IMPLEMENTING MULTI-SCALE AGRICULTURAL INDICATORS EXPLOITING SENTINELS

ATBD OF 10-M (30-M) FAPAR FOR S2 (LANDSAT-8) AND S3 (+ PROBA-V)

IMAGINES_RP2.5

ISSUE 1.0

EC Proposal Reference N° FP7-311766

Due date of deliverable: 04.12.2014

Actual submission date: 04.12.2014

Start date of project: 01.11.2012   **Duration : 40 months**

Name of lead partner for this deliverable: INRA

Book Captain:	Frédéric BARET
Contributing Authors:	Marie Weiss
	Wenjuan Li

Project co-funded by the European Commission within the Seventh Framework Program (2007-2013)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

DOCUMENT RELEASE SHEET

Book Captain:	F. Baret	Date: 04.12.2014	Sign.
Approval:	R. Lacaze	Date: 18.12.2014	Sign.
Endorsement:	I. Marin-Moreno	Date:	Sign.
Distribution:			

CHANGE RECORD

Issue/Revision	Date	Page(s)	Description of Change	Release
	04.12.2014	All	First issue	I1.00

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ALA	Average leaf angle
ATBD	Algorithm Theoretical Based Document
DART	Discrete Anisotropic Radiative Transfer
ECV	Essential Climate Variable
EU	European Union
DN	Digital Number
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FCOVER	Fraction of vegetation cover
GAI	Green Area Index
GCOS	Global Climate Observation System
GEO	Group on Earth Observations
GEOV1	VGT bioGEOphysical product Version 1 (geoland2 project)
GEOV2	VGT bioGEPphysical product Version 2 (geoland2 project)
GEOV3	PROBA-V bioGEOphysical product Version 3 (ImagineS project)
GEOGLAM	Global Agricultural Geo-Monitoring Initiative
GTOS	Global Terrestrial Observation System
HOT	Hot spot parameter
LAI	Leaf Area Index
LDAS	Land Data Assimilation System
MACCS	Multi-mission Atmospheric and Cloud Correction Software
NIR	Near Infrared spectral domain
NN	Neural Network
OLI	Operational Land Imager
PROSPECT	Proprietes spectrales (a model of leaf optical properties)
QC	Quality Control
RMSE	Relative Mean Square Error
RTM	Radiative Transfer Model
SAA	Sun Azimuth Angle
SAIL	Scattering by Arbitrarily Inclined Leaves
SLW	Specific Leaf Weight
SWIR	Short Wave Infrared
SZA	Sun Zenith Angle
THEIA	Pole Thematique Surfaces Continentales
TIRS	Thermal Infrared Sensor
TOC	Top of Canopy
USGS	United States Geological Survey
VAA	Viewing Azimuth Angle
VALERI	Validation of Land European Remote sensing Instruments
VZA	Viewing Zenith Angle

1. BACKGROUND OF THE DOCUMENT

1.1. EXECUTIVE SUMMARY

The Copernicus program is the EU response to the increasing demand for reliable environmental data. The objective of the Copernicus Land Service is to continuously monitor and forecast the status of land territories and to supply reliable geo-information to decision makers, businesses and citizens to define environmental policies and take right actions. ImagineS intends to continue the innovation and development activities to support the operations of the Copernicus Global Land service, preparing the use of the new Earth Observation data, including Sentinels missions data, in an operational context. Moreover, ImagineS aims to favor the emergence of downstream activities dedicated to the monitoring of crop and fodder production, that are key for the implementation of the EU Common Agricultural Policy, of the food security policy, and could contribute to the Global Agricultural Geo-Monitoring Initiative (GEOGLAM) coordinated by the intergovernmental Group on Earth Observations (GEO).

The main objectives of IMAGINES are to (i) improve the retrieval of basic biophysical variables, mainly LAI, FAPAR and the surface albedo, identified as Terrestrial Essential Climate Variables, by merging the information coming from different sensors (PROBA-V and Landsat-8) in view to prepare the use of Sentinel missions data; (ii) develop qualified software able to process multi-sensor data at the global scale on a fully automatic basis; (iii) complement and contribute to the existing or future agricultural services by providing new data streams relying upon an original method to assess the above-ground biomass, based on the assimilation of satellite products in a Land Data Assimilation System (LDAS) in order to monitor the crop/fodder biomass production together with the carbon and water fluxes; (iv) demonstrate the added value of this contribution for a community of users acting at global, European, national, and regional scales.

The Algorithm Theoretical Basis Document (ATBD) provides the proposed algorithm for generating biophysical variables from Landsat-8 surface reflectance data at 30 m resolution. The considered products are the following variables: LAI and FAPAR that are Essential Climate Variables (ECVs) as recognized by international organizations such as GCOS and GTOS. In addition the FCOVER variable will also be generated since it corresponds to specific needs for some users. The associated uncertainty and the quality flag are also provided.

A further version of this ATBD will include the description of the PROBA-V dis-aggregation method using Landsat-8 to provide biophysical variables per land cover class within each 300m pixel. As no Sentinel2 and -3 data are available at the time of this study, the implementation is achieved using Landsat-8 and PROBA-V data to demonstrate the capabilities of the algorithm.

1.2. SCOPE AND OBJECTIVES

One of the main objectives of ImagineS is to prepare the use of new Earth Observation data, such as the Sentinels mission data. The Landsat-8 data, launched in February 2013, has a high spatial resolution. It can be used as a proxy to test the feasibility of the algorithm at a similar spatial resolution of Sentinels data. The algorithm developed under the IMAGINES project and run in pre-operational processing chains will provide to the scientific community as well as other stakeholders including policy makers, the proper information required for several applications.

The objective of this document is to provide a detailed description and justification of the algorithm based on Landsat-8 surface reflectance products.

1.3. CONTENT OF THE DOCUMENT

This ATBD is structured as follows:

1. Overview of the algorithm. This section contains:

- A definition of the proposed products
- A brief description of the Landsat-8 instruments and data from which the products will be derived
- The outline of the algorithm

2. Description of the algorithm. This section contains:

- The input variables and output products provided by the algorithm
- The detailed algorithm used. It includes the reflectance models and input parameters required to generate the neural network, and the way to inverse the products.

1.4. RELATED DOCUMENTS

1.4.1. Inputs

Overview of former deliverables acting as inputs to this document.

Document ID	Descriptor
ImagineS_RP1.1	User Requirements Document
ImagineS_RP1.2	Service Specifications Document

ImagineS_RP2.1_ATBD   ATBD of LAI, FAPAR, FCover at 300m from PROBA-V (GEOV3)

ImagineS_RP5.1   Detailed Processing Model

1.4.2. Output

Overview of other deliverables for which this document is an input:

Document ID	Descriptor
ImagineS_D5.10	Decametric products over demonstration sites
ImagineS_D5.11	High resolution production line
ImagineS_RP6.3	Product User Manual
ImagineS_RP7.4	Validation report

1.4.3. External Documents

Document ID	Descriptor
GIOGL1_ATBD_GEOV2	Algorithm Theoretical Basis Document of LAI, FAPAR, FCover Version 2 (GEOV2) products

2. OVERVIEW

2.1. THE CONSIDERED PRODUCTS

2.1.1. 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 scale dependent [*Garrigues et al.*, 2006; *Weiss et al.*, 2000]. Note that vegetation LAI as estimated from remote sensing will include all the green contributors such as the 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 provides the same diffuse gap fraction while assuming a random distribution of leaves. The difference between the actual LAI and the 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). Note that similarly to the other variables, the retrieved LAI is mainly corresponding to the green elements: the correct term to be used would be GAI (Green Area Index) although we propose to still use LAI for the sake of simplicity.

2.1.2. FAPAR

FAPAR corresponds to the fraction of photosynthetically active radiation absorbed by the canopy. It is an essential climate variable in characterizing energy, mass and momentum exchanges between the canopy and the atmosphere and is an important input to a number of primary productivity models based on simple efficiency considerations [*McCallum et al.*, 2009; *Prince*, 1991]. FAPAR depends on canopy structure, vegetation element optical properties and illumination conditions. 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 daily integrated FAPAR values for direct radiation as well as the FAPAR value computed for diffuse conditions. To be consistent with previous FAPAR products that are considering the instantaneous FAPAR value at the time of the satellite overpass assuming only direct radiation, a study investigated the differences between alternative FAPAR definitions [*Baret et al.*, 2004]. Results show that the instantaneous FAPAR value at 10:00 solar time is very close to the daily integrated value under only direct condition. Moreover, the FAPAR for the diffuse radiation was also provided assuming that the diffuse radiation is isotropic. Note that the FAPAR refers only to the green parts (leaf chlorophyll

content higher than 15 ug cm^{-2}) of the canopy. FAPAR is relatively linearly related to the reflectance values, and is little sensitive to scaling issues [*Weiss et al.*, 2000].

2.1.3. FCOVER

It corresponds to 1- the gap fraction for nadir direction. FCOVER 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, FCOVER is only marginally scale dependent [*Weiss et al.*, 2000]. Note that similarly to LAI and FAPAR, only the green elements are considered.

2.2. LANDSAT8 INSTRUMENT CHARACTERISTICS

The Landsat-8, launched on February 11th, 2013, is a latest spacecraft of more than 40 years Landsat data archive. The characteristic of the orbit are presented in Table 1.

Characteristics	Values
Orbit type	Sun-synchronous
Orbit altitude (km)	705
Swath width (km)	185
Field of view (˚)	15
Inclination (˚)	98.2
Period (min)	99
Repeat coverage (days)	16
Equatorial descending node crossing time (hr)	10:00

Table 1: Orbit characteristics of Landsat-8

Two sensors are carried on it. The Operational Land Imager (OLI) sensor includes nine spectral bands with a spatial resolution of 30m for Bands 1 to 7 and 9. The resolution for Band 8 is 15 meters. Compared with previous Landsat instrument, two new bands were added: a deep blue band was added for coastal/aerosol studies, and a shortwave infrared band for cirrus detection. The bands characteristics of OLI sensor are presented in Table 2. The Thermal Infrared Sensor (TIRS) sensor provides two thermal bands which are useful in the surface temperature studies.

