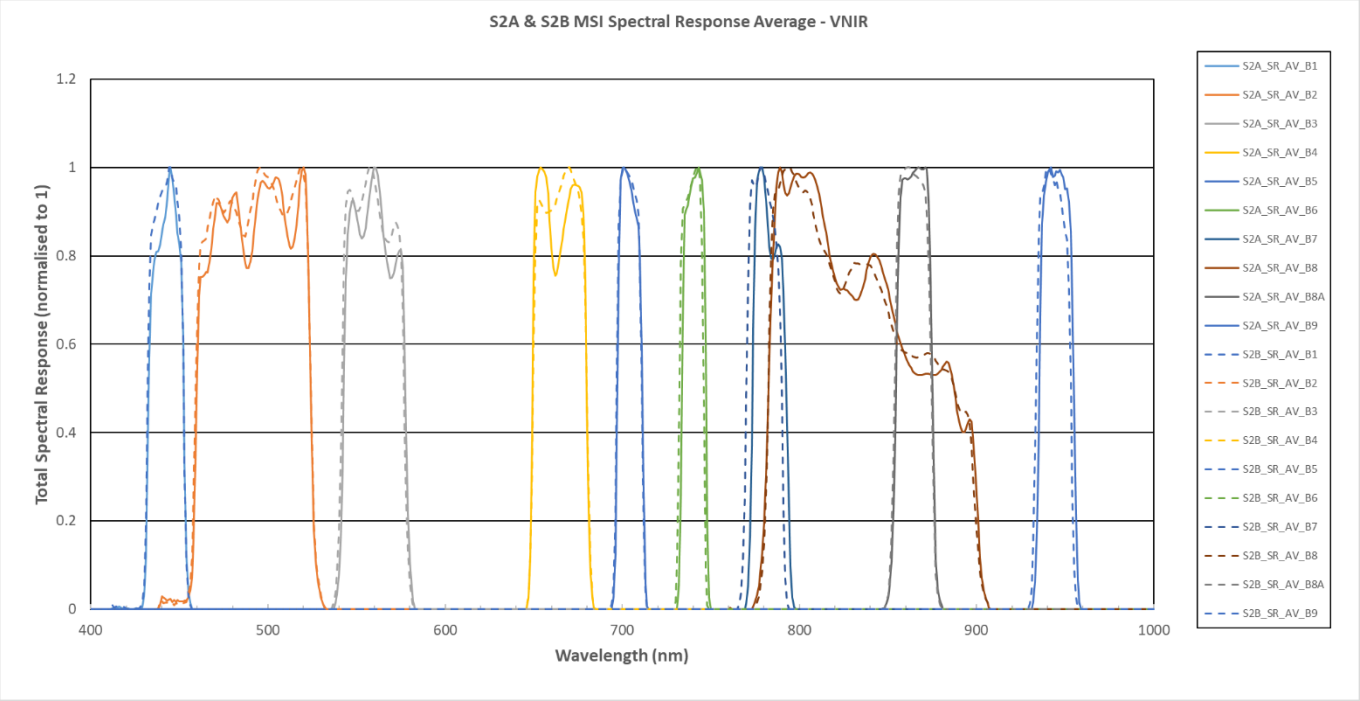


Figure 1. SENTINEL2A and 2B MSI spectral from ESA SENTINEL online web site.

2.3 JUSTIFICATION FOR THE ALGORITHM SELECTION AND DESIGN

The requirements for the selection and design of the algorithm proposed in this study for SENTINEL2 level 2 products are presented below:

- ***Explicit use of all the SENTINEL2 pertinent spectral information***. The spectral sampling of SENTINEL2 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 adequate SENTINEL2 spectral information should hopefully allow to restrain the solution space and lead to a more robust and accurate retrieval as compared to high resolution sensors such as SPOT or LANDSAT.

- ***Accuracy of the retrieval and computational efficiency***. A review of current state of the art for the estimation of biophysical variables from remote sensing data (Baret and Buis, 2007) was proposed. It shows that among the several retrieval algorithms, those based on the minimisation of the distance in the space of canopy variables provides the best accuracy on retrievals while being very efficient computationally wise. Therefore, techniques based on neural networks are selected in this study. In addition, their limitation mainly driven by the necessity to have a fixed number of input variables would not constitute any problem to process SENTINEL2 data up to level 2, if the geometrical configuration is input explicitly. Note that such techniques have already been implemented and lead to good performances of the retrieval (Baret et al., 1997; Weiss et al., 2002); (Combal et al., 2002a); (Kimes et al., 2002) (Bacour et al., 2006). It requires a learning process achieved over a training data base.
- ***No use of ancillary data difficult to derive at high spatial resolution***. As a matter of fact, retrieval of surface characteristics from remote sensing observations is an ill posed problem, leading to uncertainties in the solution (Combal et al., 2002b). The only way to regularize the problem is to use ancillary information or additional constraints on the solution to reduce its domain of variation. Spatial and temporal constraints may be used to regularize the solution as demonstrated by (Lauvernet et al., 2008) and (Kötz et al., 2005; Kötz et al., 2007) because we are mainly considering LEVEL2 products. Further, because of the generic (global) nature of the proposed vegetation products, no specific ancillary information can be efficiently used: the absence of reliable and regularly updated land cover map at a spatial resolution similar to that of SENTINEL2 from which some vegetation architecture features could be derived prevents from tailoring specific algorithm for each of the land cover classes. The proposed algorithm will therefore rely only on SENTINEL2 instantaneous observations. As a consequence, the algorithm may provide reasonably good estimates over all the cases, but certainly poorer performances as compared to algorithm specific to a given surface type for this particular surface type, although this specific algorithm would fail over most of other cases. Nevertheless, it could be possible to develop such specific algorithms and use them when the land cover is known. It would be also possible to develop corrections of the generic algorithm that will be specific to a surface type. Note also that the differences between a specific and a generic algorithm will depend on the considered variables: large differences are expected for LAI, while less sensitivity is expected for FAPAR and FVC. The genericity of the algorithm will obviously have important consequences on the generation of the training data base.
- ***Generation of the training data base***. The training data base should sample all the vegetation types and conditions that can be observed from SENTINEL2 over land surfaces. In addition it should reflect the uncertainties in the reflectance values as observed by SENTINEL2. Ideally, the training data base should therefore be made of SENTINEL2 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 within a large range of vegetation types and conditions, this simple ‘experimental’ approach is currently 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 canopy reflectance as observed within SENTINEL2 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 and water content on reflectance and transmittance,

- the background reflectance that should include in addition to a large variety of soils, litter and senescent vegetation.

- ***Quality assessment***. Quantitative and qualitative indicators should be attached to the product so that the user could properly 'weigh' the data within its application according to the confidence he puts on. This could be achieved within several ways:
  - Quality of the TOC reflectance used as input to the algorithm. This would simply correspond to the replication of indicators produced previously such as cloud occurrence, sensor problem and atmospheric correction.
  - Additional indicators based on:
    - The input out of range. This should indicate that either the input reflectances have problems (cloud contamination, poor atmospheric correction, shadow) or anyway that the application of the algorithm could result in unreliable results. This would be common to all the derived products.
    - Output out of rage: flags raised when the product appears to be out of the nominal range of variation.
    - Product uncertainty. The algorithm provides a quantitative estimation of the uncertainty associated to the product.

2.4 ALGORITHM OUTLINE

For the reasons exposed above, ***it is proposed to use neural networks for the SENTINEL2 biophysical variables estimation***. Once the neural network inputs and outputs are defined, it is very easy to change the network coefficients to take advantage of recent advances in the generation of the training data base. The processing chain will therefore be very easily upgraded if it designed for, i.e. if coefficients are stored as parameters in a well identified and documented manner.

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.

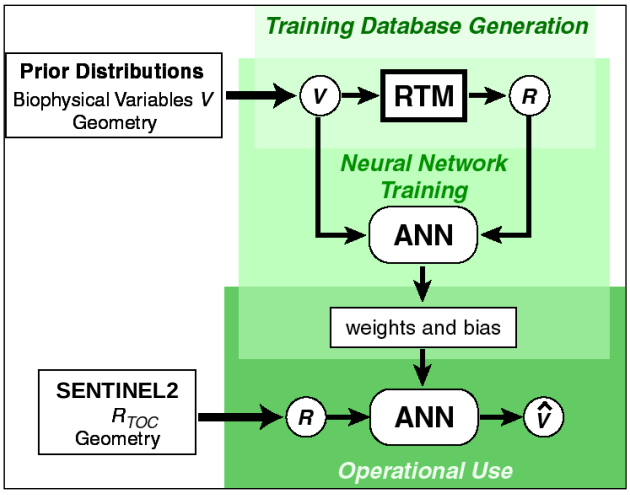


Figure 2. Flow chart showing how the products ($\hat{V}$) are generated operationally. ANN corresponds to an Artificial Neural Network characterized par its structure and its coefficients (synaptic weights and bias); R_{TOC} corresponds to the SENTINEL2 Top Of Canopy reflectance used in the operational mode and V correspond to the biophysical variable in the training data base and estimated by running the ANN over the simulated SENTINEL2 TOC reflectance and geometry.

2.4.1 Training the neural network

This process consists mainly in three steps:

- Generating training data base
- Defining the neural network architecture
- Calibrating the network.

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. It should be constituted of a representative set of top of canopy reflectances and incorporate prior information on the distribution of the input variables. The model used to simulate the learning database should represent a good compromise between the level of accuracy and the complexity of setting up the simulations: number and distribution of the input variables to describe the canopy as well as computer ressources and time.

2.4.1.2 *Designing network architecture*

It consists in defining the optimal structure (typically the number of layers and the number of neurons per layers) as well as possible transformations of the inputs and outputs such as normalization.

2.4.1.3 *Calibrating the network*

This last step corresponds to the actual training, i.e. tuning the coefficients (synaptic weights and bias) 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

Once the neural network is trained, it will be run in operational mode. 3 networks will produce in parallel estimates of the considered biophysical variables: LAI, FAPAR, Fcover. Additionally, quality assessment indicators will also be generated:

- **Input consistency with the training data base.** This represents the consistency of the measured SENTINEL2 input reflectances with those used in the training data base. The training definition domain of the inputs is therefore identified, and a flag will be raised when observations are outside the training definition domain.
- **Output consistency with expected range.** This represents the consistency of the actual network outputs (the biophysical variables) with those used in the training data base.
- **Quality indicators:** These are a replication of the previously computed quality indicators, including those related to the atmospheric correction and cloud filtering.

3 ALGORITHM DESCRIPTION

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

- The definition of the inputs and outputs,
- The radiative transfer model used to generate the learning data base
- The inversion technique
- The quality assessment

3.1 INPUTS AND OUTPUTS

3.1.1 Inputs

The neural networks will apply on instantaneous top of canopy reflectance data (level 2). All the following inputs are required for each considered pixel.

3.1.1.1 *Top of canopy reflectance.*

Individual daily observations will be used. Reflectances should be expressed in terms of reflectance factor, mainly varying between 0 and 0.7 for most land surfaces outside hot-spot or specular directions and snow or ice cover.

Two spatial resolution are considered for the generation of the biophysical products: 20m and 10m spatial resolution, and specific neural networks are trained for SENTINEL2A and SENTINEL2B. This makes four neural networks for a given biophysical variable: 20m_S2A, 20m_S2B, 10m_S2A, 10m_S2B. The 20m resolution products are generated by using eight bands B3, B4, B5, B6, B7, B8a, B11 and B12., where B3 and B4 are aggregated at 20nm (average of the 2x2 p10m pixels). This band selection is derived from the VALSE2 and SL2P project results. The 10 m resolution products are generated by using the three bands at 10m (B3, B4, B8). However, considering the limited spectral information at 10m, only LAI, FAPAR and FVC are estimated.

Acronym	Central (nm)	Width (nm)	Spatial resolution (m)	NNET 20m	NNET 10m
B1	443	20	60		
B2	490	65	10		
B3	560	35	10	Agg	x
B4	665	30	10	Agg	x
B5	705	15	20	x	
B6	740	15	20	x	
B7	783	20	20	x	
B8	842	115	10	x	x
B8a	865	20	20	x	
B9	945	20	60		
B10	1375	30	60		
B11	1610	90	20	x	
B12	2190	180	20	x	

Table 3. SENTINEL2 spectral characteristics: band centre and width, spatial resolution. Crosses indicate the bands selected to train the neural network for the 20m resolution product (NNET 20m) and the 10m product (NNET 10m).

.

3.1.1.2 *Geometry of acquisition*

Similarly to the spectral information content of the reflectance signal, the directional information must also be taken into account when training the neural networks. The cosine of the sun zenith angle (θ_s), view zenith angle (θ_v) and relative azimuth angle (φ) at the time of the image acquisition are required. The geometry of acquisition can be taken into account in two ways when using the neural nets:

1. Global Neural Network: the cosine of the sun zenith angle (θ_s), view zenith angle (θ_v) and relative azimuth angle (φ) are used as inputs to the neural networks. The simulations of the training data base must take into account the orbit characteristics of SENTINEL2 to simulate the satellite angular sampling.
2. Interpolated Neural Networks: the geometry of acquisition is not used as input to the neural networks: it is taken into account by training neural networks with simulations run for fixed values of the triplet $(\cos(\theta_s), \cos(\theta_v), \cos(\varphi))$: for each variable, a number N of neural networks is calibrated, each of them trained with a single geometric condition. When deriving the biophysical variable for a given sensor acquisition, the different neural networks are run, and the final product is estimated by interpolating the different outputs corresponding to fixed geometries into the actual sensor view angles.

We compared the results between these two implementations for LAI estimation. All the neural networks were trained with the same drawing of PROSAIL input variables (Table 5 and Table 6), except for the angles. For each simulation, the global neural net was trained with a uniform random drawing of the date and location, allowing to derive, thanks to the orbit characteristics of SENTINEL2, the corresponding geometry. For the interpolated Neural Networks, we trained 99 NNets (respectively, 27) with a single geometry configuration corresponding to sun zenith angles between 25° and 35° by step of 5°, view zenith angles between 5° and 15° by steps of 1° (respectively 5°), and relative azimuth angle between 180° and 210° by steps of 5°, which is in agreement with S2 geometry (see section 3.3).

Then, we simulated an independent test database using the same radiative transfer model (PROSAIL) with a uniform random drawing of the date and location. On one hand, we applied the global neural network, by using the actual geometry as input. On the other hand, we applied the 99 (respectively 27) NNETs and interpolated the obtained values at the actual geometry value (Figure 3). Results show that the performances of the 3 methods are comparable (Correlation coefficients of 0.83 for the 3, RMSE between 0.68 and 0.70). Moreover, the scatterplots between the global NNets and the interpolated ones shows a very good agreement (R^2=0.98), with few points a little more scattered when degrading the resolution of the view angle. This result also show that the accuracy required on the view angle is not critical. As the results are very similar, we conclude that it is much simpler to use the global NNet for the implementation in the SENTINEL2 toolbox.

However, the determination of the viewing angles for a given pixel may not be currently an easy task when using current SENTINEL2 data. Indeed, there is one viewing configuration associated to each of the 12 detectors for a given band (making 8*12 view angles per pixel). The current S2 L1C data format does not provide the information on the detector associated to the reflectance in case of overlapping between two detectors. Therefore, it is required to use an approximation of the view angle, which will not degrade the results as shown in (Figure 3, bottom). We propose to use the same approximation as in the sen2cor processor that is also implemented in the S2Toolbox to perform atmospheric corrections.