No. band	Band name	Width (nm)	Central (nm)	Spatial resolution (m)
B1	Deep blue	427-459	443	30
B2	Blue	436-528	482	30
B3	Green	512-610	561	30
B4	Red	625-691	658	30
B5	NIR	829-900	865	30
B6	SWIR-1	1515-1697	1606	30
B7	SWIR-2	2037-2355	2196	30
B8	Pan	488-692	590	60
B9	Cirrus	1340-1409	1375	30

Table 2: Landsat-8 OLI spectral characteristics

2.3. REQUIREMENTS FOR THE ALGORITHM SELECTION AND DESIGN

The objective is to develop an algorithm to estimate the LAI, FAPAR and FCOVER from the Landsat-8 observations at 30m spatial resolution. The algorithm will run at the pixel level without interactions with the surrounding pixels. The products should be associated with quality assessment flags as well as quantified uncertainties.

2.4. BACKGROUND

The European Commission within the GMES/Copernicus initiative supported the development of a series of biophysical products to be used within operational services. The first series of products were derived from kilometric resolution sensors such as VEGETATION and AVHRR, recently completed by PROBA-V. The products included LAI, FAPAR and FCOVER. The first series of product called CYCLOPES was developed within the CYCLOPES FP5 European project [*Baret et al.*, 2007]. It was tuned for the VEGETATION sensors. It was based on 2 steps. The first one was consisting in deriving dekadal top of canopy nadir like reflectances from the accumulation of observations within a fixed 30 days temporal compositing window. Then, a machine learning method (neural networks) trained over radiative transfer model simulations was used to estimate the corresponding dekadal LAI, FAPAR and FCOVER products. The products were quite well received by the community and were characterized by a very good smoothness and good performances for the low to median amount of vegetation [*Garrigues et al.*, 2008; *Weiss et al.*, 2007]. However, some limitations were identified, mainly coming by the pre-processing steps that create artefacts for the higher latitudes and a significant fraction of missing data due to the restricted length of the compositing window. In the meantime, the validation exercise was getting more mature, allowing inter-comparing the several available products.

This was the basis of the following series of products developed within the geoland2 FP7 European project: The GEOV1 products were capitalizing on the previous development efforts [*Baret et al.*, 2013]. They are expected to take benefit from the advantages of the available products, while minimizing their weaknesses. They are based on machine learning methods, but applied to the instantaneous daily observations available at the same time as the dekadal top of canopy reflectance values as derived from the CYCLOPES pre-processing steps. The radiative transfer simulations used in the CYCLOPES products were thus replaced by the actual products values derived from MODIS C5 and CYCLOPES V3.1. Significant improvements were observed as compared to MODISC5 and CYCLOPES products [*Camacho et al.*, 2013]. However, problems were remaining for the higher latitudes and a significant fraction of missing products was observed similarly to GEOV1. Within the end of geoland2 and the start of the Copernicus Global Land service, the GEOV2 was developed to improve the performances of GEOV1 and provide near-real time products required by a number of applications. GEOV2 is still based on the training of a neural network using CYCLOPES and MODISC5 products [*Verger et al.*, 2014]. However, the compositing is now achieved on the biophysical products rather than on the surface reflectance as done in CYCLOPES and GEOV1. This improved the smoothness, reduced the artefacts in the higher latitudes and dramatically decreased the fraction of missing data. Emphasis was put here on the compositing step. Finally GEOV2 was adapted to PROBA-V at 300 m resolution, resulting in GEOV3 that copes with the fact that no climatology was available at this resolution (the climatology being used in GEOV2 to fill gaps and smooth the temporal profiles).

The derivation of a product at the decametric resolution requires either the availability of an extended database of observations (or products as for the GEOVx series) or the use of radiative transfer model to simulate the products values as done in the CYCLOPES products. Because of the non availability of worldwide representative set of observations of LAI, FAPAR and FCOVER, it was decided to base the principle of the decametric products on radiative transfer model simulations. This was the basis of an algorithm proposed for Sentinel-2 [*Baret et al.*, 2009; *Verger et al.*, 2011]. The proposed algorithm is an adaptation of the previous algorithms tuned for Sentinel-2. This algorithm will deliver instantaneous products corresponding to each Landsat8 observations. It is the first step of the algorithm that will exploit jointly Landsat-8 and PROBA-V at 300m resolution to deliver near-real time products at decametric resolution.

2.5. OUTLINE OF THE ALGORITHM

The scheme of the algorithm is similar as the one proposed to Sentinel2 products. For each product, one particular neural network is trained and used. It is composed of two steps (Figure 1):

- Training the neural network

- Operational use of the neural network
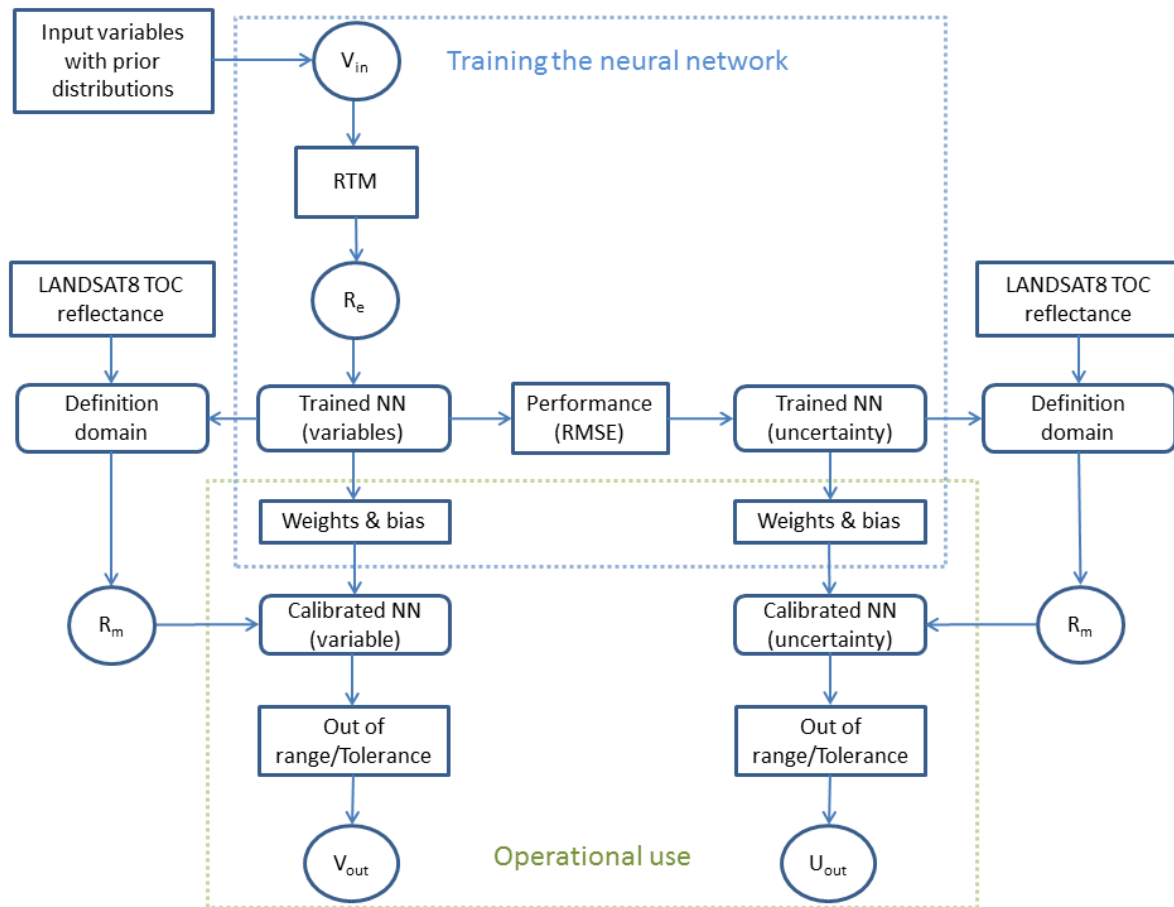
Figure 1: The flow chart of the main algorithm. V_{in} corresponds to the input variables for the model, R_e and R_m represent the estimated and measured top of canopy reflectance, respectively. V_{out} and U_{out} correspond to the output variable and associated uncertainty, respectively.

2.5.1.   Training the neural network

- **Generate the training database.** It is based on the prior information on the distribution of the input variables to the models. The generated database should represent the actual distribution of variables.
- **Designing network architecture.** It consists in defining the optimal structure and possible transformations of the inputs and outputs.
- **Calibrating the network.** This last step corresponds to the actual training, i.e. adjusting the coefficients (synaptic weights and bias) that provide the best estimates of the biophysical variables.

2.5.2. Operational use of the neural network

Three networks are generated to produce the following variables: LAI, FAPAR and FCOVER. Another three networks are trained to generate the uncertainties associated to the above variable estimates. Additional indicators are produced and are based on the following features:

- **Definition domain for the input data.** The actual Landsat-8 input reflectances should be consistent with those used in the simulated training data base. Therefore, when input reflectances 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.
- **Definition domain for the output variables.** This represents the output variables which should be within a predefined range determined by the simulated dataset and extended by given tolerance values.
- **Quality flag.** This labels the quality of the data, including the pixels contaminated by cloud, out of range or water and snow.