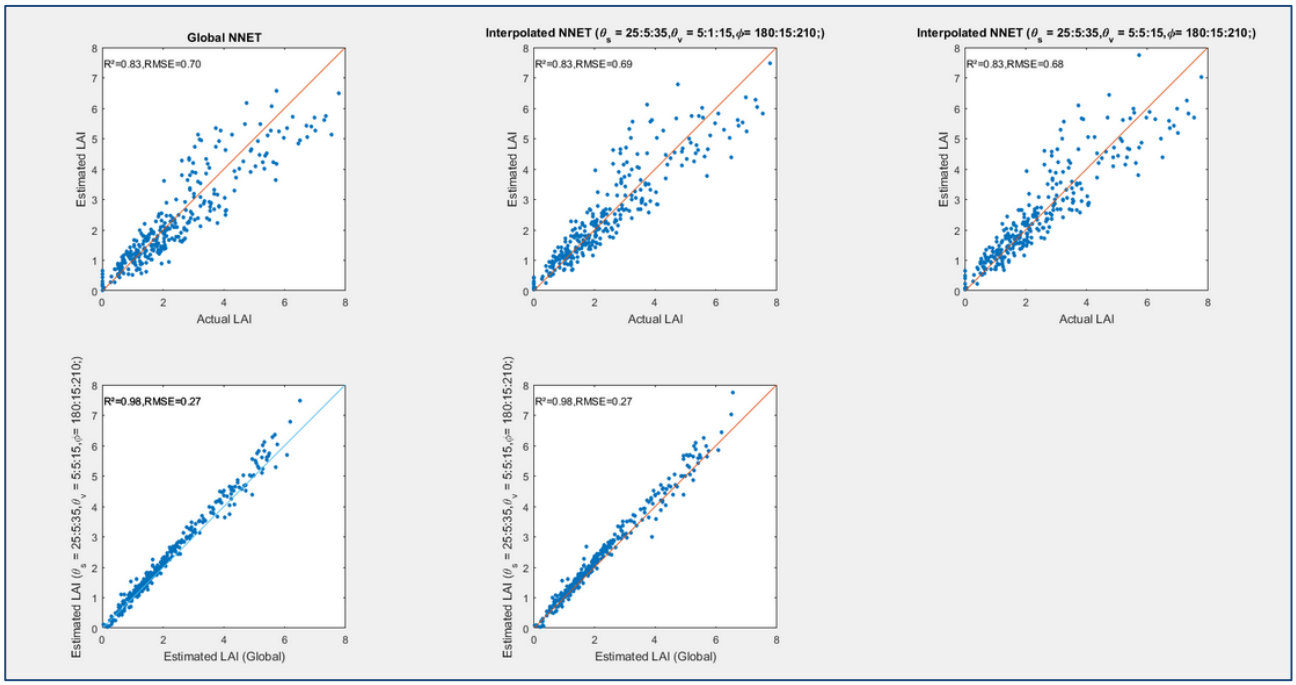


Figure 3. Comparison between the performances of a global neural network versus interpolated ones. Top: theoretical performances against a simulated test database, Bottom: scatterplot between interpolated and global neural network LAI estimates in the test database. Two interpolated neural networks are considered: 1°step or 5° step.

3.1.2 Outputs

The outputs will be provided by application of the algorithm over each pixel. They include the neural network derived LAI, FAPAR, FVC,CCC and CWC values as described previously. The proposed range of variation and resolution are presented in Table 4. In addition to the product values, quality flags are also generated.

Product	Unit	Minimum	Maximum	resolution	10m Product	20m Product
LAI	$m^2.m^{-2}$	0	8.0	0.01	xl	x
FAPAR	-	0	1.0	0.01	x	x
FVC	-	0	1.0	0.01	x	x
CCC	g/cm²	0	600	1		x
CWC	µg/cm²	0	0.55	0.0025		x

Table 4: Minimum, maximum values and associated resolution for the five products and the two spatial resolutions.

3.2 REFLECTANCE MODELS

As previously described, the algorithm is based on the training of neural networks for each product. Distinct neural networks were trained for SENTINEL2A and SENTINEL2B because of their specific spectral response. However, apart from this difference, the two training databases share exactly the

same simulation cases in terms of model inputs. Emphasis was therefore put on the generation on the training data base with attention on the assumptions associated to the physically based radiative transfer models used to generate this database. The 3 main components are 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 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 viewing directions. In addition, PROSPECT assumes that the optical properties of both leaf faces are equal.

Several authors (Jacquemoud and Baret, 1990; Fourty and Baret, 1997; 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 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 event senescent leaves having disorganized mesophyll with large amount of air spaces.
- C_{ab} Leaf Chlorophyll content (µg.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 90 µg.cm^{-2}.
- C_{dm} Leaf dry matter content (g.cm^{-2}). Dry matter absorbs over the whole spectral domain, and its effect is 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_w Leaf water content (g.cm^{-2}). Several studies showed that the relative water content could be approximated to a value close to 75 % for the green leaves. This allows linking the water (C_w) and the dry matter (C_{dm}) contents together.
- C_{bp}. Leaf brown pigment content (relative units). Baret et al (2002) reported that chlorophyll and brown pigments are exclusive, i.e. green and non-green elements (senescent leaves, branches, stems) are spatially dissociated. The canopy structure model should therefore include at least green and non-green elements. Green leaves will have no brown pigments and senescent leaves will have no chlorophyll pigments.

Bacour et al (2002b) and Le Maire (2002) have analysed the sensitivity of the radiometric response both at the leaf and canopy level. 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 et al., 1996) or DISORD (Myneni 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 et al., 1999) or radiosity (Gerstl and Borel, 1992), (Borel et al., 1991), (Chelle et al., 1997), (Soler 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 ground level.
- The time associated to the model computation.

It is thus proposed to use a reflectance model that is computer efficient and uses a small number of input variables. The SAIL radiative transfer model (Verhoef, 1984, 1985) is widespread in the remote sensing community for the estimation of vegetation biophysical variables. The canopy is described as a 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, 1991).

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 *products*, all the green vegetation layers have to be accounted for in the computation of these variables. Therefore, if the understory is green (including lichens and moss), it will not be considered as the background here and will be included within the green vegetation layer. The background reflectance may thus correspond to soil, litter, water and snow. However, because of the particularities of water and snow backgrounds, these will not be included in this study. Indeed, when these particular cases will be encountered when applying the algorithm on SENTINEL2 data, the 'input out of range' quality indicator should be flagged.

3.2.3.1 *The background brightness concept*

The background reflectance, for a given wavelength, will depend on the background type (snow, soil type, litter, water), geometrical illumination and view conditions (Ω), roughness (z) or moisture (H). Note also that there is a continuum between soil background and water (which is always above soil!).

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 not to depend on 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:

$$\rho_b(\lambda,\Omega_i,H_j,z_k)=Bs\cdot\rho_b(\lambda,\Omega_l,H_m,z_n) \qquad \textbf{Equation 1}$$

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 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:

$$\rho_b(\lambda_1,\Omega_i,H_j,z_k)=a(\lambda_1,\lambda_2)\cdot\rho_b(\lambda_2,\Omega_i,H_j,z_k)+b(\lambda_1,\lambda_2) \qquad \textbf{Equation 2}$$

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

$$\rho_b(\lambda_1,\Omega_l,H_m,z_n)=a(\lambda_1,\lambda_2)\cdot\rho_b(\lambda_2,\Omega_l,H_m,z_n)+b(\lambda_1,\lambda_2) \qquad \textbf{Equation 3}$$

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 1 and Equation 2:

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

Identifying Equation 4 to Equation 3 provides the condition under which Equation 2 is valid:

$$Bs\cdot b(\lambda_1,\lambda_2)\approx b(\lambda_1,\lambda_2) \qquad \textbf{Equation 5}$$

which is true either for $Bs\approx 1$ or $b(\lambda_1,\lambda_2)\approx 0$. Experimental and theoretical results (Baret 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, 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 (Bacour et al., 2002a; Weiss 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
- A reference soil reflectance spectrum.

3.2.3.2 *Background reference spectral variation*

The background spectral database must represent the different background types with special attention to soils and litter.

Note that 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 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 comparable to that of soil background (Gausman et al., 1975), (Biard and Baret, 1997),

(Chen and McKyes, 1993). Few litter and vegetation residues were measured at the laboratory to get some reasonable representation of these background types. Because of the similarity between litter, residues and soil reflectance, these are finally aggregated within the 'soil' background category.

3.3 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 canopy reflectance for the 8 SENTINEL2A and 8 SENTINEL2B bands considered
- Addition of uncertainties to the simulated top of canopy reflectance values previously simulated.

3.3.1 Radiative transfer model

The top of canopy reflectances are simulated with the PROSPECT+SAIL model described earlier (Figure 4). The coupled model allows as well the computation of the secondary biophysical variables such as *FVC*, *FAPAR*. The model inputs 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), derived from the SENTINEL2 orbit characteristics and swath,
- the background reflectance spectrum, as described earlier,
- the primary biophysical variables related to leaf optical properties (*N*, C_{ab}, *Cw*, C_m, and C_{bp}) and to the canopy structure (*LAI*, *ALA, HOT,* and *Bs*). Their associated distribution law will be specified hereafter.

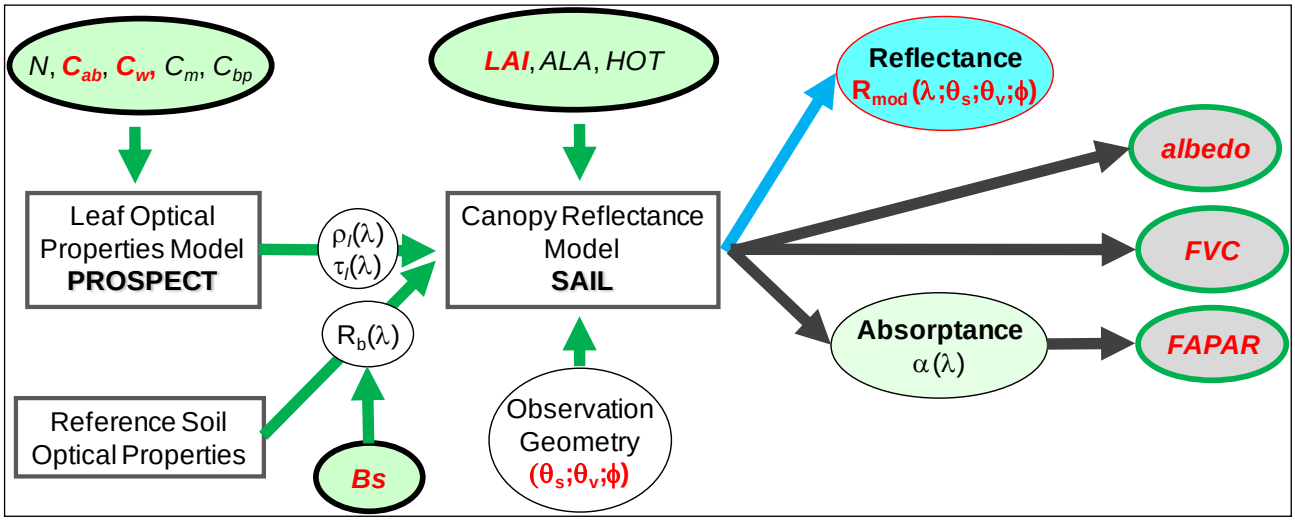


Figure 4: The coupled PROSPECT+SAIL model to generate the training database made of TOC reflectances and corresponding biophysical variables.

3.3.2 Distribution of the vegetation input variables

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 variable distributions are derived from available information on the actual distribution of the variables. Table 5 presents the range of variation and the actual distribution used for the input variables of the vegetation and background.

3.3.2.1 *Using results from the literature for leaf characteristics*

One of the main problems in generating a learning data base is to get the proper distributions for the input model variables if realistic distributions are required. Unfortunately, the distribution of most variables is not precisely known. For leaf optical properties, a brief literature review allows to characterize the distributions for chlorophyll (Cab) and dry matter contents (Cdm).

Results show that C_{ab} et C_{dm} distributions are roughly Gaussian (Figure 5 and Figure 6). However, they depend on the vegetation type:

- for chlorophyll (Figure 5), the mode is close to 40 µg.cm-2 for non evergreen vegetation types, while it is close to 70 µg.cm-2 for evergreen broadleaf forests (the Hawaiian data set reported in Feret et al (2008). Note that the chlorophyll content definition considered here is consistent with the input of the PROSPECT model, i.e. it includes chlorophyll a and b as well as carotenoids.
- for dry matter contents (Figure 6), the mode is close to 0.005 g.cm-2 for non evergreen broadleaf forests, and close to 0.012 for evergreen broadleaf forests.

The distribution derived from a compilation of literature data (Figure 6 left) with few broadleaf evergreen forest samples is very consistent with the observations of Feret et al (2008) (Figure 6 right).

Note however that the compilation of data presented here might not represent realistically the actual distribution of values around the globe. Anyway, the differences found between vegetation types should indicate that training specific algorithm for each vegetation type might improve the performances.

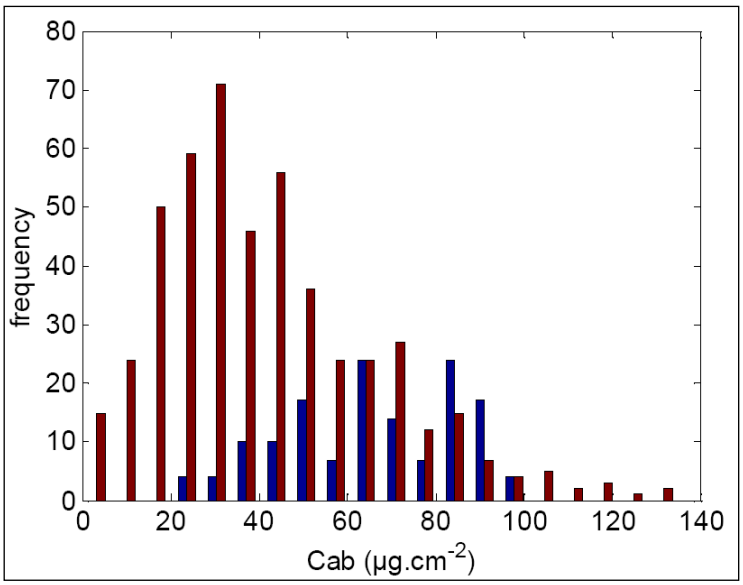


Figure 5: Distribution of leaf Chlorophyll content as described by Feret et al (2008). Values of Chlorophyll include chlorophylls a and b as well as carotenoids. Distribution in blue corresponds to data mainly from tropical forests (Hawai data set in Feret's paper). Distribution in red corresponds to other species.

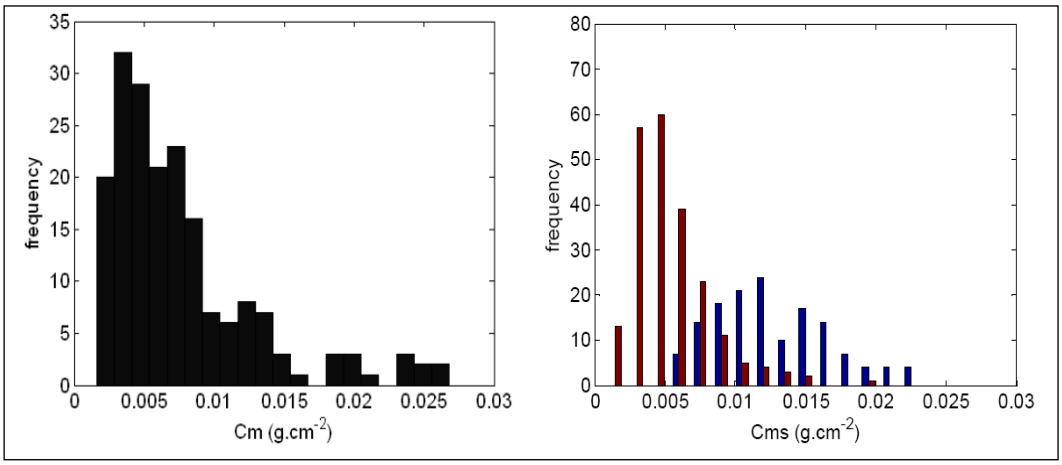


Figure 6: On the left, distribution of dry matter contents (Cdm) derived from literature review: (Prado and De Moraes 1997; Poorter and Evans 1998; Reich, Ellsworth et al. 1998; Poorter and De Jong 1999). On the right, distributions reported by Feret et al (2008): distribution in blue corresponds to data mainly from tropical forests (Hawai data set in Feret's paper). Distribution in red corresponds to other species.

The water content was tied to the dry matter content assuming that the green leaves have a relative water content close to 75%.

3.3.2.2 *Soil characteristics*

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 et al., 1992; Liu et al., 2002). Considering the brightness concept will allow increasing the diversity in actual soil properties. Only 7 soil spectra were selected among the 1500 measured ones available to represent the range of spectral shapes observed within a good accuracy (Figure 7). 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

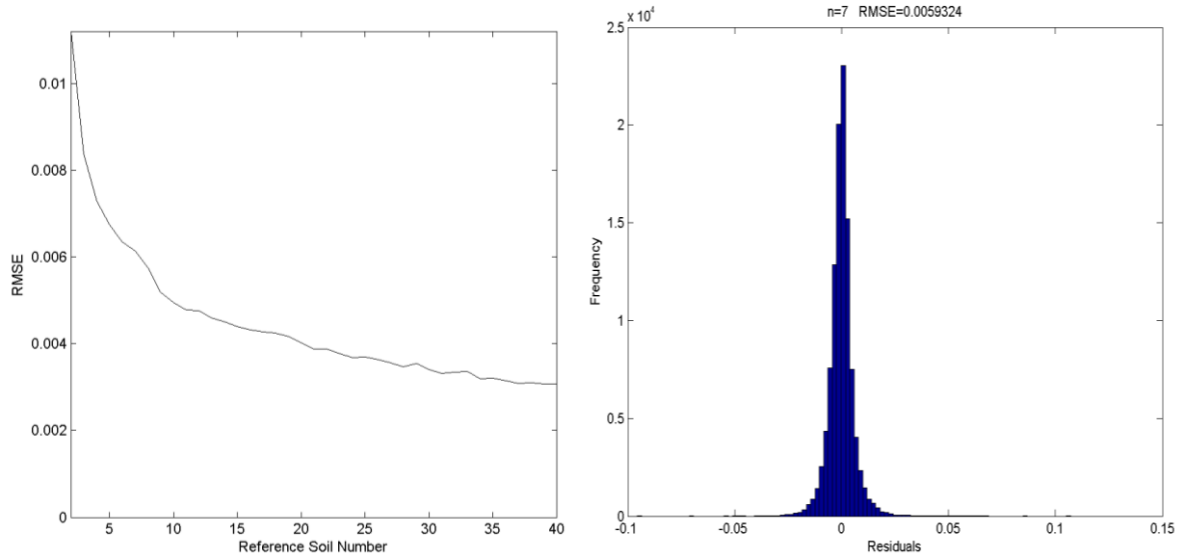


Figure 7. RMSE values associated to the reconstruction of all the soils investigated for SENTINEL2 bands (left) and distribution of the residuals when considering 7 soils.

The corresponding soil spectra are presented in Figure 8. Note that here all the spectra were normalized so that they provide the same reflectance value when averaged over the SENTINEL2 bands.

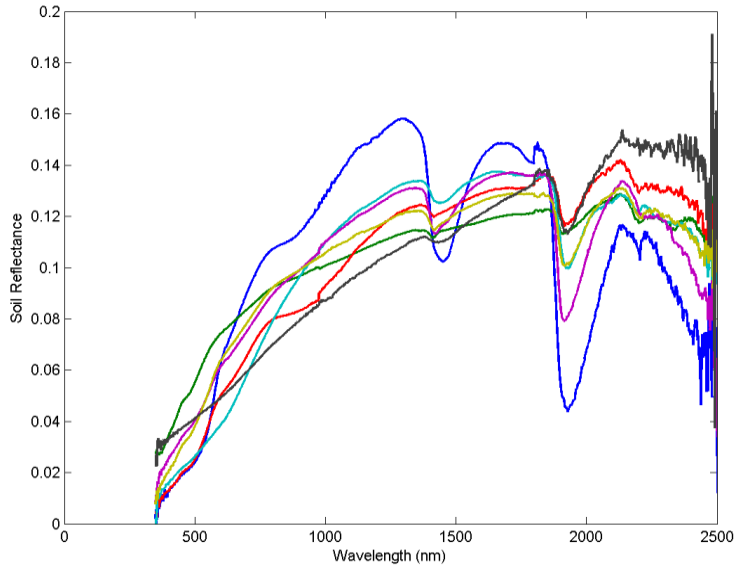


Figure 8. The 7 standard soil reflectance spectra in SENTINEL2 spectral domain as measured by Liu et al (2002).

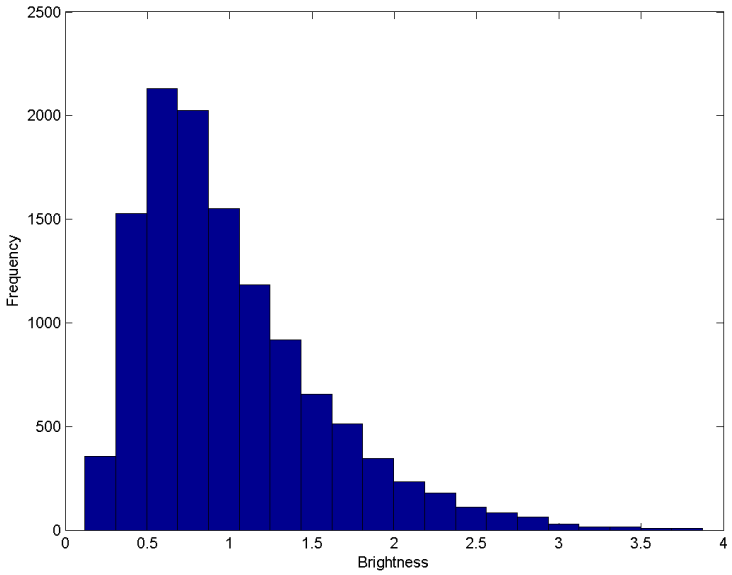


Figure 9. Distribution of brightness values as observed over Liu's data base (Liu et al., 2002) when using 7 reference spectra.

3.3.2.3 *Canopy variables*

3.3.2.1 *Leaf Area Index*

(Scurlock et al., 2001) have compiled about 1000 LAI experimental values over 300 original-source references. These values correspond to nearly 400 unique field sites during the 1932-2000 period. Figure 10 shows that the measured LAI is quite uniformly distributed between the 1-7 values while low LAI values (<1) are not frequently measured. On the other hand, some very high but not very frequent LAI values are represented in this data set (up to 40).

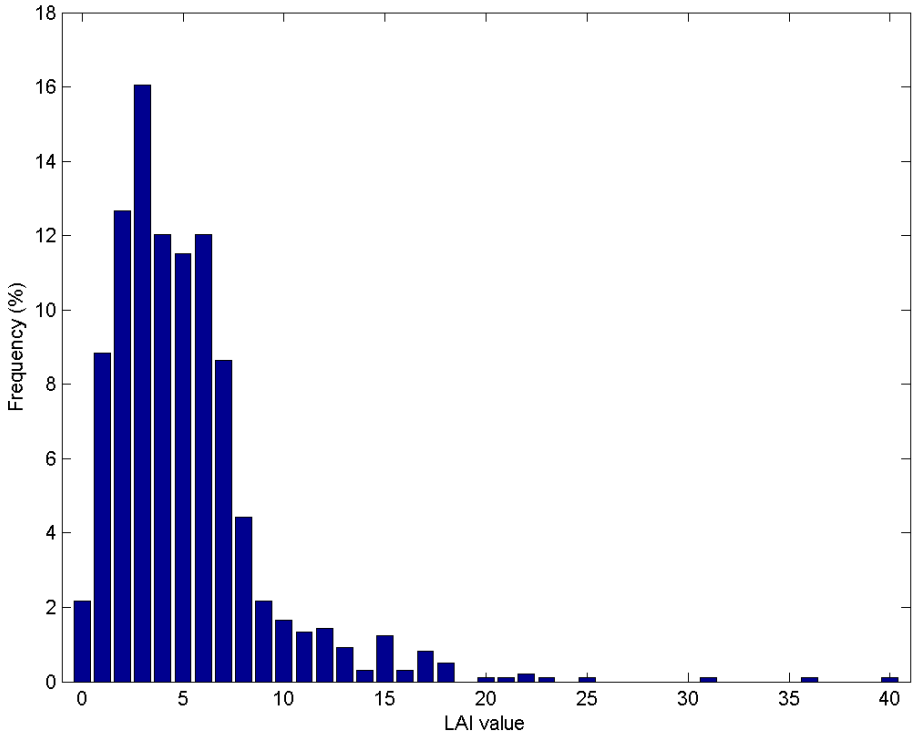


Figure 10. Distribution of leaf area index from the 1008 ground LAI measurements provided by Scurlock et al (2002)

However, although this data set represents a large amount of experimental data, it does not represent well the actual LAI distribution over the globe for the following reasons:

- The measurements are achieved when some vegetation is present during the experiments. Therefore, bare soils (LAI=0) are not taken into account here, whereas they represent about 30% of the emerged surface (Masson et al., 2003; Baret et al., 2006).
- The measurements used in this study are performed using different methods and LAI devices leading to different estimates of this variable. They all correspond to LAI but under different assumptions (effective or true LAI) and accuracy (Weiss et al., 2004; Garrigues et al., 2008)
- Finally the distribution derived from this study is certainly not representative of the actual distribution at the Earth surface over a year since the measurements correspond to punctual dates during the vegetation cycle depending on the objectives of the authors of the studies.

We compared the distribution obtained from Scurlock et al (2001) with the one derived from the VALERI data set (http://www.avignon.inra.fr). This latter corresponds to an ensemble of measurements performed between 2000 and 2008 over various vegetation types, including the main biomes, and is used for the validation of remote sensing products (Figure 11). The results presented

here correspond to the local measurements of effective LAI performed at the 20m scale. The VALERI distribution is significantly different from that of Scurlock's since true LAI values are most of the time higher than effective values. Further, the VALERI data set includes a significant fraction vegetation types with low LAI values such as crops shrubs and savannahs which also represent a significant fraction of the earth surface (about 30%). Therefore, it may better correspond to what is expected for the algorithm training data base although the spatial and temporal sampling are very limited.

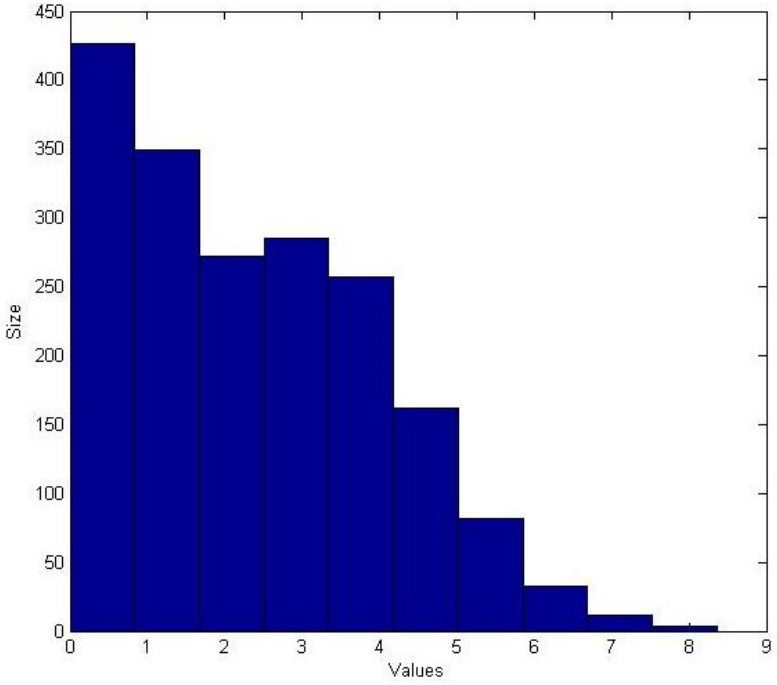


Figure 11. Effective LAI distribution over the VALERI sites

Therefore, the only way to access the actual distribution of LAI over the Earth Surface would be to use remote sensing estimates over a representative sample of the Earth surface (Masson et al., 2003; Baret et al., 2006). Such data set is currently available only at 1km spatial resolution (MODIS, CYCLOPES, GLOBCARBON) which might very significantly depart from the required distribution at 20 m resolution, particularly regarding the geostatistical characteristics of landscapes, with typical ranges of few hundreds of metres (Garrigues et al., 2006b).

In addition, the distribution used for LAI may also depend on that of the other variables, particularly for the larger LAI values corresponding to near 'saturation' conditions. In these conditions, the neural network will adjust a model that will pass in the middle of the cloud of possible solutions. To prevent underestimations due to more values with low LAIs, it was decided to use log-normal distributions that may agree quite well with those presented in Figure 10 and Figure 11 and fixing the upper limit to LAI=15. The log-normal distribution was slightly modified to increase the frequency of low LAI values by adding 15% of cases within a uniform distribution law for the 15% lower LAI values.

3.3.2.2 *Average Leaf Angle (ALA)*

Very few measurements of the average leaf angle measurements have been reported within the scientific community. The VALERI project provides some information on the expected distribution of LAI at 20 m spatial resolution. Figure 12 shows that low ALA values (up to 40°) are not very frequent. A peak is observed at ALA=60° which corresponds to a spherical distribution. The ALA was derived from the inversion of a gap fraction model measured either from hemispherical photos or LAI2000 devices allowing estimating concurrently both effective LAI (Figure 11) and ALA (Weiss et al., 2004).

However, due to the limited spatial and temporal sampling associated, with in additional possible biases due to the retrieval technique, these distributions might be only considered as indicative.

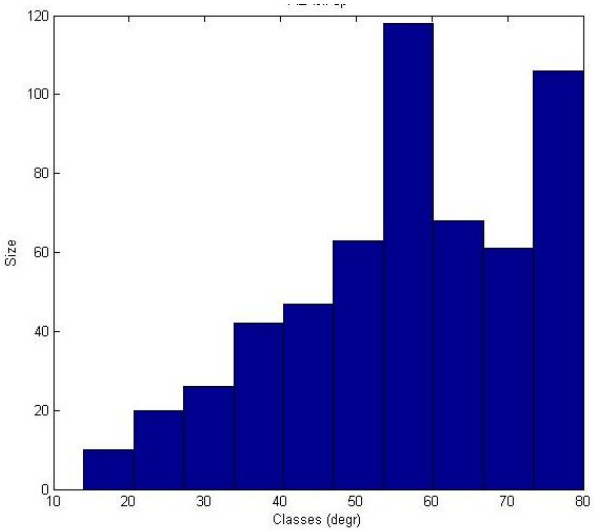


Figure 12. Distribution of the average leaf angle (ALA) over the VALERI sites

3.3.2.3 *Hot Spot Parameter (HOT)*

It is not easy to derive the hot spot parameter distribution at the Earth surface from the literature since it is a variable which is only used in radiative transfer modelling. Moreover, this variable has an impact only when remote sensing measurements are acquired near the principal plane. Figure 13 shows the distribution of typical values of the hot spot parameter for some crops provided by (Baret et al., 2009), most of the values are between 0.05 and 1. Extreme values are obtained for crops with specific architecture: low values for row planted trees (olive, peach) while the highest value is observed for young vineyards, lettuce and sugar beets.

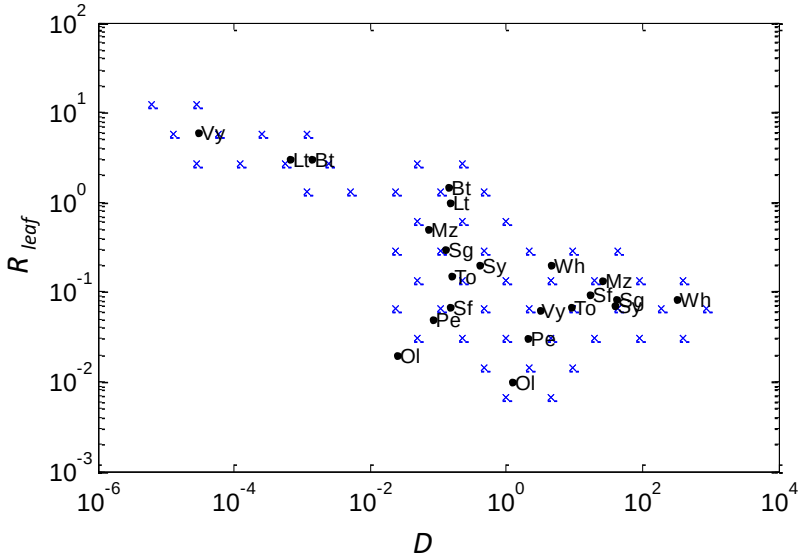


Figure 13. Relative leaf size (hot spot parameter) as a function of plant densities from (Baret et al., 2009). Typical values for 3D simulated scenes (crosses) and for actual canopies (circles): Wheat (Wh), Maize (Mz), Sorghum (Sg), Vineyards (Vy), Sunflower (Sf), Soybean (Sy), Tomatoe (To), Olive tree (Ol), Peach (Pe), Lettuce (Lt) and Beet (Bt).

3.3.2.4 *Co-distribution between variables*

Almost no information is available on possible co-distributions between variables. However, it is likely that some of the variables are linked together. For example, a very dense forest canopy will never be associated to low chlorophyll content and planophile leaf orientation. For this reason, we proposed to restrict the range of variation for some variables as a function of LAI value. This will be simply achieved by assuming that the range of variation linearly changes with LAI between $V_{min}(0)$ (respectively $V_{max}(0)$) and $V_{min}(LAI_{max})$ (respectively $V_{max}(LAI_{max})$) as illustrated by Figure 14. LAI_{max} is the maximum LAI value considered.

This will be simply achieved by assuming that the variable dynamics linearly varies with LAI according to:

$$(V-V_{min}(0))/(V_{max}(0)-V_{min}(0)) = (V^*-V_{min}(LAI))/(V_{max}(LAI)-V_{min}(LAI))$$

$$\text{With : } V_{min}(LAI)=V_{min}(0)+LAI*(V_{min}(LAI_{max})-V_{min}(0))$$

$$\text{and } V_{max}(LAI)=V_{max}(0)+LAI*(V_{max}(LAI_{max})-V_{max}(0))$$

where V* is the value of variable V after linking its distribution to that of LAI. The values defining the co-distributions are specified in Table 6. They were derived empirically, assuming that large LAIs corresponded to a restricted range of the other variables.

Note that on other complementary way to relate the distribution between input variables will consist in filtering the simulated outputs (reflectance, FAPAR) as it will be demonstrated in the following.