3. THE RETRIEVAL ALGORITHM

3.1. INPUT DATA

3.1.1. Top of Canopy reflectance

The top of canopy Landsat-8 reflectance data is generated based on the MACCS algorithm (Feuview ans Ruffel, 2013), and are thus corrected from atmospheric, adjacency and terrain effects. A strict cloud mask should be applied to the reflectance data to remove the pixels contaminated by clouds and clouds shadows. Reflectances should be expressed in terms of reflectance factor, mainly varying between 0 and 0.7 for most land surfaces outside the hot-spot or the specular directions and snow or ice cover. Four bands are used: B3 (green), B4 (red), B5 (NIR) and B6 (SWIR-1). These bands are selected to be consistent with PROBA-V spectral characteristics and the 300m-biophysical product algorithm [ImagineS_RP2.1_ATBD_LAI300m]. And the green band is used as it is highly correlated with the vegetation responses. Figure 2 shows the relative spectral response on each Landsat-8 band used in this study and Figure 3 shows the corresponding band of PROBA-V.
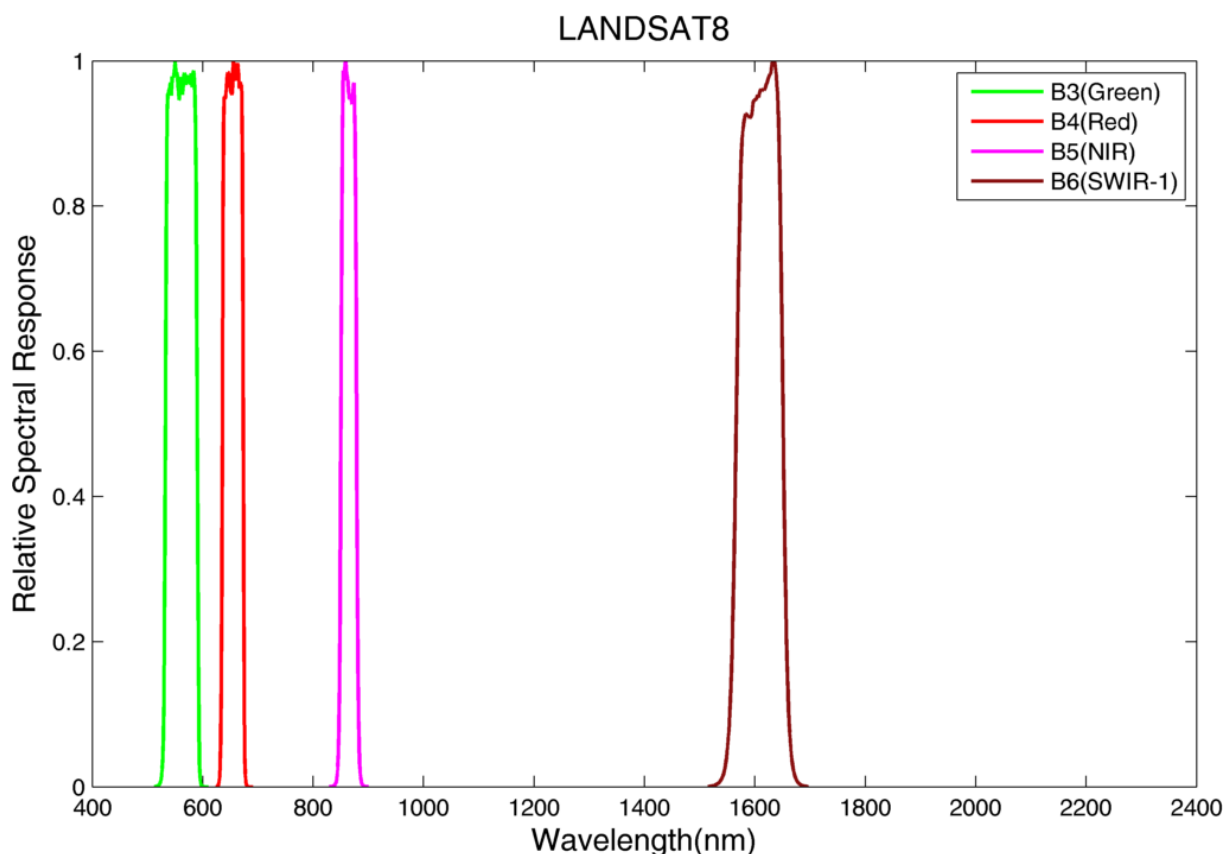
Figure 2. The relative spectral response of Landsat-8 bands used in this study.
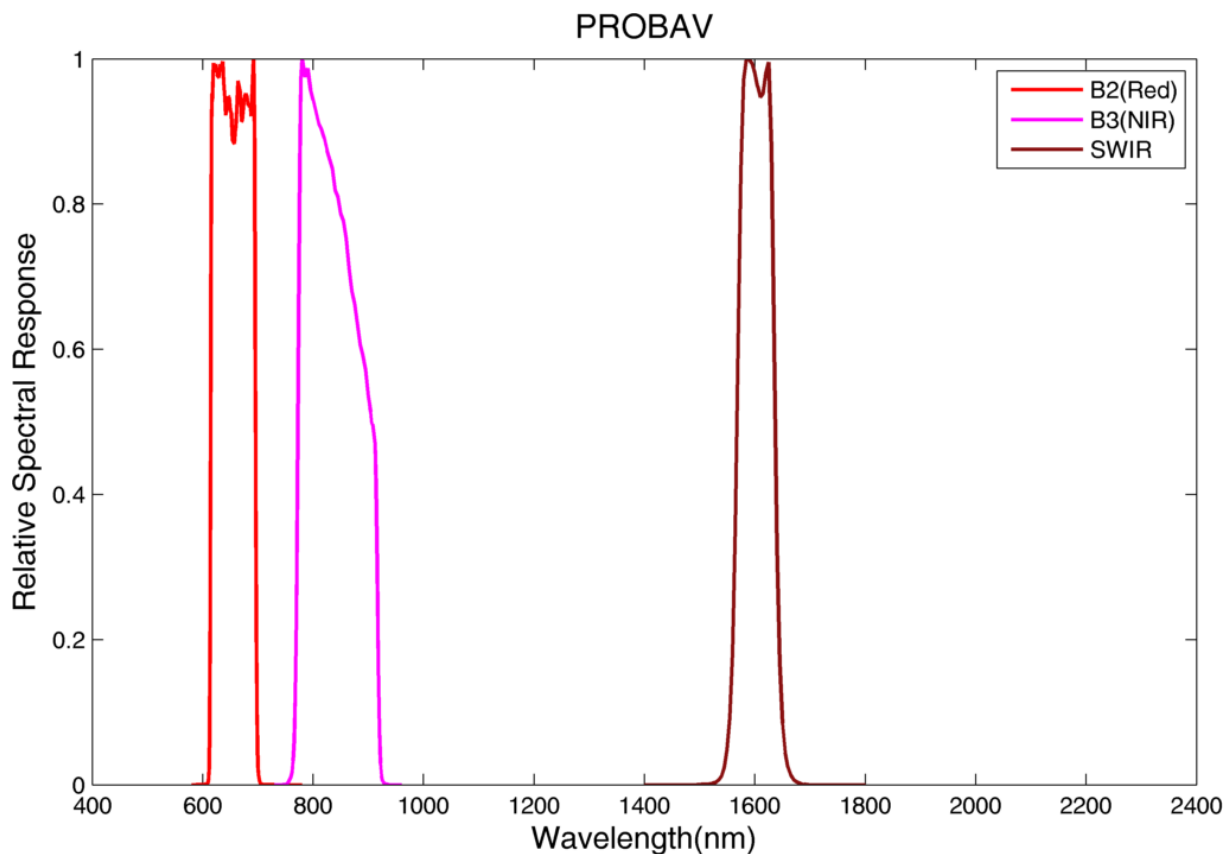
Figure 3. The relative spectral response of PROBA-V red, NIR and SWIR bands.

3.1.2.   Geometry of acquisition

The geometry information is required as input to the neural network for the three variables LAI, FAPAR and FCOVER. It includes:

- The sun zenith angle (SZA)
- The sun azimuth angle (SAA)

The viewing geometry is assumed independent from the date and location. The view zenith angle (VZA) is set to nadir, and the view azimuth angle (VAA) is null.

3.2. OUTPUTS

The output data includes seven layers:

- Three layers for LAI, FAPAR and FCOVER products
- Three layers for the uncertainty of each product
- One layer of qualitative quality indicators

3.2.1.   The products

The output products are provided by application of the algorithm over each pixel on each Landsat-8 measurements. They include the neural network derived LAI, FAPAR and FCOVER as described previously. The minimum and maximum values proposed are presented in Table 3. They are in agreement with what was proposed for the GEOV2 products [GIOGL1_ATBD_GEOV2] and GEOV3 products [ImagineS_RP2.1_ATBD].

Products	Physical Minimum	Physical Maximum	Max DN value	Scaling factor
LAI	0	7	210	30
FAPAR	0	0.94	235	250
FCOVER	0	1	250	250

Table 3. The minimum, maximum values and associated resolution for LAI, FAPAR and FCOVER products.

3.2.2.   Products uncertainty

The uncertainty values associated to each biophysical variable are also provided in separated layers. The range of variation of the uncertainties is presented in Table 4.

Products	Physical Minimum	Physical Maximum	Max DN value	Scaling factor
LAI_uncer	0	1.25	250	200
FAPAR_uncer	0	0.2	40	200
FCOVER_uncer	0	0.2	40	200

Table 4. The minimum, maximum values and scaling factors of the uncertainties for LAI, FAPAR and FCOVER.

3.2.3. Qualitative quality indicators

This layer provides information about the quality indicators (Table 5).