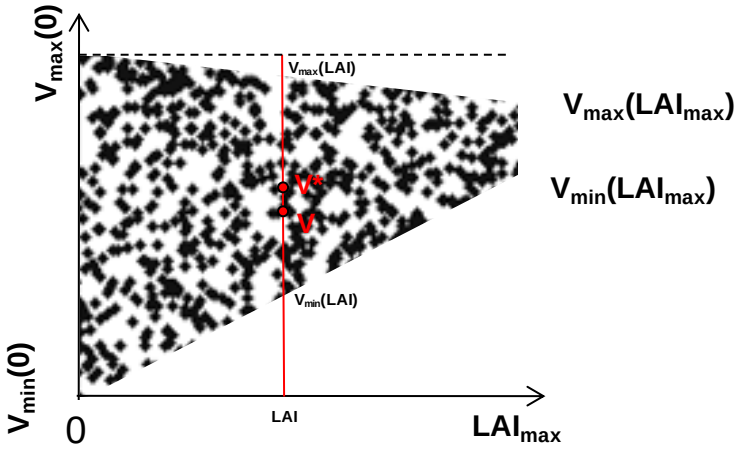


Figure 14. Scheme explaining the way the distribution of variable V is linked to that of LAI.

3.3.3 Actual generation of the inputs for the training data base

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 et al., 2002b) 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. We have simulated 41472 cases.

	Variable	Minimum	Maximum	Mode	Std	Nb_Class	Law
Canopy	LAI	0.0	15.0	2.0	3.0	6	Gauss
	ALA (°)	30	80	60	30	3	gauss
	Hot	0.1	0.5	0.2	0.5	1	gauss
Leaf	N	1.20	1.8	1.50	0.30	3	gauss
	Cab (µg.m^{-2})	20	90	45	30	4	gauss
	Cdm (g.m-2)	0.0030	0.0110	0.0050	0.0050	4	gauss
	Cw_Rel	0.60	0.85	0.75	0.08	4	gauss
	Cbp	0.00	2.00	0.00	0.30	3	gauss
Soil	Bs	0.50	3.50	1.20	2.00	4	uni

Table 5. Distribution of the input variables of the radiative transfer model used to generate the training data base. Truncated Gaussian, log-normal or uniform distribution laws are used, characterized by the mode, the standard deviation (s), and minimum and maximum values. The number of classes for each variable is presented (Nb_Class).

	Variable	Co_Distribution	$V_{min}(0)$	$V_{max}(0)$	$V_{min}(LAI_{max})$	$V_{max}(LAI_{max})$
Canopy	ALA (°)	Yes	30	80	55	65
	Hot	Yes	0.1	0.5	0.1	0.5
Leaf	N	Yes	1.20	2.20	1.3	1.8
	Cab (µg.m^{-2})	Yes	20	90	45	90
	Cdm (g.m-2)	Yes	0.0030	0.0110	0.0050	0.0110
	Cw_Rel	Yes	0.60	0.85	0.70	0.80
	Cbp	Yes	0.00	2.00	0.00	0.20
Soil	Bs	Yes	0.50	3.50	0.50	1.20

Table 6. Values used for the co-distributions of the variables with LAI (see Figure 14).

The sampling scheme is based on a full orthogonal experimental plan (Bacour et al., 2002b). This consists of identifying 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 5.

The following scheme was used to generate the training data base:

1. **Date.** Randomly draw a date. a whole year (365 days) is used. The year is decomposed into 4 periods that are successively sampled.

2. **Location.** The location is randomly selected within -56° and + 83° latitudes corresponding to the mission requirements. However above + 70° latitude the illumination conditions are very poor most of the time, impacting the accuracy of radiance measurements. In addition, the high sun zenith angles experienced induce strong atmospheric effects due to the increase optical path.

Once date and location are sampled, the corresponding geometrical configuration is derived using SENTINEL2 orbit characteristics and swath (Table 1). Note that special patterns are observable, due to the orbit constraints.

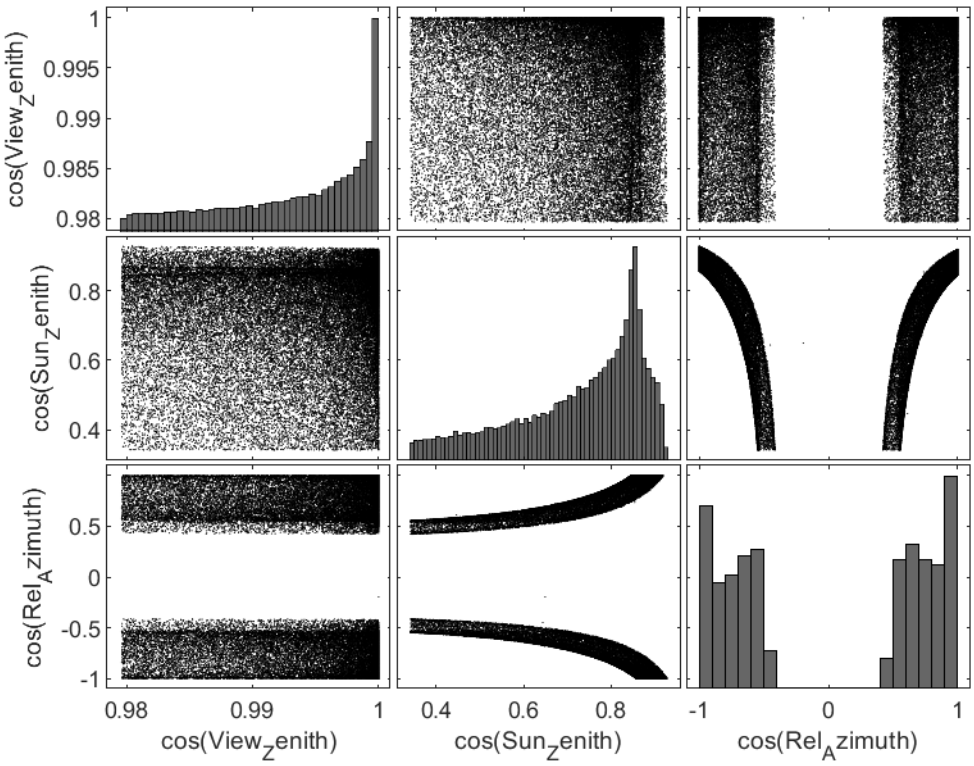


Figure 15: Distribution of values of observational characteristics used to simulate the learning data base.

3. Background variables.

- **Background type:** The background type is derived by randomly selecting one of the 7 soil background types (Figure 8).
- **Brightness**: The *Bs* coefficient is randomly drawn according to a truncated Gaussian distribution centred on *Bs*=1.0 (Table 5). The larger frequencies for the lower Bs values are explained by the co-distribution with LAI values: bright soils are not expected under very dense vegetation.

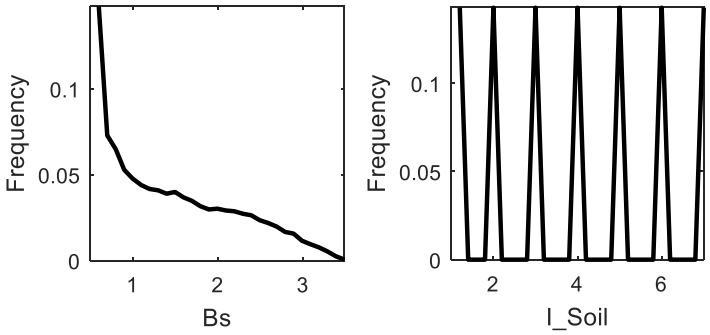


Figure 16. Distribution of the soil brightness (Bs, left) and soil type (right, I_Soil) used.

4. **Canopy variables.** The following distributions were used (Table 5):
   - ***LAI*:** The distribution of LAI values follows a gaussian distribution with a mode at LAI=2, and a relatively wide distribution (σ=3). This allows sampling significantly low and high values of LAI. The higher LAI values were truncated at LAI=15. This particular selection of distribution of LAI values improved the saturation problem by introducing a significant amount of very high LAI values in the training process. The distribution law was further modified to increase the frequency of low LAI values as explained previously.
   - ***ALA*:** The average leaf inclination angle is assumed to follow a truncated Gaussian distribution centred over the spherical widely represented one. The distribution is tied to the LAI, assuming that for large LAI values, leaf angle distribution was close to a spherical one. The peak observed between 55° and 65° is due to the co-distribution constraints for high LAI values.
   - ***HOT*:** The hot spot parameter follows a truncated Gaussian distribution.

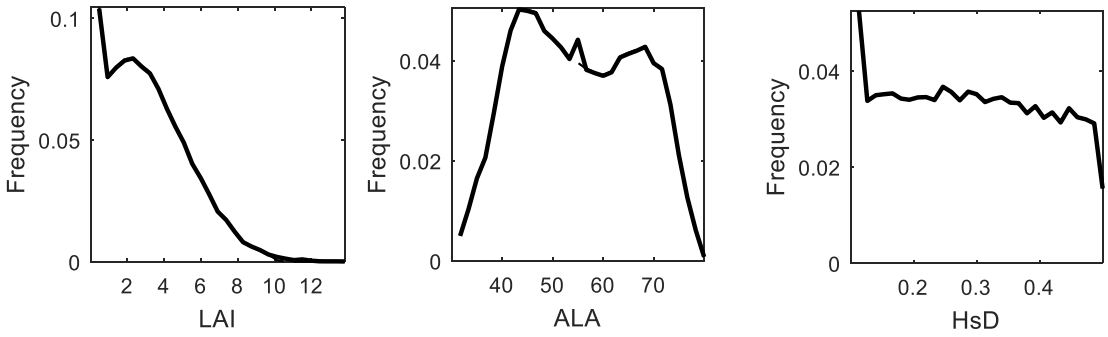


Figure 17. Distribution used for canopy variables LAI, ALA and Hot (HsD).

5. **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 5) The artefacts observed in the distributions are due to the co-distribution constraints applied for high LAI values (Table 6).

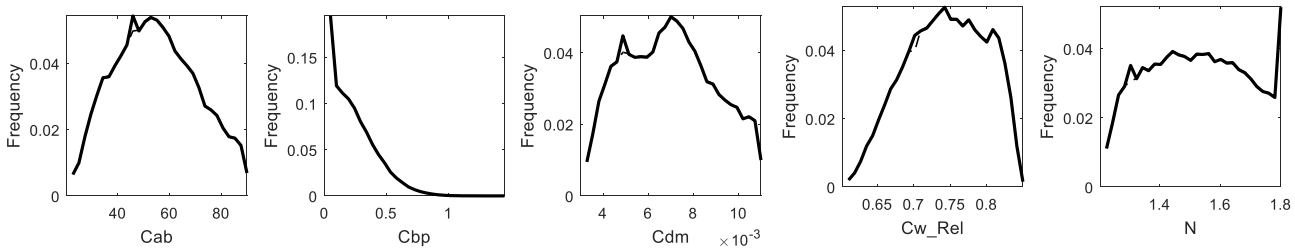


Figure 18. Distribution of the leaf characteristics used.

At the end of the simulation process, a total number of 41472 cases were simulated. Note that this number of cases is above the minimum 10 000 cases as required for a medium complexity problem and should allow good training performances for this more complex problem. This data set are then be split in two parts with a random selection process:

- **Training**: 2/3 of the simulations are affected randomly to the training of the neural network
- **Testing and Hyper-specialization**: 1/3 of the simulations are used for the hyper-specialization control and evaluation of theoretical performances.

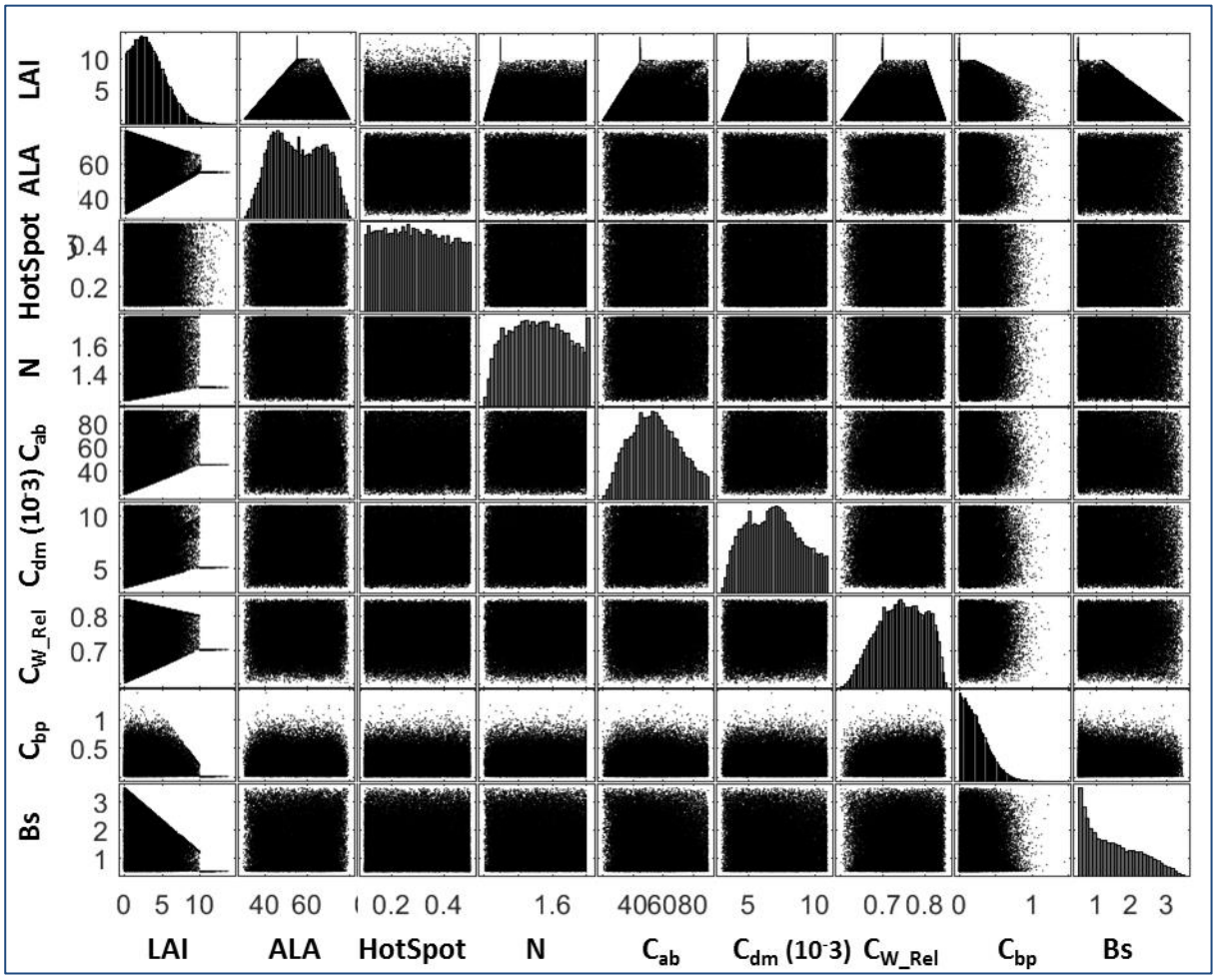


Figure 19. Distribution and co-distributions of the input soil, canopy and leaf characteristics used to populate the training data base.

6. Simulation of the top of canopy reflectance for the SENTINEL2 bands

The previously derived table of input variables is used to simulate the corresponding SENTINEL2A and 2B top of canopy reflectance in the 3 10m-bands and 8 20m-bands using the PROSPECT+SAIL model.

A simple uncertainty model was used to better describe actual SENTINEL2 characteristics as well as the ability of the radiative transfer model used to represent actual reflectances. They are described in Table 7. The uncertainties are computed according to:

$$R^*(\lambda)=R(\lambda)(1+(MD(\lambda)+MI)/100)+AD(\lambda)+AI$$

Where $R(\lambda)$ is the raw simulated reflectance, $R^*(\lambda)$ is the reflectance contaminated with noise, MD is the multiplicative wavelength dependant noise, MI is the multiplicative wavelength independent noise, AD is the additive wavelength independent noise, and AI is the additive wavelength independent noise.

The uncertainties attached to the radiative transfer model mainly derive from the representation of canopy architecture and leaf and soil background optical properties which is difficult to estimate. A posterior estimation will be issued using the reflectance mis-match criterion as computed over actual SENTINEL2 data.

Sensor	SENTINEL2			
Input Bands	Add. Dep. (AD)	Add. Ind. (AI)	Mult. Dep. (MD) (%)	Mult. Ind. (MI) (%)
Band3	0.01	0.01	2	2
Band4	0.01		2	
Band5	0.01		2	
Band6	0.01		2	
Band7	0.01		2	
Band8	0.01		2	
Band8a	0.01		2	
Band11	0.01		2	
Band12	0.01		2	

Table 7. Characteristics of the uncertainties model used.

Figure 20 shows the distribution and co-distribution of simulated reflectances for SENTINEL2A (very similar figures were obtained for SENITNEL2B). Bands 3 and 5 appear very strongly correlated. The same is observed for bands 6 and 7 with 8a, and in a lesser way bands. For each band, the distributions are roughly Gaussian.