QC	Qualitative quality indicator	Meaning
1	Land / Water and snow	QC(1)=0: Water or snow (heritage from Landsat surface reflectance QC) QC(1)=1: Land
2	No Clouds / Clouds or clouds shadow	QC(2)=0: Contaminated by clouds or clouds shadow (heritage from Landsat surface reflectance clouds mask layer) QC(2)=1: No clouds or clouds shadow contamination
3	Input Valid / Input out of range	QC(3)=0: the input reflectance or geometry data are out of the definition domain due to cloud or cloud shadow contamination, poor atmospheric correction QC(3)=1: all input and output data are within the definition domain
4	Input Valid / Input out of range	QC(4)=0: the output variables are out of definition domain QC(4)=1: all input and output data are within the definition domain

Table 5. Qualitative quality indicators

3.3. DETAILED DESCRIPTIONS

3.3.1. Reflectance models

The algorithm is based on the neural network training for each product. The first step is to select pertinent models to generate the training data base. Physically based radiative transfer models need 3 main components that are described separately in the following.

3.3.1.1. Canopy radiative transfer models

The use of pure 3D models such as DART 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 or radiosity radiative transfer description

and applied to a representative sample of biomes and conditions would be ideal. However, it might be difficult to implement for three main reasons:

- The necessity to describe a very large range of realistic canopy architectures. This requires a huge effort in canopy architecture and optical properties measurements at the ground level. No such knowledge is currently available.
- The time associated to the model computation.
- The increase number of inputs required to get a more realistic description of canopy architecture should be balanced by prior information to regularize the ill posed and underdetermined radiative transfer model inversion process.

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] is widespread in the remote sensing community for the estimation of vegetation biophysical variables. It assumes the canopy as a turbid medium for which leaves are randomly distributed. Canopy structure is characterized by LAI, the average leaf angle (ALA) assuming an ellipsoidal distribution [*Campbell*, 1986], and hot spot parameter (HOT) [*Kuusk*, 1991].

3.3.1.2. Leaf optical model

The PROSPECT model [*Jacquemoud and Baret*, 1990] with the updated absorption coefficients proposed by [*Fourty and Baret*, 1997] is used as its versatility and good performances in simulating the leaf optical properties.

This model has been successfully validated over broadleaf types [*Fourty and Baret*, 1997; *Jacquemoud and Baret*, 1990; *Newnham and Burt*, 2001]. 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.*, 2004]. The variables required as input to the PROSPECT model are listed as following:

- N : leaf mesophyll structure index. It varies between 1.0 for the most compact leaves (such as young cereal leaves) up to 3.5 for thick leaves with well-developed spongy mesophyll or even senescent leaves having disorganized mesophyll with large amount of air spaces.
- C_{ab} : Leaf chlorophyll content ($\mu g.cm^{-2}$). 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 varies between 0 and 100 $\mu g.cm^{-2}$, although a threshold value of 15 $\mu g.cm^{-2}$ has been proposed to consider a leaf as 'green'.
- C_{dm} : Leaf dry matter content ($g.cm^{-2}$). Dry matter absorbs 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 used by physiologists. C_{dm} typically varies from 0.002 up to 0.02 g.cm^{-2}.

- 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). The chlorophyll and brown pigment content are studies to be exclusive, i.e. green and non-green elements (senescent leaves, branches, stems) are spatially dissociated. Green leaves will have no brown pigments and senescent leaves will have no chlorophyll pigments. C_{bp} value typically varies from 0 for green leaves, up to 3.5 for the senescent dark brown leaves.

3.3.1.3. 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 elements 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 and will be included within the green vegetation layer. The background reflectance may thus correspond to soil, litter, water and snow. However, as the water and snow are excluded from this study, they will not be involved in the products.

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:

$$\rho_b(\lambda,\Omega_i,H_j,Z_k) = Bs \cdot \rho_b(\lambda,\Omega_l,H_m,Z_n) \qquad (1)$$

where *Bs* is the brightness parameter that does not depend on wavelength λ, but depends on all the other factors (Ω,H,Z).

The brightness concept allows describing the spectral variation of a given background when the brightness parameter which is independent on wavelength and a reference soil reflectance spectrum are provided. This concept has been successfully used in previous studies [*Bacour et al.*, 2002].

3.3.2. Generation of the training database

The quality of the training database directly influences the accuracy of the generated products. Two steps need to be completed:

- Generate the input variables database considering the distribution of each variable
- Simulate the TOC reflectance on Landsat-8 bands considering the uncertainties

3.3.2.1. Generate the input variables database

To better simulate the top of canopy reflectance close to the actual one, the distribution of the input model variables need to be described with the aim to achieve a good representativeness of the actual global distributions. The variable distributions are derived from available information of previous studies. Moreover, although no much information is available on possible co-distribution between variables, 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 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 4 and Eqs. (2) – (4). LAI_{max} is the maximum LAI value considered.
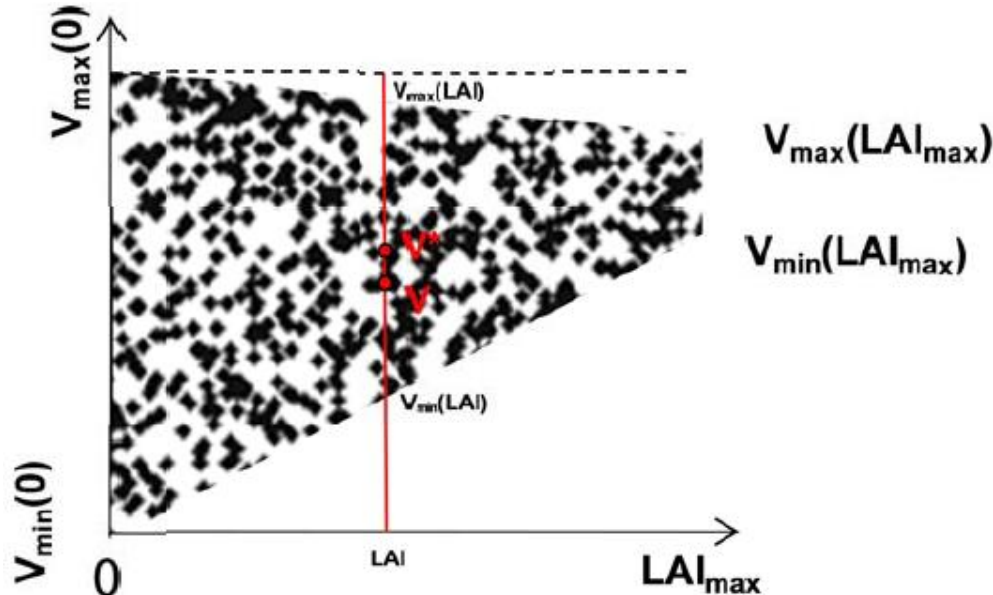

Figure 4. Scheme explaining the way the distribution of variable V is linked to that of LAI.

$$(V - V_{min}(0)) / (V_{max}(0) - V_{min}(0)) = (V^* - V_{min}(LAI)) / (V_{max}(LAI) - V_{min}(LAI)) \quad (2)$$

$$V_{min}(LAI) = V_{min}(0) + (LAI_{max} - LAI) * (V_{min}(LAI_{max}) - V_{min}(0)) \quad (3)$$

$$V_{max}(LAI) = V_{max}(0) + (LAI_{max} - LAI)* (V_{max}(LAI_{max}) - V_{max}(0)) \quad (4)$$

where V* is the value of variable V after linking its distribution to that of LAI. Table 6 lists the range of variation and the actual distribution used for the input variables of the canopy model, leaf optical model and background reflectance model and Table 7 presents the co-distribution of each variable with LAI. Figure 5 shows the distribution and co-distribution of each variable.

- **LAI**: The distribution of LAI values follow a log-normal pattern characterized by a mode of 2 and a large standard deviation of 2. This allows sampling extremely low and high LAI values. The higher LAI values were truncated at LAI of 15. The value was determined to improve the saturation problem by introducing a significant amount of very high LAI values in the training process.
- **ALA**: The average leaf inclination angle is assumed to follow a truncated Gaussian distribution centered 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.
- **HOT**: The hot spot parameter follows a truncated Gaussian distribution.
- **Leaf optical properties**: As little knowledge is available on the actual leaf characteristics, truncated Gaussian distributions were used for all these variables.
- **Brightness**: The *Bs* coefficient is randomly drawn according to a truncated Gaussian distribution centered on *Bs* = 1.0. 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.

	Variable	Minimum	Maximum	Mode	Std	Nb_Class
Canopy	LAI	0.0	15.0	2.0	2.0	6
	ALA (°)	30	80	60	20	3
	HOT	0.1	0.5	0.2	0.5	1
Leaf	N	1.20	1.80	1.50	0.30	3
	C_{ab} (µg.m-2)	20	90	45	30	4
	C_{dm} (g.m-2)	0.003	0.011	0.005	0.005	4
	Cw_Rel	0.60	0.85	0.75	0.08	4
	C_{bp}	0.00	2.00	0.00	0.30	3
Background	Bs	0.50	3.50	1.20	2.00	4

Table 6. Distribution of the input variables of the radiative transfer model used to generate the training database. Truncated Gaussian, log-normal and uniform distribution laws are used, characterized by the mode, standard deviation (std), minimum and maximum values. The number of classes for each variable is shown (Nb_class). The viewing conditions are set to nadir.