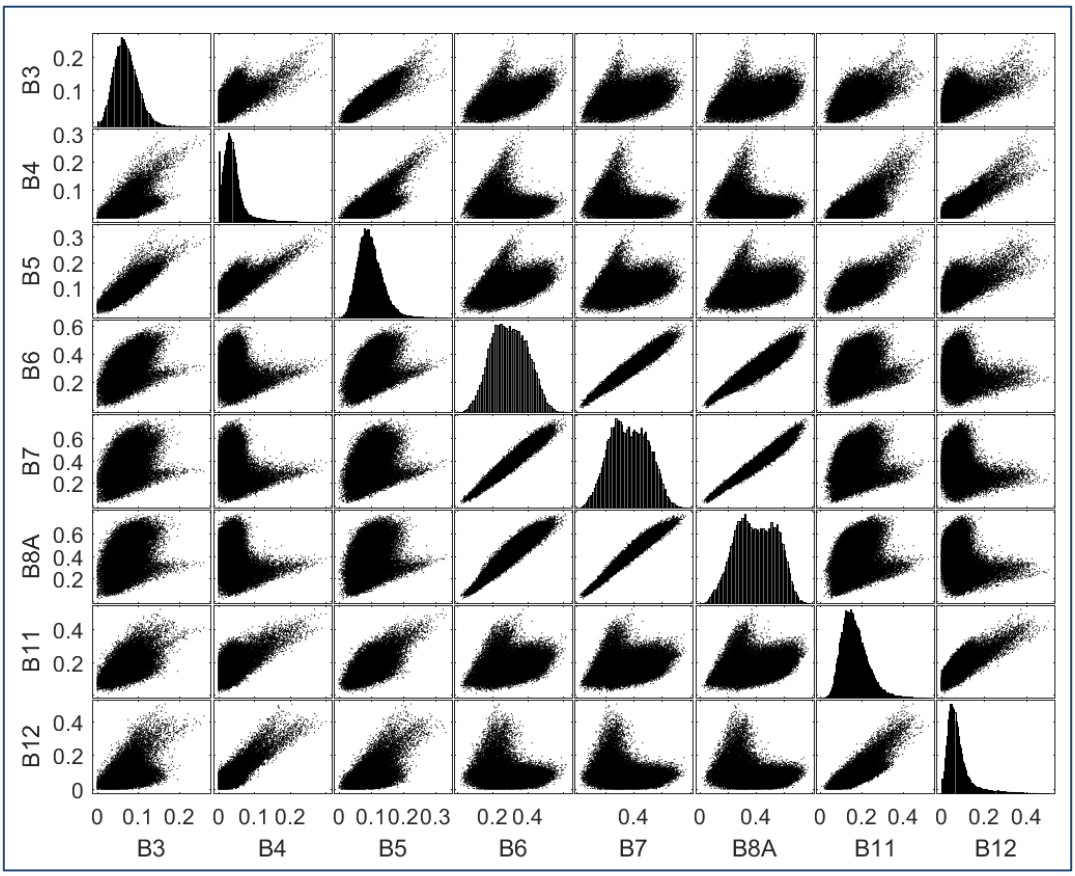


Figure 20: Distribution and co-distribution of the training database simulated reflectances, in the 8 SENTINEL2B bands (20m resolution).

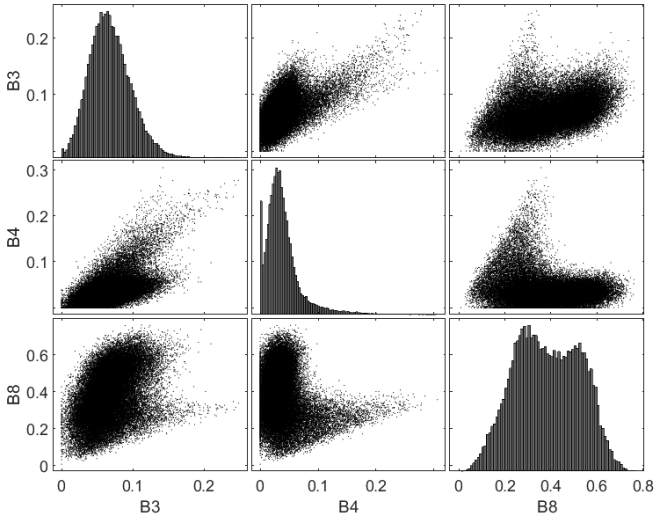


Figure 21: Distribution and co-distribution of the training database simulated reflectances, in the 3 SENTINEL2A bands (20m resolution).

3.3.4 Consequence on the distribution of the output variables

3.3.4.1 *Distribution of the output variables*

Figure 22 shows the distribution and co-distribution of the output variables. They are highly non Gaussian. Co distributions are of interest since they demonstrate that there is no single relationship between one output variable and another. Note that, as expected, the CWC and CCC are quite linearly correlated to LAI. Further, some relationships have already been investigated in detail over a range of vegetation types such as the LAI-FAPAR relationship. It may be the basis to check the consistency of the simulated outputs with regards to the expected behaviour.

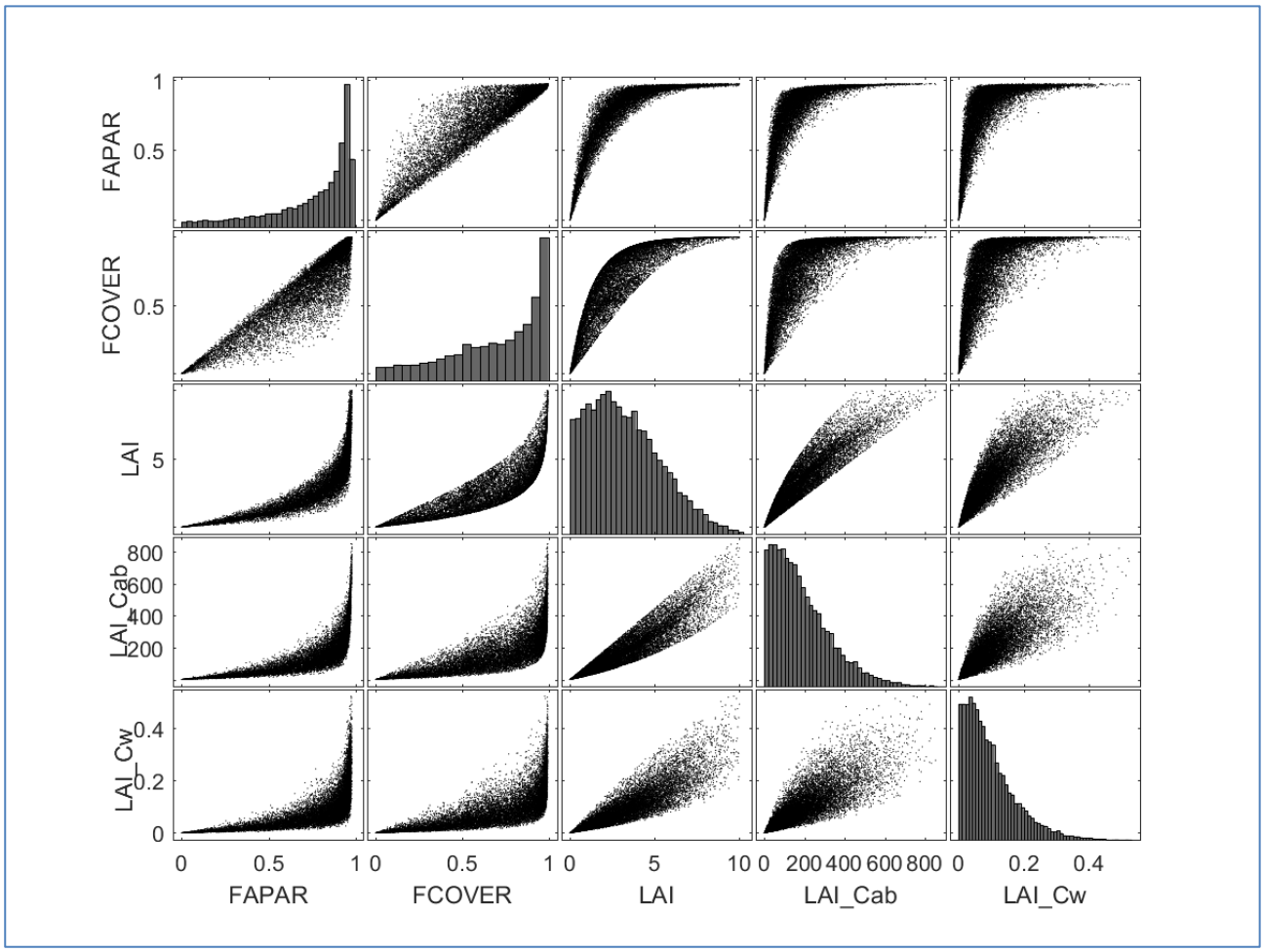


Figure 22. Distribution and co-distribution of the output variables.

3.3.4.2 *Relationship between LAI and FAPAR*

A number of experimental observations report that the relationship between LAI and FAPAR follows an exponential law with extinction coefficients showing a limited range of variability. Figure 23 (left) shows results derived from the compilation of VALERI 3x3 km² sites covering a large range of biome types Figure 23 right shows similar relationship as simulated for MODIS products using 3D radiative transfer model run for a range of canopy types. Both LAI-FAPAR relationships show very good

agreement. The distribution of cases simulated for SENTINEL2 products shows also a good consistency.

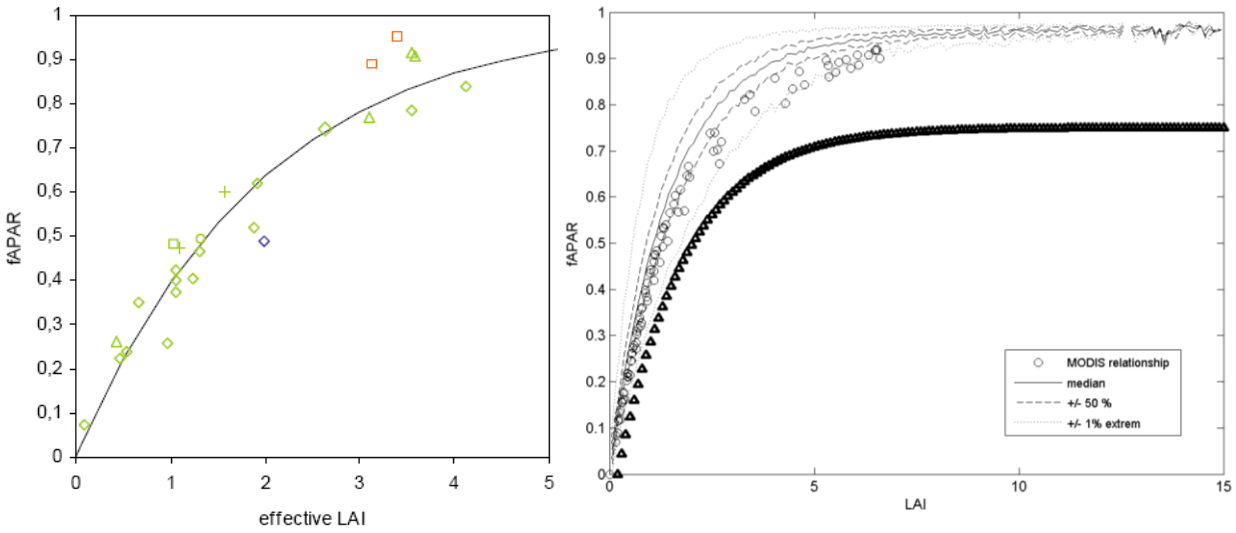


Figure 23: On the left, relationship between LAI and FAPAR as established over the ensemble of VALERI sites. Symbols correspond to the sites. The solid line corresponds to the best fit model. On the right, symbols correspond to MODIS C4 relationship between LAI and FAPAR as simulated by the 3D radiative transfer model for the several vegetation types considered. The several lines correspond to the distribution within the learning data base. The line with triangles represents a threshold under which values are not expected.

3.4 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 et al., 1986) is one of the most common neural networks used to solve our radiative transfer model inversion problem.

3.4.1 Normalization of the input and output values

The inputs (SENTINEL2A or SENTINEL2B TOC reflectance in three or eight bands depending on the considered spatial resolution, and geometry) and output (the biophysical variable considered) values are first normalized according to Equation 6. Such data transformation is performed mainly to increase the performances of convergence of the training algorithm.

$$X^* = 2*(X-XMin)/(XMax-XMin)-1 \qquad \text{Equation 6}$$

Where X* is the normalized input, X the original value, Xmin and Xmax respectively the minimum and maximum values.

3.4.2 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.

Although is it possible to train a single network for all the variables considered here, previous tests have shown that the training is more easy when targeting a single variable. This decreases significantly the complexity of the neural network architecture associated to multiple output variables. Further, these tests showed also that only marginal lost consistency between the variables to be estimated was observed when using individual neural nets for each variable as compared to a single neural net for all the variables.

The neural networks investigated that way are thus composed of (Figure 14):

- For the 20m resolution NNET: one input layer made of the 11 normalized input data ($\cos(\theta_s)$, $\cos(\theta_v)$, $\cos(\phi)$, and the TOC reflectances in the 8 SENTINEL2 wavebands). For the 10m resolution NNET: one input layer made of the 6 normalized input data ($\cos(\theta_s)$, $\cos(\theta_v)$, $\cos(\phi)$, and the TOC reflectances in the three 10m spatial resolution SENTINEL2 wavebands)
- one hidden layers with 5 neurons with tangent sigmoid transfer functions.
- one output layer with a linear transfer function.

Note that this simple network requires 65 (35) synaptic coefficients and 6 bias coefficients to adjust for 20m (10m) resolution.

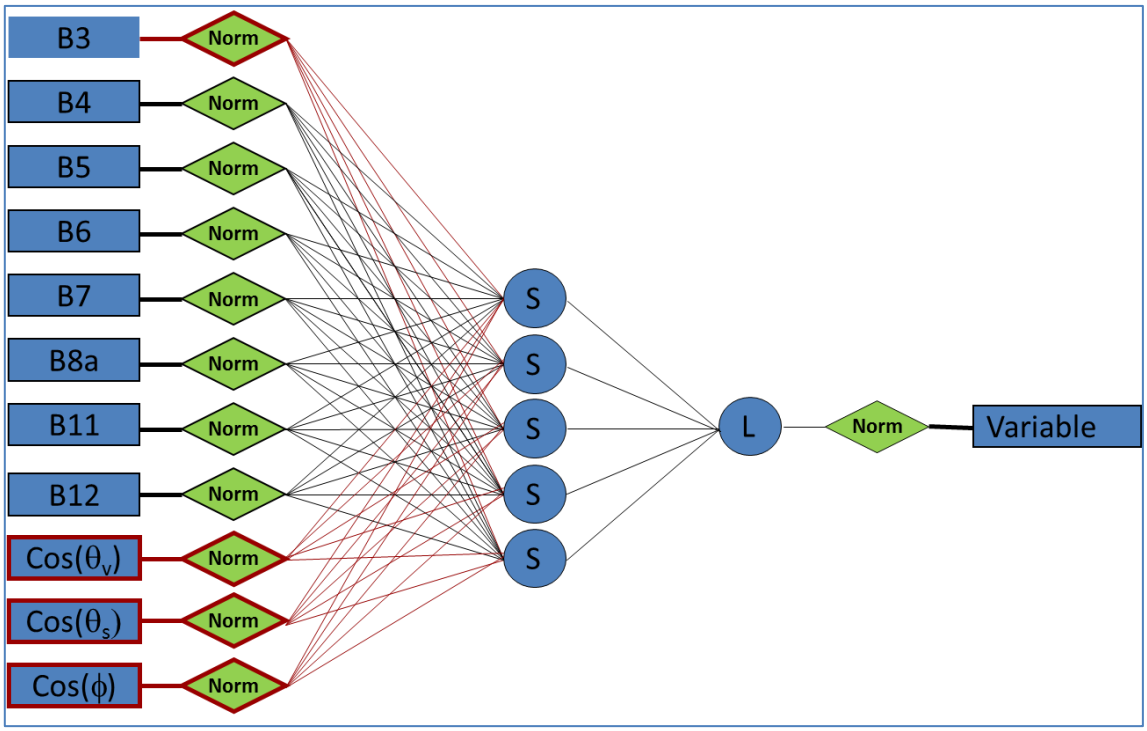


Figure 24. Neural network architecture developed for the estimation of the biophysical variables considered from the eight SENTINEL2 bands (20m spatial resolution) and the 3 angles defining the geometry of observation. The network is made of 1 hidden layer of 5 neurons and 1 linear output neuron. The 'Norm' symbols correspond to the normalization process as described by Equation 10. Symbols 'S' and 'L' correspond respectively to the sigmoid (tansig) and linear transfer functions of the neurons.

3.4.3 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 losing its capacity to describe the general features to be extracted. When this happens, 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. Finally, the network kept for implementation uses is the one that provides the best performances over the test data set.