	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	No	1.2	2.2	1.30	1.8
	C_{ab}	Yes	20	90	45	90
	C_{dm}	Yes	0.003	0.011	0.0050	0.0110
	C_{w_Rel}	Yes	0.6	0.85	0.70	0.80
	C_{bp}	Yes	0.00	2.00	0.00	0.20
Background	Bs	No	0.5	3.50	0.50	1.20

Table 7. Co-distribution of input variables with LAI.
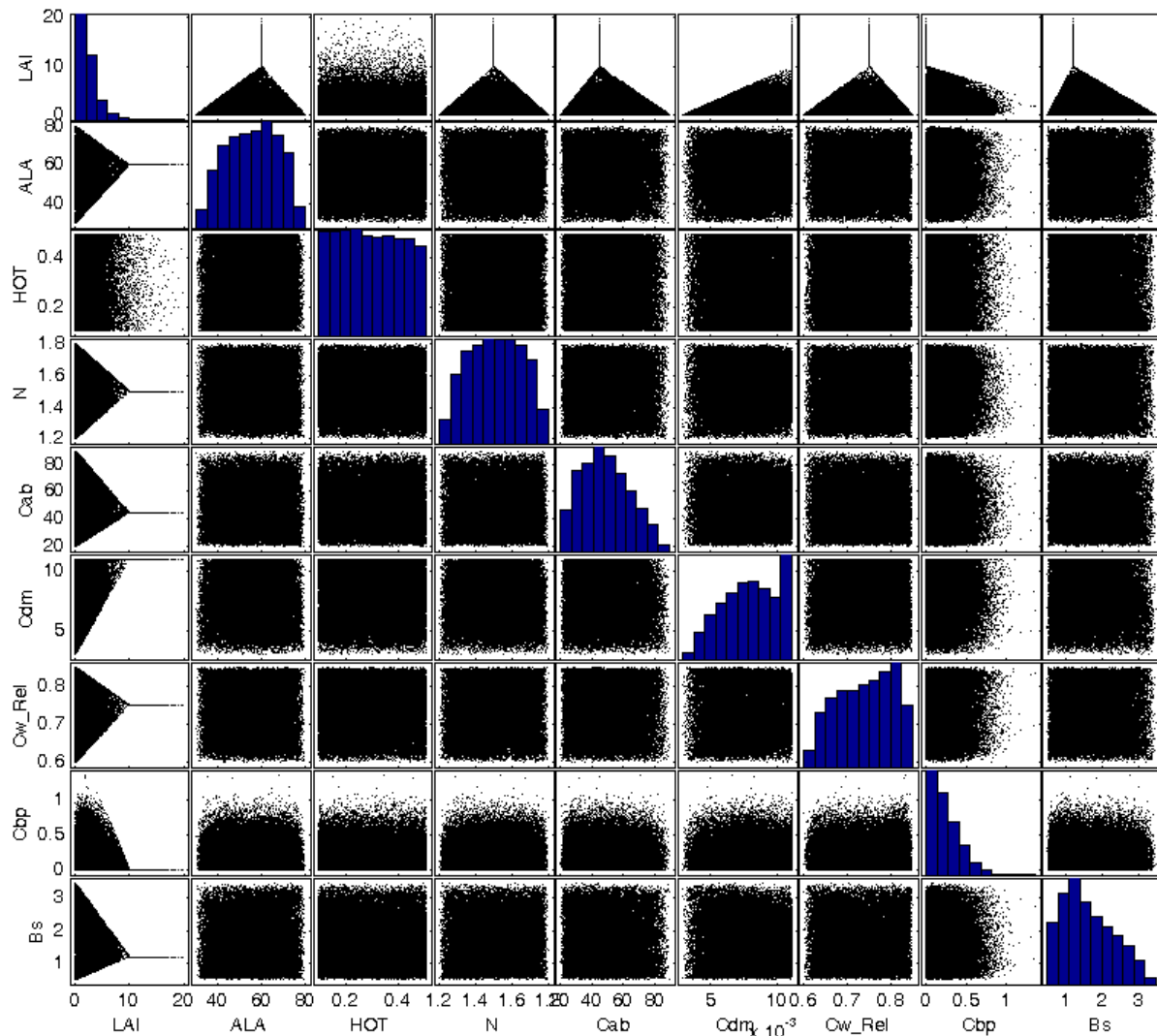

Figure 5. Distribution and co-distributions of the input canopy, leaf and background variables.

The sampling scheme is based on a full orthogonal experimental plan [*Bacour et al.*, 2002]. This consists to identify classes of values for each variable. Then all the combinations of classes are sampled once. Finally the actual values of each variable are randomly drawn within the range of variation defined by the corresponding class, according to the distribution law specified for the variable considered. This process allows accounting for all the interactions, while having the range of variation for each variable densely and quasi-randomly populated.

At the end of the simulation process, a total number of 41472 cases were simulated. The size of the training database is decided based on the complexity of the problem. Previous studies [*Combal et al.*, 2002] have shown that for a medium complexity problem, a training dataset close to 10000 cases was satisfactory. In the current case that corresponds to a more complex algorithm, the size of the training data base should be increased. This data set will then be split in two parts with a random selection process:

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

3.3.2.2. Simulating the TOC reflectance on each Landsat-8 band

From the input variables presented above, the top of canopy reflectance on each Landsat-8 band was simulated using the PROSPECT+SAIL model.

An uncertainty model was used to describe actual Landsat-8 characteristics as well as the ability of the radiative transfer model used to represent actual reflectances. They are described in Table 8. The uncertainties are computed according to:

$$R^*(\lambda) = R(\lambda)\ (1+(MD(\lambda)+MI)/100)+AD(\lambda)+AI \qquad (5)$$

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

Bands	Landsat-8			
	AD	AI	MD (%)	MI (%)
Band 3	0.01	0.01	2	2
Band 4	0.01	0.01	2	2
Band 5	0.01	0.01	2	2
Band 6	0.01	0.01	2	2

Table 8. Characteristics of the uncertainties model used.

Figure 6 shows the distribution and co-distribution of simulated top of canopy reflectance on each Landsat-8 band. The B3 (green) and B4 (red) and B6 (SWIR) bands are highly correlated as expected. Conversely, B5 (near infrared) does not show strong correlations with the three other bands.
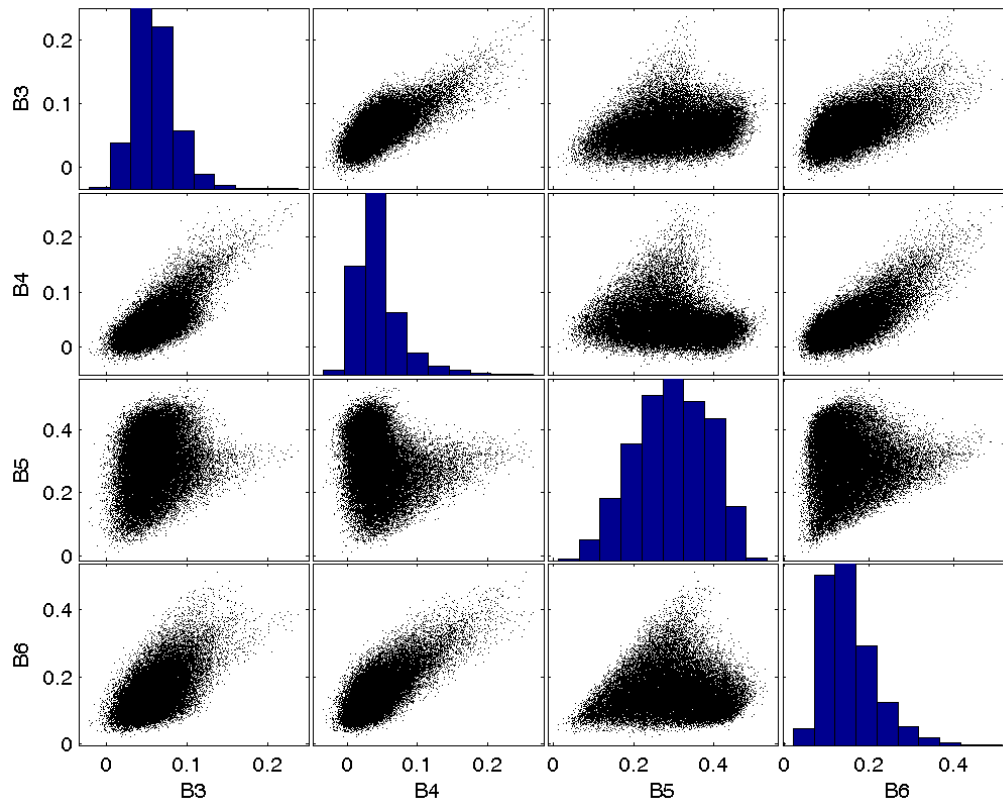

Figure 6. The distribution and co-distribution of the simulated top of canopy reflectance on each Landsat-8 band.

3.3.3.   Training the neural network

Neural network are widely used in the retrieval of satellite biophysical products [*Baret et al*., 2007]. In this study, a dedicated back-propagation artificial neural network [*Rummelhart et al.,* 1986] was trained. The structure of the network is shown in Figure 7. It includes three layers:

- one input layer composed of the normalized input data
- one hidden layer composed of 5 neurons with tangent sigmoid transfer functions.
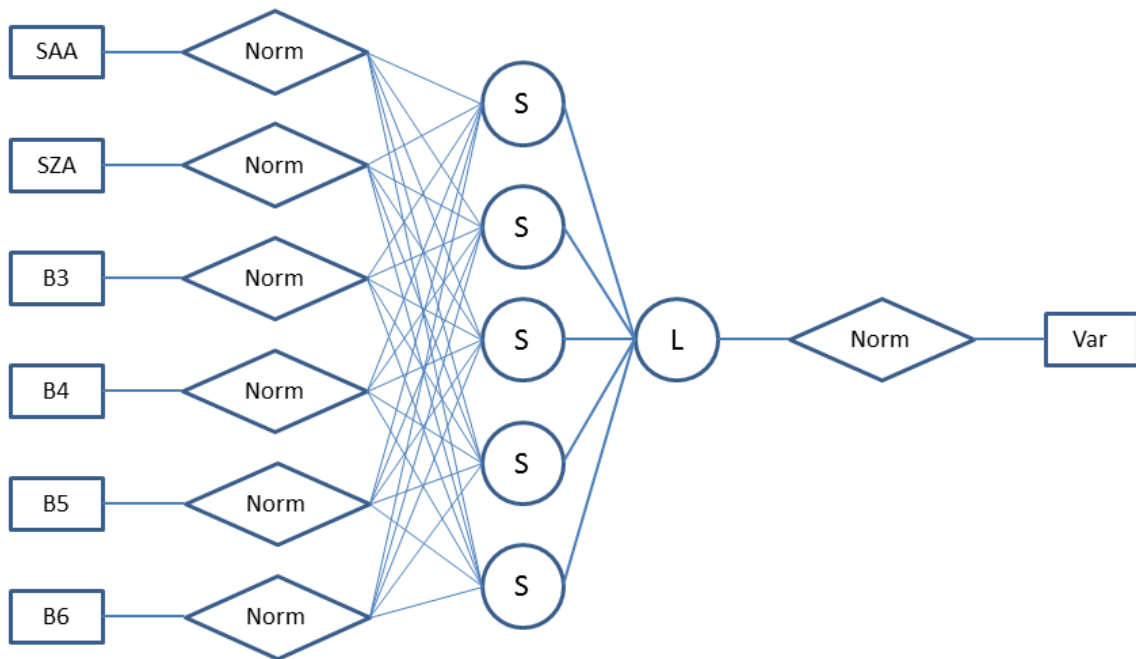- one output layer with a linear transfer function.