3.4.4 Theoretical Performances

Theoretical performances allow checking whether no major problems occur during the training process and provide a first glance on the capacity of SENTINEL2 to access the considered variables. The theoretical performances of the networks were evaluated over the test dataset which is a fraction of the simulated training data base (1/3) not used in the calibration of the neural networks. It consists

mainly in computing simple statistics such as RMSE values and exploring the dependency of residuals on the variable itself.

Product	Product	S2A -10m	S2A-20m	S2B-10m	S2B-20m
LAI	R^2	0.71	0.82	0.71	0.82
	RMSE	1.13	0.90	1.15	0.90
fAPAR	R^2	0.92	0.95	0.91	0.95
	RMSE	0.072	0.054	0.072	0.055
FVC	R^2	0.95	0.98	0.95	0.97
	RMSE	0.059	0.041	0.059	0.042
CCCµg/cm²	R^2		0.84		0.84
	RMSE		57.99		57.22
CWCg/cm²	R^2		0.84		0.85
	RMSE		0.031		0.023

Table 8. Theoretical performances of the neural networks trained for each variables and using the spectral bands available at 10m resolution, and the spectral bands at 20m resolution.

Results show that the training of the networks was quite efficient, with relatively small RMSE values. FAPAR (Figure 26) and FVC (Figure 27) show the best performances as expected, with the larger RMSE values observed for the medium values of the products. The algorithms are unbiased as expected (Figure 25 to Figure 29. No early saturation effect is observed as a function of the product value: LAI seems to be well estimated up to values of LAI=6 although uncertainties increase with LAI values (Figure 25). The same applies for CCC and CWC since they are directly correlated with LAI (Figure 28 and Figure 29). The addition of spectral bands at 20m resolution increases the performances of the neural network for all the variables. As expected, there is almost no differences between S2A and S2B NNET performances as the spectral sensitivities are very similar.

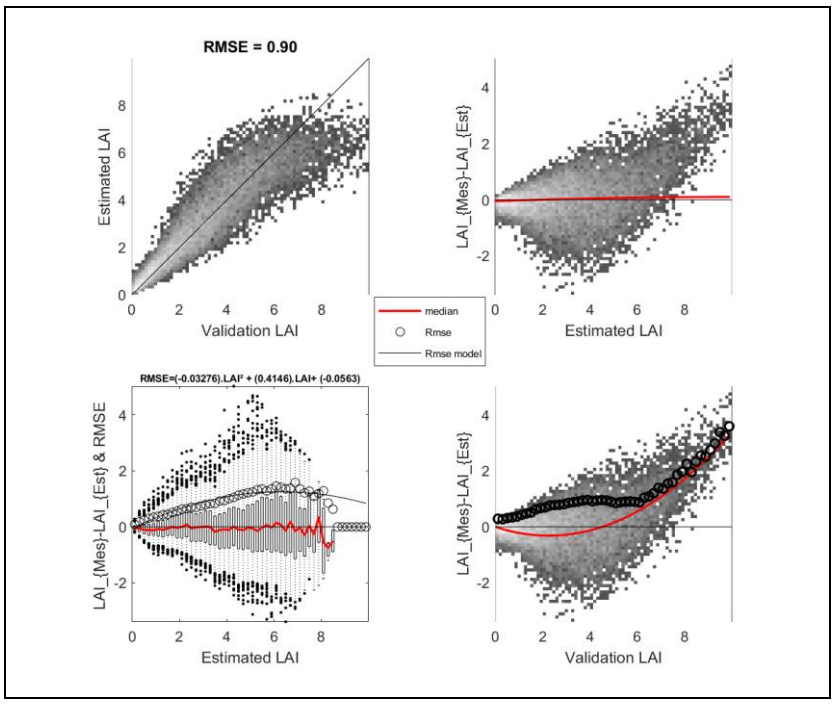


Figure 25. Theoretical performances of the neural networks for LAI on the test database (S2B-20m)

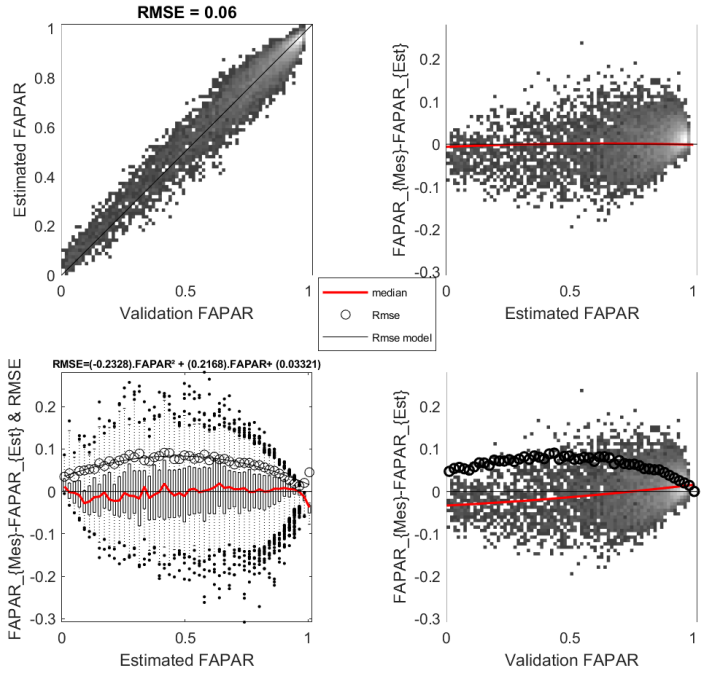


Figure 26. Theoretical performances of the neural networks for FAPAR on the test database (S2B-20m)

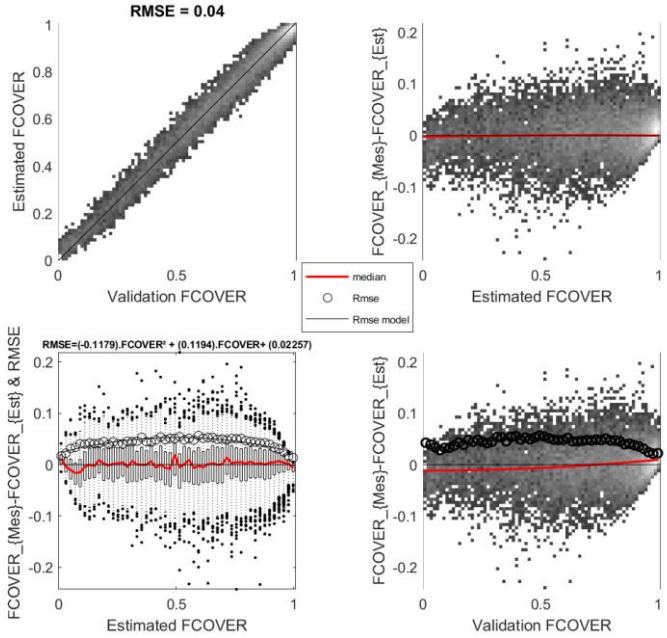


Figure 27. Theoretical performances of the neural networks for FCOVER on the test database (S2B-20m)

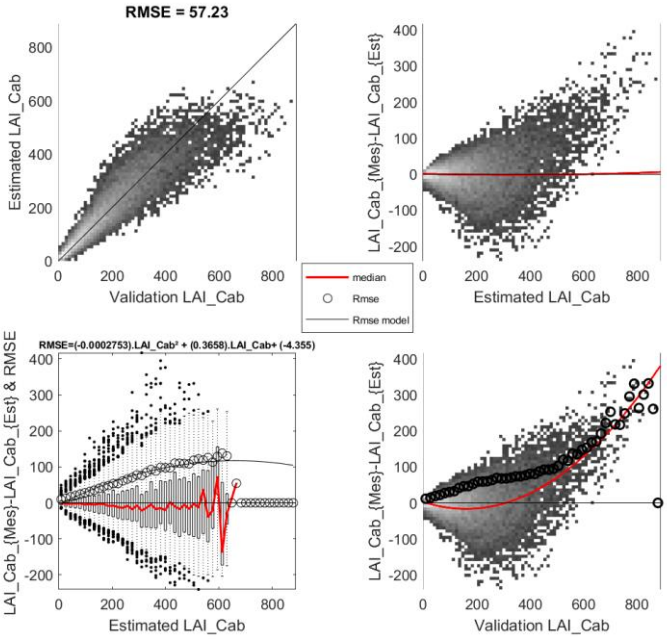


Figure 28. Theoretical performances of the neural networks for CCC on the test database (S2B-20m)

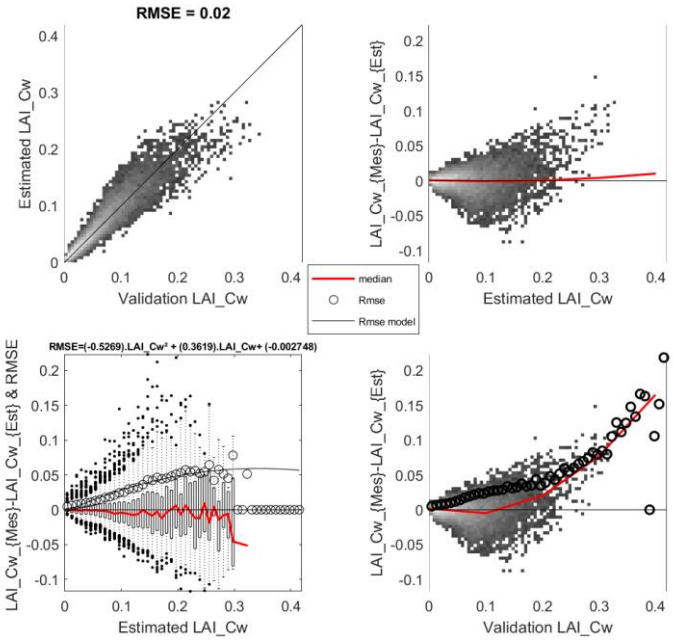


Figure 29. Theoretical performances of the neural networks for CWC on the test database (S2B-20m)

The intercomparison of LAI estimates between the different neural networks also show that a better consistency is reached when considering the same inputs (3 bands or 8 bands) of the neural networks (Figure 30, Bottom), with no bias observed (Figure 31, Bottom). This is mainly explained by the fact that similar information is used to assess the LAI, as the main differences between the neural networks training are due to the sensor spectral sensitivity. The S2A and S2B 3 bands neural networks provide more consistent results, with a constant scattering around the 1:1 line (except for very low values of LAI). Conversely, the S2A and S2B 8 bands neural networks are very consistent for LAI below 3 while a higher scattering is observed for high LAI values. This suggest that the training of the 8 bands neural networks may be less stable and thus more sensitive to the neuron weight initialization or to the neural networks architecture. A solution to better stabilize the 8 band neural network would be to consider an ensemble approach by training a number N of neural networks and keeping the median value of the estimates. However, this would require more computational power, which can not be considered from an operational point of view at this time. When comparing the 3bands and 8bands neural networks for the same sensor (Figure 30 and Figure 31, Top), a bias is observed with a systematic LAI overestimation by the 3 bands neural networks as compared to the 8 bands for high LAI, showing that, as confirmed by Table 8, the red-edge added in the 8 band neural network allow to get better estimates for high LAI values.

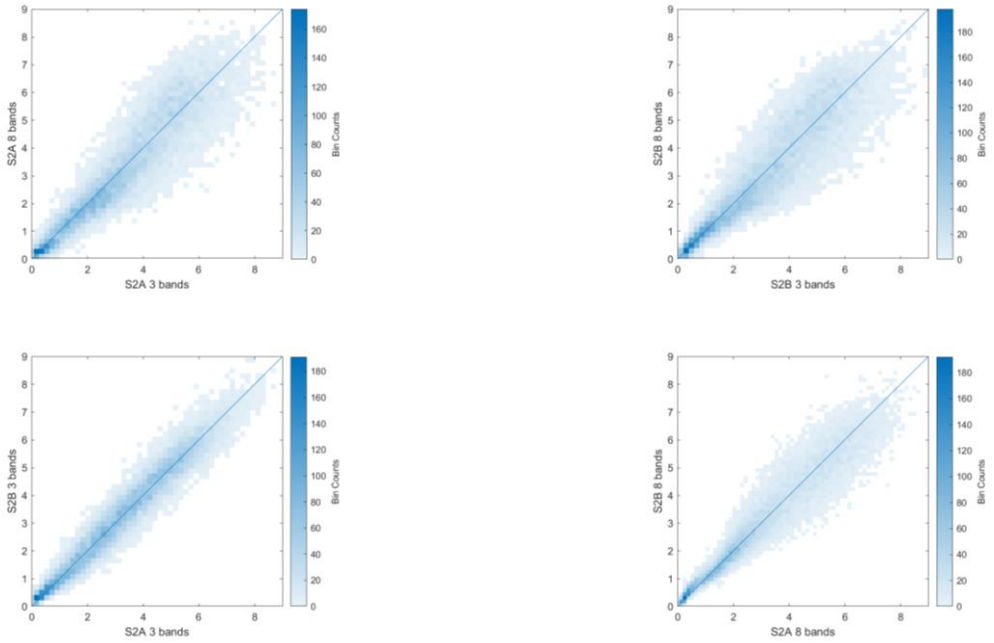


Figure 30. Intercomparison between the LAI estimates over the simulated validation dataset. Top Left: SENTINEL-2A 8 bands versus SENTINEL-2A 3 bands, Top Right: SENTINEL-2B 8 bands versus SENTINEL-2B 3 bands, Bottom left: SENTINEL 2A 3 bands versus SENTINEL-2B 3 bands, and bottom right: SENTINEL 2A 8 bands versus SENTINEL-2B 8 bands

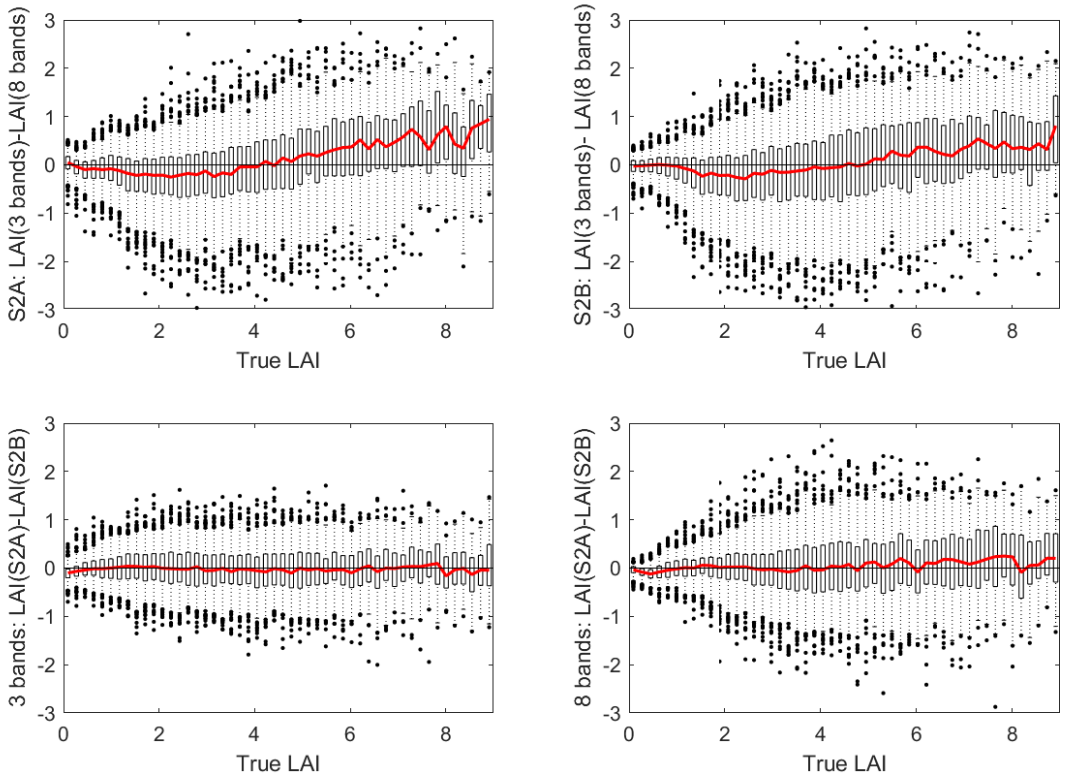


Figure 31. Distribution of the residuals between LAI estimates as a function of the true LAI value over the simulated validation database. Top Left: SENTINEL-2A 3 bands - SENTINEL-2A 8 bands, Top Right: SENTINEL-2B 3 bands - SENTINEL-2B 8 bands, Bottom left: SENTINEL 2A 3 bands - SENTINEL-2B 3 bands, and bottom right: SENTINEL 2A 8 bands - SENTINEL-2B 8 bands. Bars correspond to 25%, dots to individual values below 1% or over 99%. The solid red line corresponds to the median value.