Figure 7: Structure of the neural network. SZA represent the solar zenith angle. 'Norm' represents the normalization process of the input and output variables. Neurons 'S' and 'L' correspond to tangent-sigmoid and linear transfer functions, respectively.

The inputs layer of the network includes the simulated surface reflectance for each band and the geometrical configurations, and the outputs are LAI, FAPAR and FCOVER. The inputs and output variables are firstly normalized to increase the performances of convergence of the training algorithm (Eq. (6)).

$$X^* = 2 * (X-X_{min})/(X_{max} - X_{min}) - 1 \qquad (6)$$

where X^* is normalized values, X is the input values to the neural network, X_{max} and X_{min} are the maximum and minimum of the inputs values.

The neural network is described by its architecture, i.e. 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:

$$y = Tansig (x) = 2/(1+exp(-2x))-1 \qquad (7)$$

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

For each variable, five networks were trained in parallel with different initial solutions, and the best one is selected based on the smallest RMSE between the outputs and the 'true' biophysical variables in the validation data set.

3.3.4. Definition domain and outliers remove

The definition domain is defined as the range of the input and output variables. The definition domain for the reflectance data, input geometry and output variables are defined as following:

- **Reflectance**: $(\rho_i^m, \rho_j^m) \in D(\rho_i^s, \rho_j^s)$, where ρ^m is the actual TOC reflectance, and ρ^s is the simulated TOC reflectance from the training dataset. D represents the definition domain determined by the co-distribution of simulated reflectance on two bands, i and j. Figure 8 shows the definition domain region of simulated reflectance on Landsat-8 band combinations.
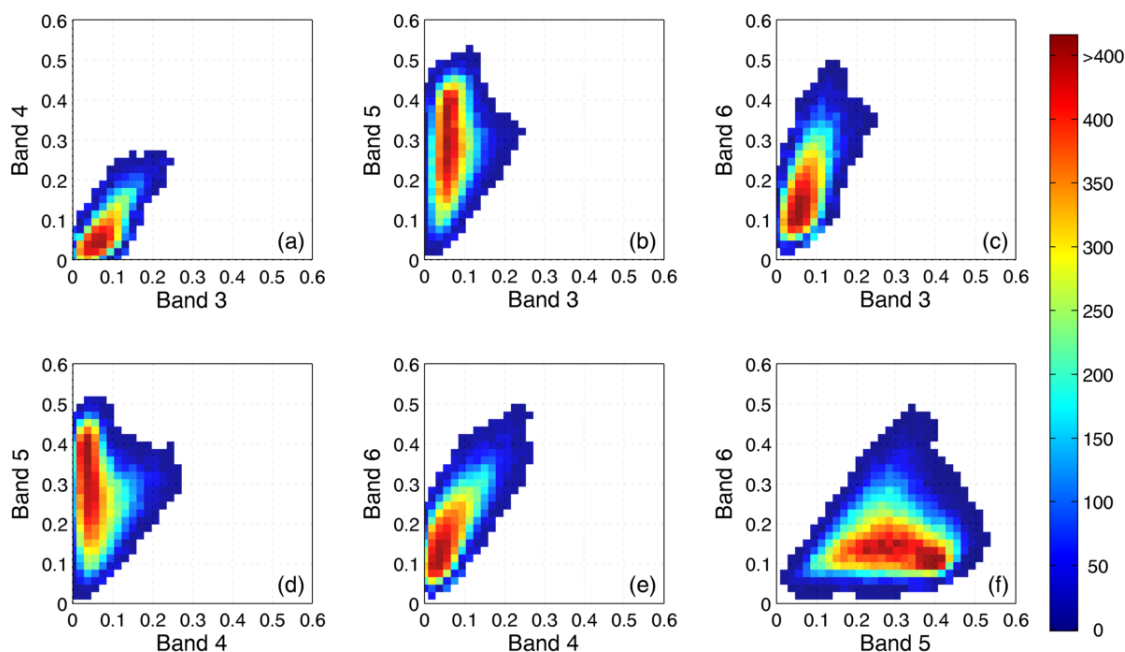

Figure 8: Definition domain region (colorful region) of simulated TOC reflectance on Landsat-8 bands.

- **Solar angles**: SZA < 60˚, no restrict on SAA.
- **Output variables**: X_{min} – Tol ≤ X ≤ X_{max} + Tol, where Xmin and Xmax are predefined minimum and maximum value of each variable, Tol corresponds to the tolerance of the output variable. Table 9 shows the X_{min}, X_{max} and Tol for each variable.

	X_{min}	X_{max}	Tol
LAI	0	7	0.2
FAPAR	0	0.94	0.05
FCOVER	0	1	0.05

Table 9: The definition domain for each output variable. X_{min} and X_{max} represent the minimum and maximum value, and Tol corresponds to the tolerance.

In the actual inversion, pixels with input or output values out of the definition domain are labeled as 'out of range'.

3.3.5. Product uncertainties

Based on the training dataset and validation dataset, the theoretical performances of the algorithm can be calculated from the RMSE between the estimated and actual biophysical values. A specific neural network is trained for each product to relate the estimated uncertainties to the input reflectance and observation geometry values. Figure 9 shows the theoretical performances of each variable. It shows that all estimated products are well correlated to the validation datasets (R>0.88). The performance improves from LAI (RMSE=0.83), FAPAR (RMSE=0.07) and FCOVER (RMSE=0.05).

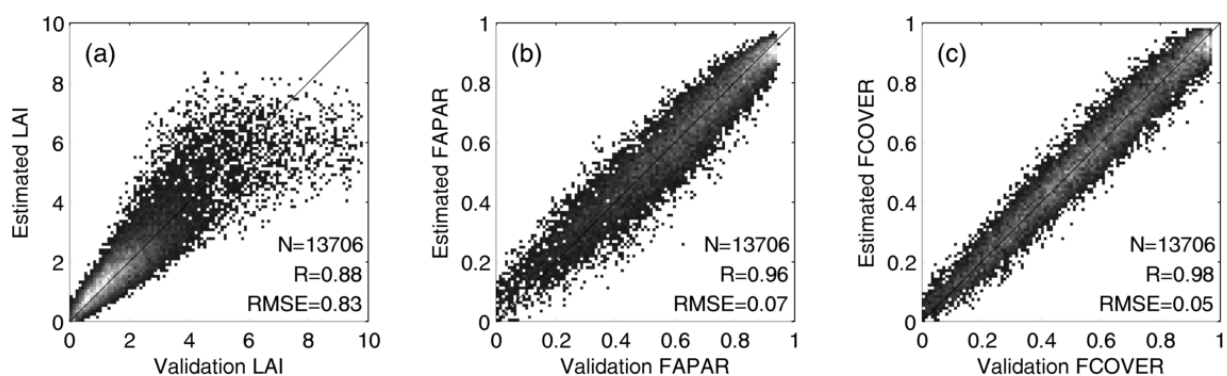**Figure 9: The theoretical performances of (a) LAI, (b) FAPAR, (c) FCOVER evaluated over the validation dataset.**

3.4. LIMITATIONS

The derived Landsat-8 LAI is the effective LAI because the foliage clumping is not accounted. This may lead to an underestimation of the foliage amount for the aggregated vegetation such as forest. This underestimation will be severely for highly clumped coniferous forests and evergreen forests. Further work will be done to consider the clumping effects on this decametric scale.

4. PRELIMINARY RESULTS

At the time of writing, we have satellite data associated with field data over only one site. The analysis will be completed when more sites will be available.

4.1. STUDY AREA

A study area is selected to perform the algorithm. This region is located in the South West (SouthWest) of France (43.52˚ N, 1.18˚ E). Figure 10 shows the location and land cover types in the region.
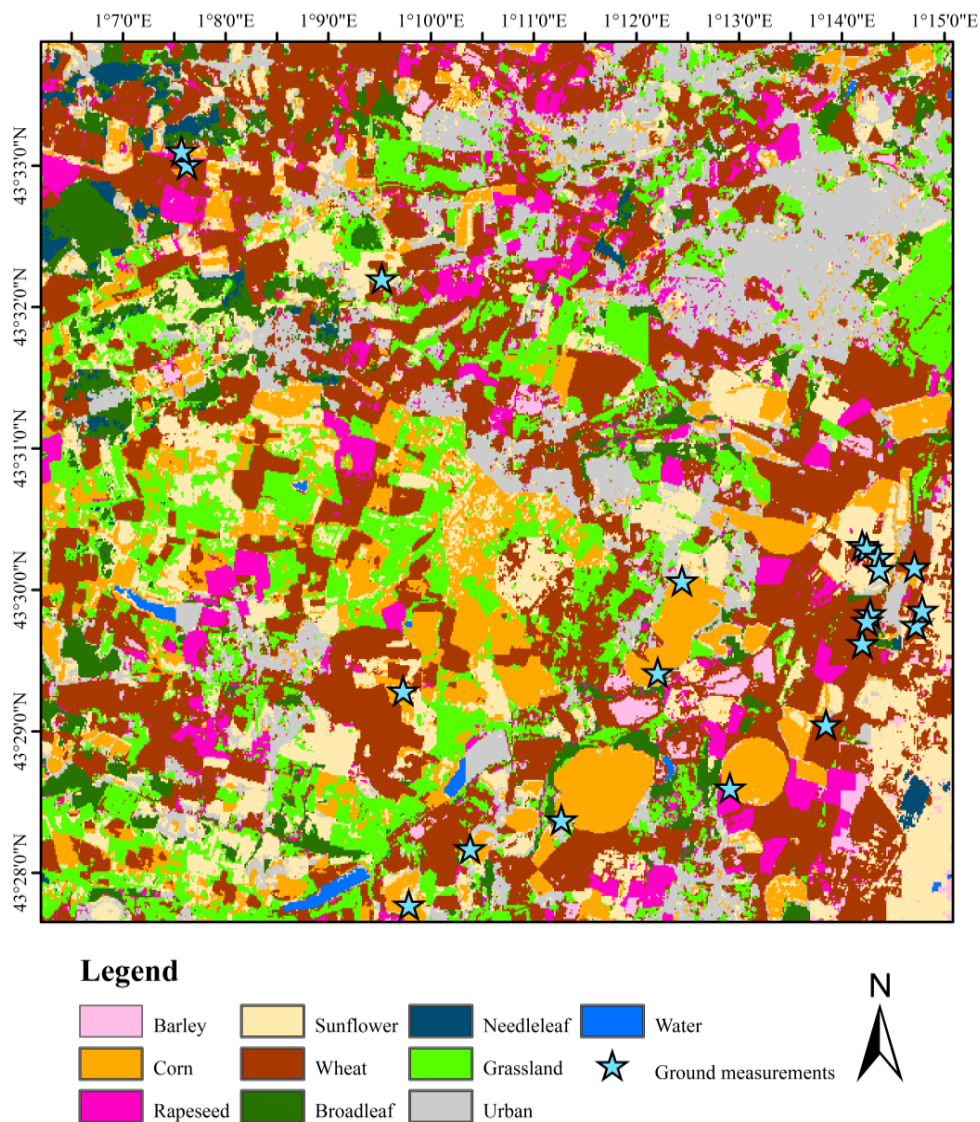

Figure 10: The land cover types of the SouthWest site in a Landsat-8 image (April 14th, 2013).

LAI, FAPAR and FCOVER are generated based on the earlier described algorithm from the top of canopy reflectance data from Landsat-8 in 2013 as derived from the MACCS processing chain.

4.2. SPATIAL CONSISTENCY

Figures 11 to 13 show the spatial variation map of three variables in 2013.
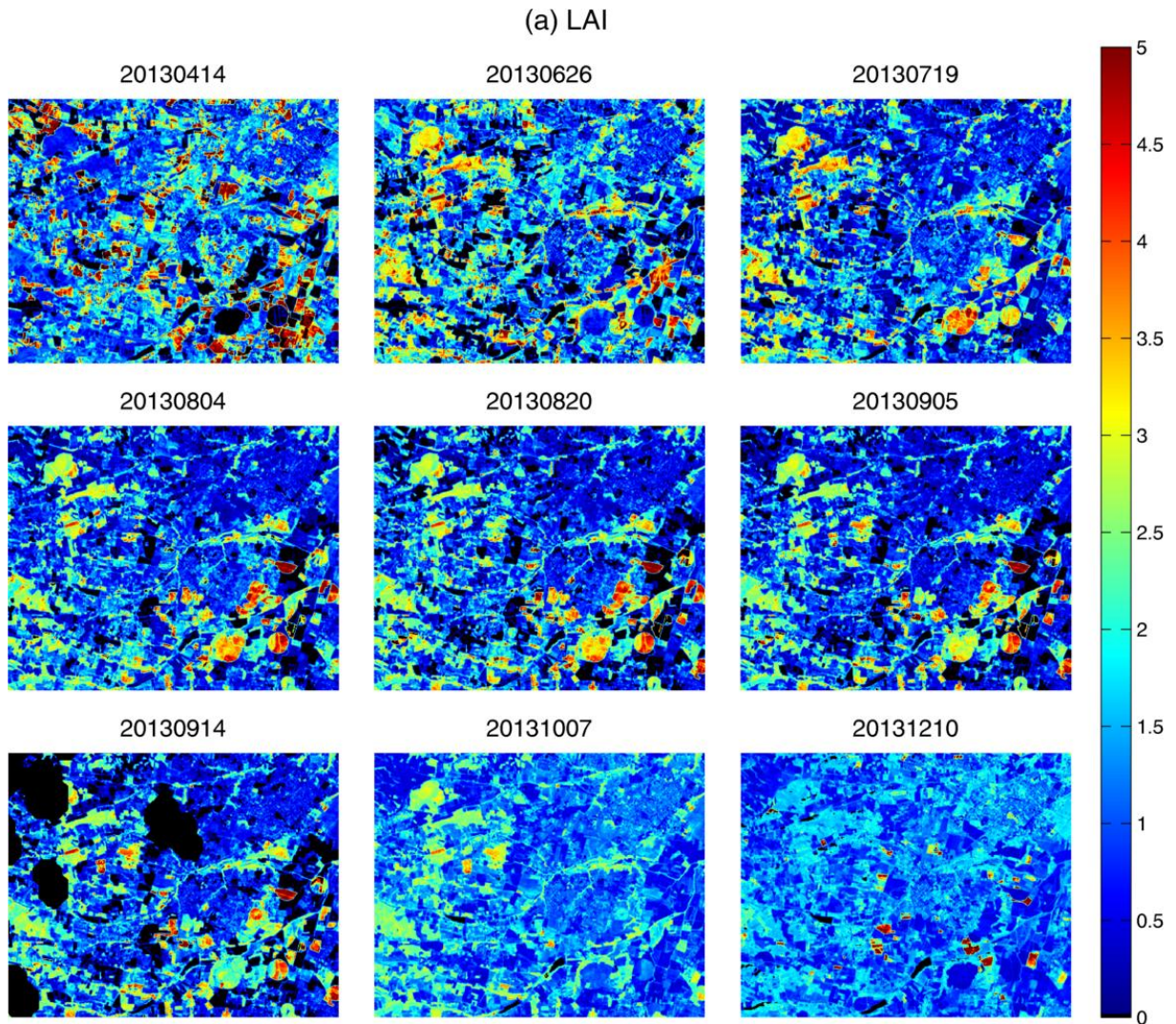

Figure 11: The spatial variation of Landsat-8 LAI over SouthWest site in 2013. Black regions represent the pixels contaminated by clouds or clouds shadow, or out of the definition domain or the output range.
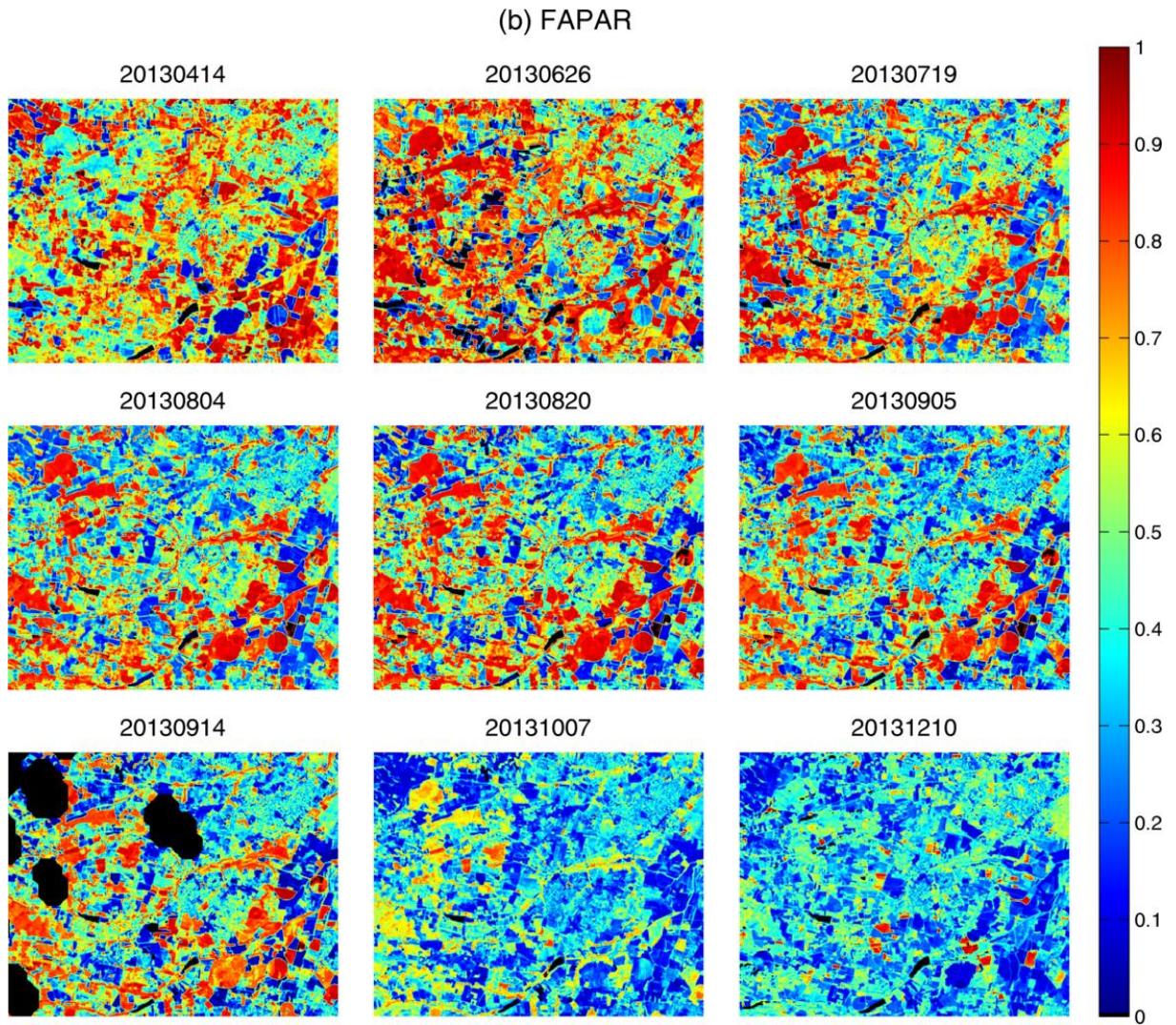
(b) FAPAR