3.5 FIELD OF APPLICATION OF THE NNET ALGORITHM

The algorithm developed to estimate the biophysical variables was built on radiative transfer simulations. Although some noise was considered during the training step (section 3.3), the Neural network algorithm may not be accurate in the following circumstances:

a. Unexpected noise in the reflectance inputs due to mis-cloud detection or significant atmospheric residual noise
b. Surfaces that do not correspond to vegetative areas.
c. Very specific soil backgrounds with sparse vegetation, such as bright sand (areas near desert) or inundated areas (rice crops for example)
d. Heterogeneous areas (mixed pixels) or canopy with a strong influence of the architecture, which would break the turbid assumption of the radiative transfer model used to build the neural network (forests, orchards,…)

To identify pixels that may be affected by this kind of errors, we developed two different flags. The first one consists in checking if the input reflectances, as well as the value of the variables estimated by the neural networks are consistent with what was used to train the database (section 3.5.1 for items a, b, and c). The second one consists in characterizing the heterogeneity of the pixels (section

3.5.2), indicating that the high heterogeneity could indicate a loss of accuracy in estimating the biophysical variables.

3.5.1 Definition Domain

3.5.1.1 *Input out of range flag*

When inputs are outside the convex hull defined by the simulated reflectance values of the training data base, i.e. the definition domain, then a specific 'input out of range' flag is raised. The convex hull is approximated by a hypercube with the same dimensions as those of the inputs of the neural network (i.e. 8 for SENTINEL2). Each dimension corresponding to a specific input of the neural network varies within the minimum and maximum values (Table 9). The range is split into 10 equal classes of values (see Figure 32 in the simple case of 2 inputs). However, because it was checked that geometrical configuration was more or less evenly distributed with regards to reflectance values in the considered bands, the geometrical configuration was not entered as input for the definition domain to simplify the process and make the code running faster. Results show that less than 50% of the cells the 8 dimension hypercube corresponds to expected reflectance observations over the surface.

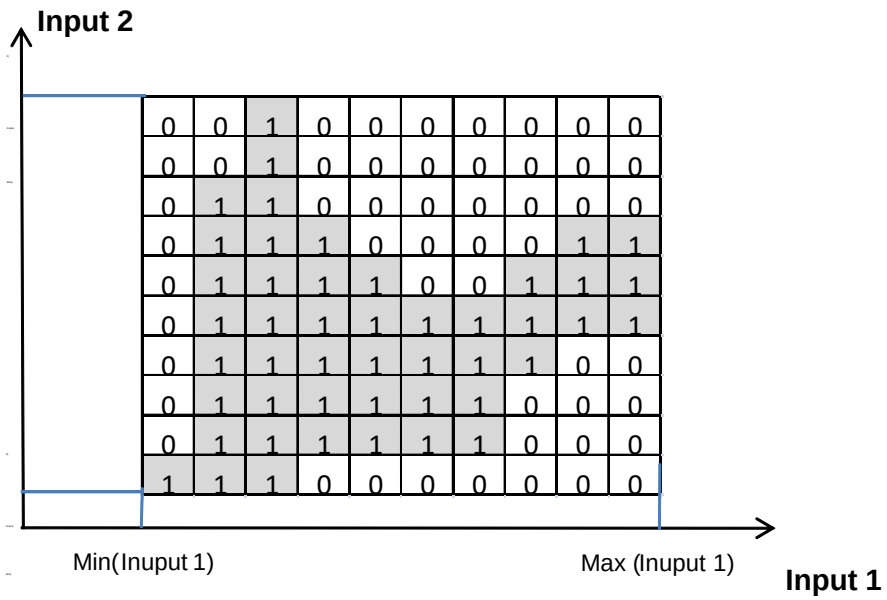


Figure 32. Schematic representation of the convex hull in the case of 2 dimensional inputs of the network. The convex hull is approximated by a regular grid. The cells with ones (the gray cells) correspond to cases of possible input range (i.e. inside the convex hull) while cells with zeros correspond to inputs out of range (outside the convex hull).

	Min	Max
Band3	0	0.25
Band4	0	0.31
Band5	0	0.32

Band6	0	0.60
Band7	0	0.74
Band8a	0	0.78
Band11	0	0.51
Band12	0	0.50

Table 9. Minimum and maximum input values (bounding box) of the definition domain for inputs from SENTINEL2A TOC reflectance (20m). Note that these values vary from one dataset to another and the user has to referred to the excel datasheet delivered with the neural network design description.

The algorithm used to check whether the inputs are outside the convex hull is the following:

- Check if the inputs are outside the bounding box, i.e. if the value of an input is either lower than the corresponding minimum or larger than the maximum values defined in Table 9. If this is the case, the 'input out of range' flag is raised.
- If inputs are within the bounding box, then, check if values of the inputs are in a non-valid cell, i.e. a cell with '0' (Figure 32). If this is the case, the 'input out of range' flag is raised. Note that a pixel flagged as 'input out of range' solely indicates that the algorithm has potentially failed to accurately estimate the biophysical variable interest, and that the obtained value should be checked further.

3.5.1.2 *Output out of range flag.*

In the case where the ANN provides biophysical variable estimates outside their definition range, the corresponding product value will be set to the closest bound of the range, i.e. either the minimum or the maximum accepted values. However, because of the several sources of uncertainties associated to the inputs and the algorithm calibration, a tolerance is set before raising the 'output out of range' flag: Three cases are possible for each product:

1. The product value is within expected physical range of variation (Table 10): It is considered valid.
2. The value is within the tolerance limits (Table 10) but higher (lower) than the physical maximum (minimum). The value is considered valid but set to the physical maximum (minimum).
3. The value is outside the tolerance limits: it is considered invalid.

	Tolerance	P_{min}	P_{max}
LAI	0.2	0	8.0
FAPAR	0.1	0	0.94
FVC	0.1	0	1.0
CCC	15	0	600

CWC	0.015	0	0.55

Table 10. Tolerance, and Minimum (P_{min}) and maximum (P_{max}) values admitted for the products.

3.5.2 Characterization of the heterogeneity of the pixel environment

The heterogeneity of a surface is often characterized using textural properties, such as with the Gray Level Occurrence Matrix (GLCM) and the associated Haralick features (Haralick et al., 1973). However, this type of method requires a significant computation time that does not allow operational implementation. For this reason, we considered a simple algorithm, initially developed by Wettle et al. (2004) to retrieve environmental noise (that comprises both instrument radiometric noise, atmospheric residual, and scene composition) from hyperspectral images for aquatic remote sensing. The method simply consists in computing the standard deviation of the pixel value over a growing size area (2x2 pixels, then 3x3, …, nxn) and compute the slope of the regression line with null intercept between those standard deviation. A null slope indicates that the area is very homogeneous while a higher slope indicates a more heterogeneous area.

To evaluate this method, we applied it on temporal series of SENTINEL2 images, over the BELMANIP2.1 site network in 2018 (Baret et al., 2006; Weiss et al., 2014b). This network (445 sites) was initially designed for the evaluation of medium resolution sensor products (Weiss et al., 2014a) and corresponds to 3x3km² sites composed of a single biome type (cropland, shrubs & savannas, needleleaf forests, deciduous broadleaf forest, evergreen broadleaf forests). For each of the image, we extracted the 3x3km² area, applied the neural network trained with the three bands (10m resolution) to estimate the leaf area index, and computed the LAI standard deviation over 20x20m², 30x30m², and 40x40m² areas. Figure 33 shows the distributions of the standard deviation values as a function of the considered area and for the different biome type defined in BELMANIP 2.1. As expected, the distribution of the standard deviation values does not vary significantly when increasing the size of the area of interest. Conversely, the distributions differ between biomes with crops, bare soils and savanna presenting much lower standard deviation values than forests. Indeed forest correspond to more heterogeneous areas and also are often characterized by high LAI which are less accurately estimated (see the noise observed in the scatterplot shown in Figure 25).

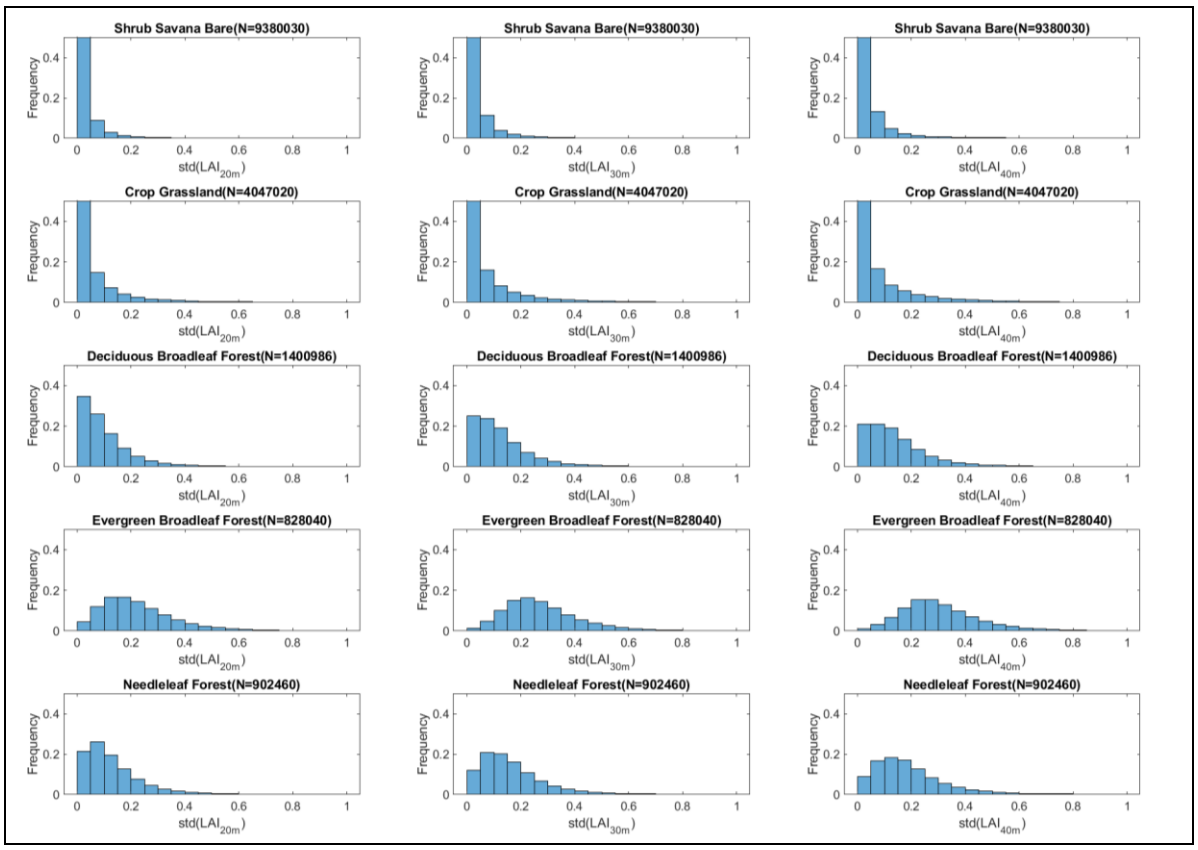


Figure 33. Distribution of the LAI standard deviation over 20x20m² (left), 30x30m², 40x40m², and 50x50m² (right) over the BELMANIP2.1 sites during year 2018. Each line is dedicated to a given biome type (top to bottom: shrub savanas and bare soils, crops and grasslands, deciduous broadleaf forests, Evergreen broadleaf forests and needleleaf forests. N is the number of pixels considered over the 2018 period and the BELMANIP2.1 sites.

The distribution of the slope between the LAI standard deviations computed over increasing areas (from 20x20m² to 40x40m²) differ for low vegetation and forests, suggesting that, as expected, forest areas are more heterogeneous than the other vegetation biomes. However, a low percentage of pixels having large slope values can still be observed in croplands, mainly due to the high heterogeneity of pixels located at the border of the fields. Moreover, as the BELMANIP2.1 sites were designed for medium resolution product validation, some of them are affected by the presence of buildings or very small urban areas that may affect the computation of the slope. The same applies for the edges of the fields for cropland areas. To further check this assumption, we applied the standard deviation slope algorithm computation over two SENTINEL2 tiles located in the South-East (Farming, Delta of the Rhone river) and South West (Montaneous area of Pyrenees) of France at one date. The standard deviation slope presents very high values at the field borders while it is quite low within fields. Conversely, values computed over the forested areas remain high.

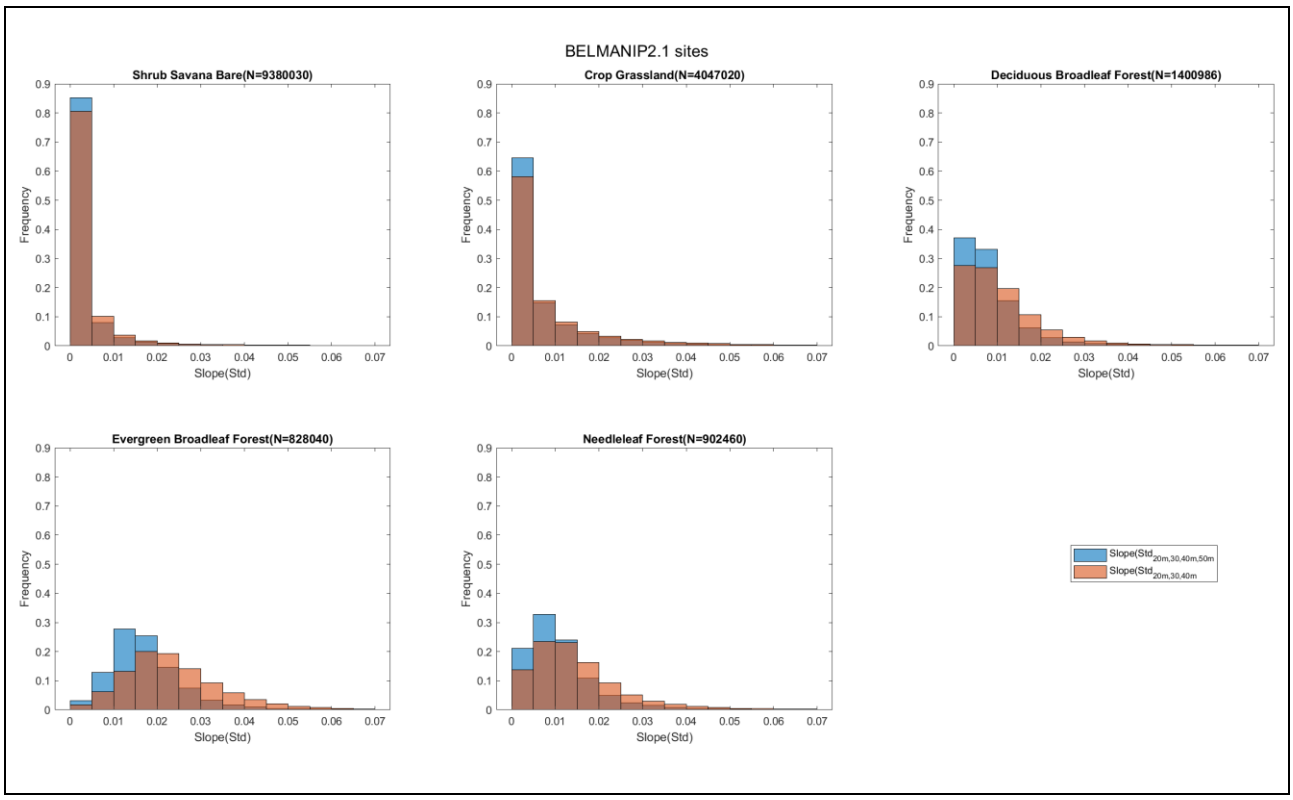
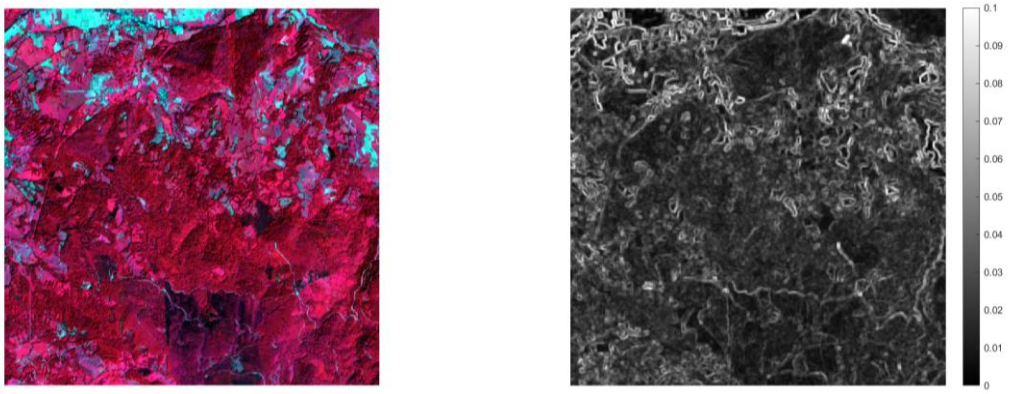


Figure 34. Distribution of the slope between LAI standard deviation computed over 20x20m², 30x30m², 40x40m², and 50x50m²) over the BELMANIP2.1 sites during year 2018 for given biome types N is the number of pixels considered over the 2018 period and the BELMANIP2.1 sites.



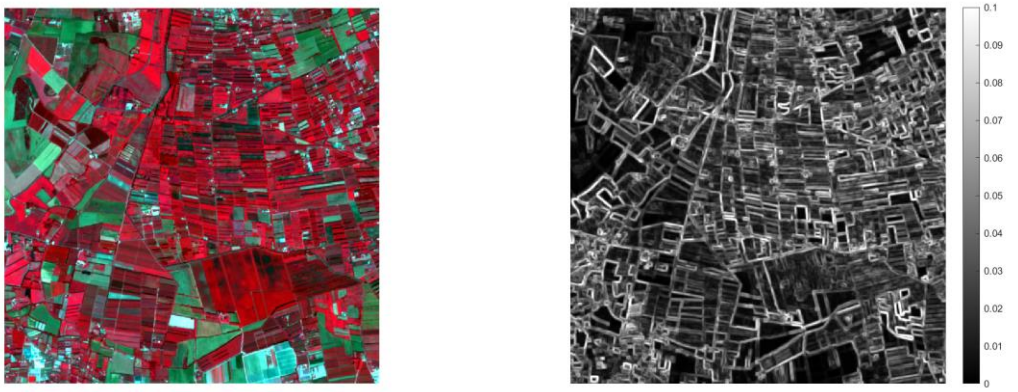


Figure 35. Left: Sentinel2 images extract over a mountainous area (South West of France, Pyrenees, top) and farming area (South East of France – Camargue, bottom). Right standard deviation slope computed over the area.

We therefore propose to provide the value of the slope associated to the standard deviations computed over the products over 2x2, 3x3, and 4x4 pixels as an additional indicator of the quality of the estimation, a high slope value indicating a higher heterogeneity and, thus a higher probability to lower the accuracy of the LAI estimation.

4 PRACTICAL CONSIDERATIONS FOR THE ALGORITHM IMPLEMENTATION

4.1 REQUIREMENTS FOR EASILY UPDATING THE ALGORITHM

To be able to easily upgrade the algorithm in the processing chain, all the coefficients used for the neural networks, normalization, quality flags and uncertainties should be set in parameter tables that can just be changed when updates will be available. This will occur possibly several times within a short period before launching the operational processing chain, when new product values will be available and when possible problems in the early versions will be quickly identified.

This information is available in excel files for each product: 'Algo_S2_VX_SL2T_VVV.xlsx ', where X satnds for the version of the algorithm, and VVV stands for the vegetation variable investigated, i.e. VVV=[LAI, FAPAR, FVC, CCC, CWC].

4.2 ALGORITHM IMPLEMENTATION

The different steps are described hereafter:

1. **Normalization of the inputs**: for the 12 inputs X, the following normalization equation must be applied:

$$X^*=2.(X-X_{min})/(X_{max}-X_{min}) - 1$$

where X* is nornalized input value, and Xmin and Xmax are computed over the neural network training data set. These values are provided in the Excel file under sheet 'Normalisation'

2. **Run the neural network**

The neural networks will be described by its architecture, *i.e*, the the number of hidden layers and the output layer. Each layer is described by its number of neurons, associated weight and biases and transfer function.

For the neurons of the hidden layers, the transfer function is a tangent sigmoid function given by:

$$Y=Tansig(x) = 2/(1+exp(-2x))-1$$

While for the output layer the transfer function is linear (y=x).

For each neural net (one per product and per sensor), tables will be provided given the weight and biases for each neurons. Example for the NNT provided for LAI is provided hereafter. Sheet 'Weights' in the Excel file contains the weights, biases and neural network structure information.

3. **Denormalization of the output**

It simply consists in applying the inverse function used for input normalization:

$$Y=0.5.(Y^*+1)*(Y_{max}-Y_{min}) +Y_{min}$$

where Y* is normalized output value issued from the NNT, and Ymin and Ymax are computed over the neural network training data set. These values are in the sheet 'Normalisation' in the Excel file contains the 'denormalisation' information.

4. Generate quality indicator for definition domain:

Generate the quality indicator QA coded over 3 bits,

QA=0 0 0 : data is OK

QA=0 0 1 : input out of range (provide product value if within tolerance)

QA=0 ,1 0 : output out of range (provide product value if within tolerance)

QA= 0 1 1 : input and output out of range (product value = fill value)

QA= 1 0 0 : bad quality of input values (depending on the TOC reflectance S2 product (provide product value if within tolerance)

5. Generate the heterogeneity indicator:

Coded over 8 bits: 0- 255, Scale Factor=2550 (slope of standard deviation between 0 and 0.1)

5 CONCLUSION

This ATBD describes the algorithm used to compute *LAI, FAPAR*, *FVC*, Canopy Chlorophyll and water content from SENTINEL2A and SENTINEL 2B top of canopy reflectance data that will be implemented in the SENTINEL2 Toolbox. The same methodology was applied to LANDSAT8 sensor (using exactly the same training database as for SENTINEL2A and 2B), to get compatible product estimation.

If the principles of the algorithm will be kept unchanged, it is likely that updates of the coefficients will be made when issues will be identified, data on actual models of uncertainties refined (including residual atmospheric effects). It is therefore important to keep the algorithm easy to upgrade. Refinement could be also made with actual SENTINEL2 top of canopy reflectance, allowing to filter the training data base so that only the simulated cases too different from the observed ones would be eliminated. This should result in better training and performances.

The proposed algorithm is based on 1D radiative transfer models associated with strong assumptions, particularly regarding canopy architecture (turbid medium model). All the variables derived from such algorithms should be seen as effective, i.e. the variables that would correspond to the measured satellite signal reflected by a canopy verifying all the assumptions made through the radiative transfer models. Depending on the variable, this may lead to differences with ground values that may be accessed from field measurements. Furthermore, the algorithm is 'generic', i.e. it should apply to any type of vegetation with reasonable performances. However, to better match the specificities of given canopies, either simple correction could be calibrated, or more specific algorithm could be developed. Results have also shown that despite a lower accuracy, especially for high LAI, a better consistency is found between 3bands neural network estimates than 8 bands. Ensemble approaches could be further considered to get more stable estimates. Anyway, when considering temporal series, it is highly recommended to consider compositing and gap filling techniques to improve the consistency of the temporal series.

Two neural networks per sensor were designed by considering either the 10m resolution bands (green, red, near infrared) or the 20m resolution (8 bands). When considering heterogeneous areas, we highly recommend to keep the spatial resolution unchanged and use the 3 band neural networks to avoid resampling applied to the 20m bands that may induce artefacts on LAI images showing some 20x20m² patches. One strong assumption embedded in any single pixel retrieval algorithm as this one, is that the pixel targeted belongs to a landscape patch presenting enough homogeneity (at the pixel scale) preventing unexpected loss or gain of radiation fluxes. Therefore, it can be applied for larger resolution than 20m. For forests with large crowns, or any pixel showing strong heterogeneity such as pixels at the intersection between two different vegetation patches, results may be uncertain. This extends also to pixels where the neighbouring ones are very different. We proposed an algorithm allowing to qualify these heterogeneous pixels and a flag has been added to indicate them.

6 REFERENCES

Asner, G.P., Wessman, C.A., Schimel, D.S., & 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., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data: principles and validation. *Remote Sensing of Environment, 105*, 313-325

Bacour, C., Jacquemoud, S., Leroy, M., Hautecoeur, O., Weiss, M., Prévot, L., Bruguier, N., & Chauki, H. (2002a). Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne PolDER data. *Agronomie, 22*, 555-566

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

Baret, F., & Buis, S. (2007). Estimating canopy characteristics from remote sensing observations. Review of methods and associated problems. In S. Liang (Ed.), *Advances in Land Remote Sensing: System, Modeling, Inversion and Application* (pp. 171-200): Springer

Baret , F., De Solan, B., & Weiss, M. (2009). PAI estimates from digital photos at 57.5° zenith angle over wheat crops. *Agricultural and Forest Meteorology, soumis (Février 2009)*

Baret, F., Jacquemoud, S., & Hanocq, J.F. (1993). The soil line concept in remote sensing. *Remote Sensing Reviews, 7*, 65-82

Baret, F., Koetz, B., & Bruguier, N. (2002). WP 1400: Characterization of maize structure and optical properties. In (p. 6). Avigon (France): INRA-CSE,

Baret, F., Leroy, M., Roujean, J.L., Knorr, W., Lambin, E., & Linderman, M. (2003). CYCLOPES User Requirement Document. In. Avignon: INRA-CSE,

Baret, F., Morissette, J., Fernandes, R., Champeaux, J.L., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., & Nickeson, J.E. (2006). Evaluation of the representativeness of networks of sites for the validation and inter-comparison of land biophysical products. proposition of the CEOS-BELMANIP. *IEEE Trans. Geosc. Remote Sens., 44*, 1794-1803

Baret, F., Weiss, M., Leroy, M., Hautecoeur, O., Santer, R., & Begue, A. (1997). Impact of surface anisotropies on the observation of optical imaging sensors. In: INRA Bioclimatologie,

Biard, F., & Baret, F. (1997). Crop residue estimates using multiband reflectance data. *Remote Sensing of Environment, 59*, 530-536

Borel, C.C., Gerstl, S.A.W., & 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., & Bouatouch, K. (1997). Nested radiosity for plant canopies. *The Visual Computer, Nov1997-Mar1998*, 1-24

Chen, J.M., Menges, C.H., & Leblanc, S.G. (2005). Global mapping of foliage clumping index using multi-angular satellite data. *Remote Sensing of Environment, 97*, 447-457

Chen, Y., & 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., Baret, F., & Weiss, M. (2002a). Improving canopy variable estimation from remote sensing data by exploiting ancillary information. Case study on sugar beet canopies. *Agronomie, 22*, 205-215

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., & Wang, L. (2002b). 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., & Beale, M. (1998). Neural network user's guide. In T. Mathworks (Ed.), *Matlab User's guide*

España, M., Baret, F., Aries, F., Chelle, M., Andrieu, B., & Prévot, L. (1999). Modeling maize canopy 3D architecture. Application to reflectance simulation. *Ecological modeling, 122*, 25-43

Fourty, T., & Baret, F. (1997). Amélioration de la précision des coefficients d'absorption spécifique de la matière sèche et des pigments photosynthétiques. In (p. 35). Avignon: INRA Bioclimatologie,

Garrigues, S., Allard, D., Baret, F., & Weiss, M. (2006a). Influence landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data. *Remote Sensing of Enviroment, 105*, 286-298

Garrigues, S., Allard, D., Baret , F., & Weiss, M. (2006b). Quantifying Spatial Heterogeneity at the Landscape Scale Using Variogram Models. *Remote Sensing of Enviroment, 103*, 81-96

Garrigues, S., Shabanov, N.V., Swanson, K., Morisette, J.T., Baret, F., & Myneni, R.B. (2008). Intercomparison and sensitivity analysis of Leaf Area Index retrievals from LAI-2000, AccuPAR, and digital hemispherical photography over croplands. *Agricultural and Forest Meteorology, 148*, 1193-1209

Gascon, F., & Berger, M. (2007). Sentinel-2 Mission Requirements Document In ESA (Ed.) (p. 30): ESA,

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

Gausman, H.W., Gerbermann, A.H., Wiegand, C.L., Leamer, R.W., Rodriguez, R.R., & Noriega, J.R. (1975). Reflectance differences between crop residues and bare soils. *Soil science society of America proceedings., 39*, 752-755

Gerstl, S.A.W., & Borel, C.C. (1992). Principles of the radiosity method versus radiative transfer for canopy reflectance modeling. *IEEE Transactions on Geoscience and Remote Sensing, 30*, 271-275

Govaerts, Y.M., & 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*, 493-505

Green, D.S., Erickson, J.E., & Kruger, E. (2003). Foliar morphology and canopy nitrogen as predictors of light use efficiency in terrestrial vegetation. *Agricultural and Forest Meteorology, 115*, 163-171

Haralick, R.M., Shanmugam, K., & Denstien, I. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern., 3*, 610–621

Houlès, V., Mary, B., Machet, J.M., Guérif, M., & Moulin, S. (2001). Do crop characteristics available from remote sensing allow to determine crop nitrogen status? In G. Grenier, & S. Blackmore (Eds.), *3rd European Conferance on Precision Agriculture* (pp. 917-922). Montpellier: Agro Montpellier,

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

Jacquemoud, S., Baret, F., & Hanocq, J.F. (1992). Modeling spectral and directional soil reflectance. *Remote sensing of the Environment, 41*, 123-132

Kimes, D.S., Gastellu-Etchegorry, J.P., & Esteve, P. (2002). Recovery of forest canopy characteristics through inversion of complex 3D model. *Remote Sensing of Environment, 79*, 320-328

Kötz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment, 95*, 115-124

Kötz, B., Kneubuehler, M., Huber, S., Schopfer, J., & Baret, F. (2007). Radiative transfer model inversion based on multi-temporal CHRIS/PROBA data for LAI estimation. In, *ENVISAT symposium*. Montreux: ESA,

Kuusk, A. (1991). The hot-spot effect in plant canopy reflectance. In (pp. 1-16): Tartu University,

Lauvernet, C., Baret, F., Hascoët, L., Buis, S., & Le Dimet, F.X. (2008). Multitemporal-patch ensemble inversion of coupled surface-atmosphere radiative transfer models for land surface characterization. *Remote Sensing of Enviroment, 112*, 851-861

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. In (p. 45). Orsay (France): Laboratoire d'Ecologie Végétale, université Paris-Sud,

Liu, W., Baret, F., Gu, X., Tong, Q., Zheng, L., & Zhang, B. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment, 81*, 238-246

Masson, V., Champeaux, J.L., Chauvin, F., Meriguer, C., & Lacaze, R. (2003). A global database of land surface parameters at 1km resolution in meteorological and climate models. *Journal of Climate, 16*, 1261-1282

Myneni, R.B., Asrar, G., & 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., & 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*

Rummelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning internal representations by error propagation. In D. Rummelhart, & J. Mc Clelland (Eds.), *Parallel data processing* (pp. 318-362). Cambridge, MA (USA): M.I.T. press

Sanz, C., Espana, M., Baret, F., Weiss, M., Vaillant, L., Hanocq, J.F., Sarrouy, C., Clastre, P., Bruguier, N., Chelle, M., Andrieu, B., & Zurfluh, O. (1997). Bi-directional characteristics of leaf reflectance and transmittance: measurement and influence on canopy bi-directional reflectance. In G. Guyot, & T. Phulpin (Eds.), *7th International Symposium on physical measurements and signatures in remote sensing* (pp. 583-590). Courchevel (France): Balkema,

Scurlock, J.M.O., Asner, G.P., & Gower, S.T. (2001). Worldwide Historical Estimates and Bibliography of Leaf Area Index, 1932-2000. In. Oak Ridge, Tennessee, U.S.A.: Oak Ridge National Laboratory,

Soler, C., F., S., Blaise, F., & de Reffye, P. (2001). A physiological plant growth simulation engine based on accurate radiant energy transfer. In (p. 31). Montbonnot-Saint-Martin (France): INRIA,

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

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

Weiss, M., Baret, F., Block, T., Koetz, B., Burini, A., Scholze, B., Lecharpentier, P., Brockmann, C., Fernandes, R., Plummer, S., Myneni, R., Gobron, N., Nightingale, J., Schaepman-Strub, G., Camacho, F., & Sanchez-Azofeifa, A. (2014a). On Line Validation Exercise (OLIVE): A Web Based Service for the Validation of Medium Resolution Land Products. Application to FAPAR Products. *Remote Sensing, 6*, 4190-4216

Weiss, M., Baret, F., Leroy, M., Hautecœur, O., Bacour, C., Prévot, L., & Bruguier, N. (2002). Validation of neural net techniques to estimate canopy biophysical variables from remote sensing data. *Agronomie, 22*, 547-553

Weiss, M., Baret, F., Myneni, R., Pragnère, A., & 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

Weiss, M., Baret, F., Smith , G.J., Jonckheered, I., & Coppin, P. (2004). Review of methods for in situ leaf area index determination, part II: Estimation of LAI, errors and sampling. *Agricultural and Forest Meteorology, 121*, 37-53

Weiss, M., Baret, F., Verger, A., Lacaze, R., D.Ramon, L.Wandrebek, & B.Smets (2014b). Near real time, global, LAI, FAPAR and cover fraction products derived from PROBA-V, at 300m and 1km resolution. In, *Global Vegetation Monitoring and Modeling (GV2M)*. Avignon, France

Wettle, M., Brando, V.E., & Dekker, A.G. (2004). A methodology for retrieval of environmental noise equivalent spectra applied to four Hyperion scenes of the same tropical coral reef. *Remote Sensing of Environment, 93*, 188-197

Zarco-Tejada, P.J., Miller, J.R., Harton, J., Hu Baoxin, N., T.L., Goel, N., Mohammed, G.H., & Sampson, P. (2001). Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies. *Remote Sensing of Environment, submitted*