Figure 12: The spatial variation of Landsat-8 FAPAR over SouthWest site in 2013.
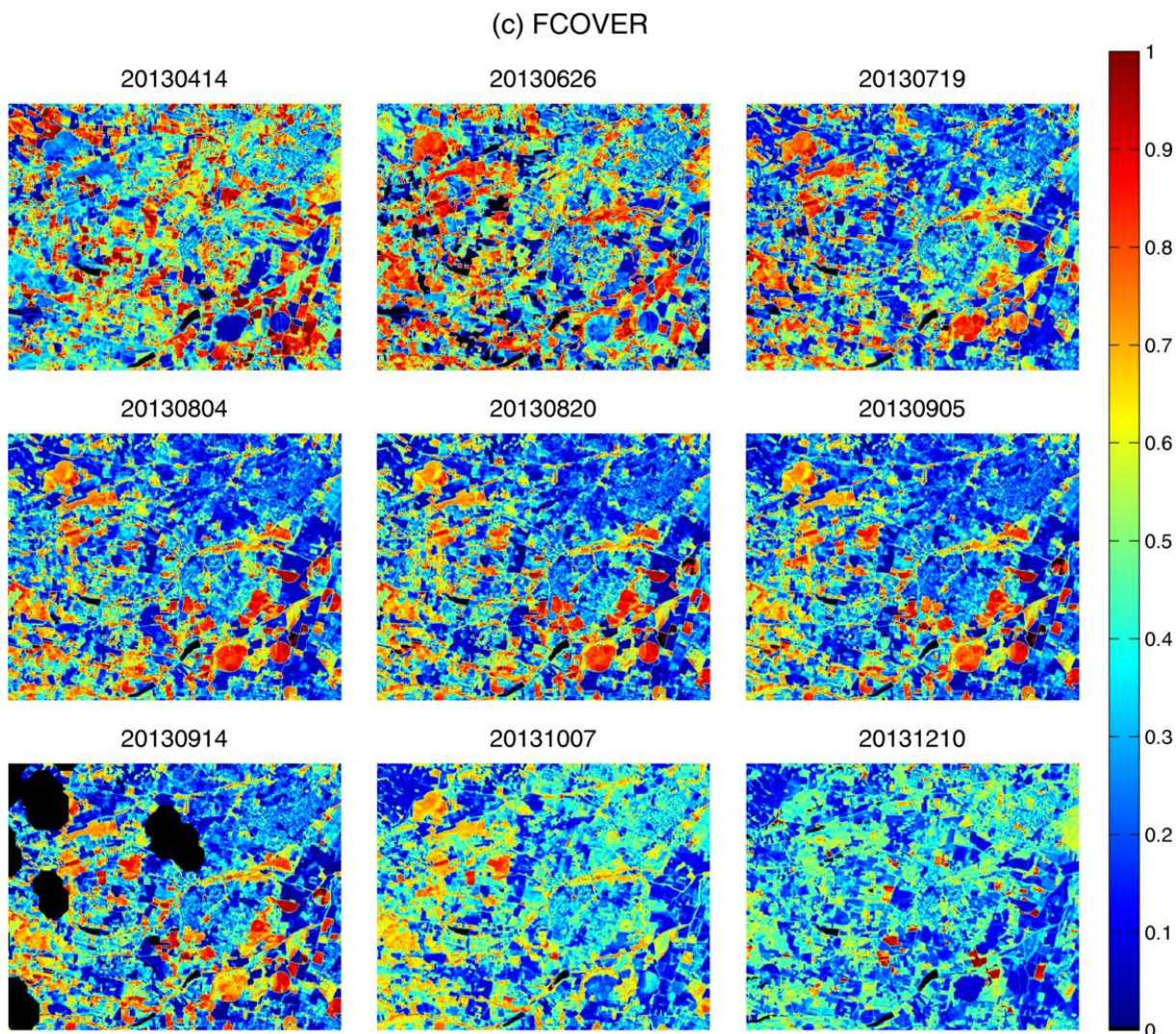
(c) FCOVER

Figure 13: The spatial variation of Landsat-8 FCOVER over SouthWest site in 2013.

4.3. TEMPORAL CONSISTENCY

To generate a whole year complete seasonal profile, the biophysical variables were also derived from the SPOT4/Take5 images (http://spirit.cnes.fr/take5) based on the same algorithm described above. The generated SPOT4 variables were resampled into the same spatial resolution as Landsat-8. The profile was generated from the mean value of all pixels in this study area. Figures 14 to 16 show the generated profiles for LAI, FAPAR and FCOVER. Generally, the derived variables on each biome exhibit a clear seasonal variation.
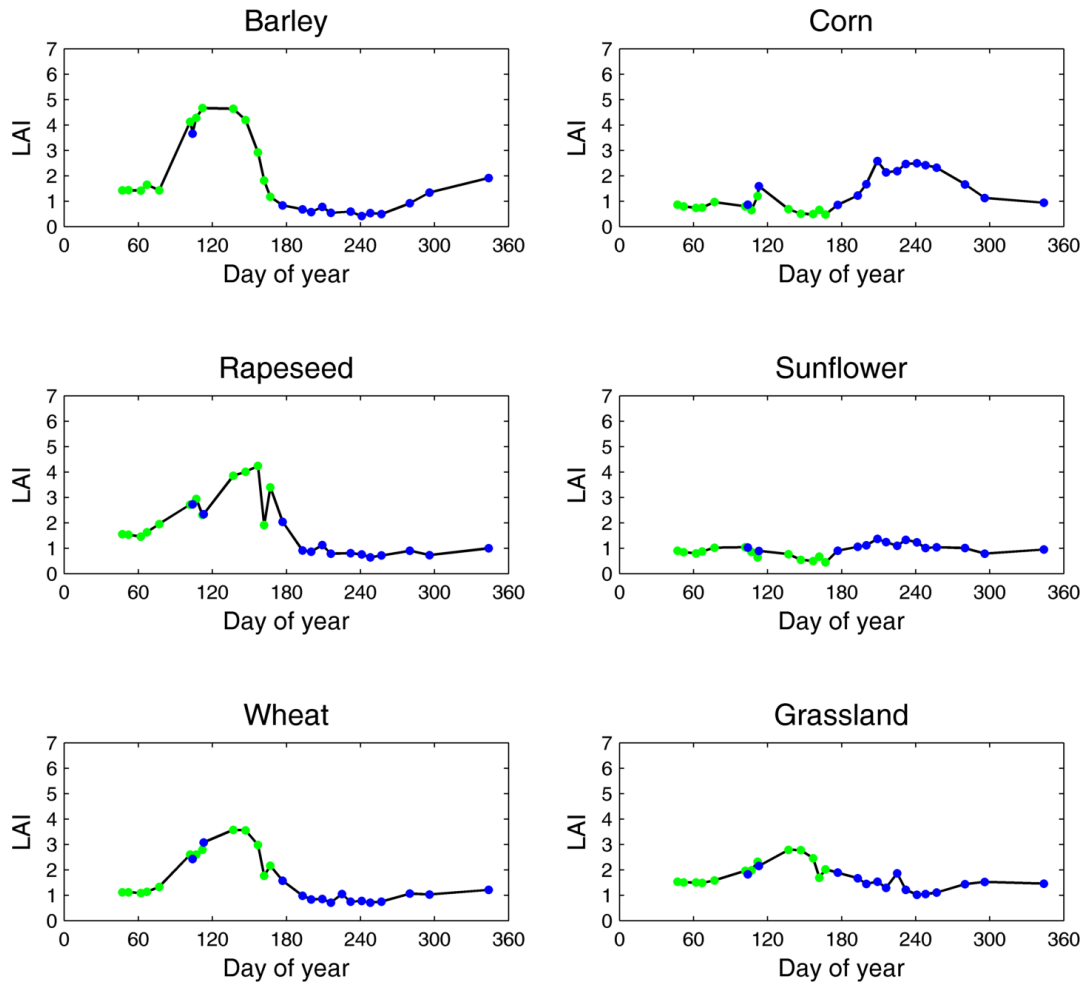
Figure 14: The mean temporal variation of the LAI estimated from Landsat-8 and SPOT4, 2013. The green and blue dots represent the estimations from SPOT4 and Landsat-8, respectively.

Figure 15: The temporal variation of the FAPAR estimated from Landsat-8 and SPOT4, 2013. The green and blue dots represent the estimations from SPOT4 and Landsat-8, respectively.
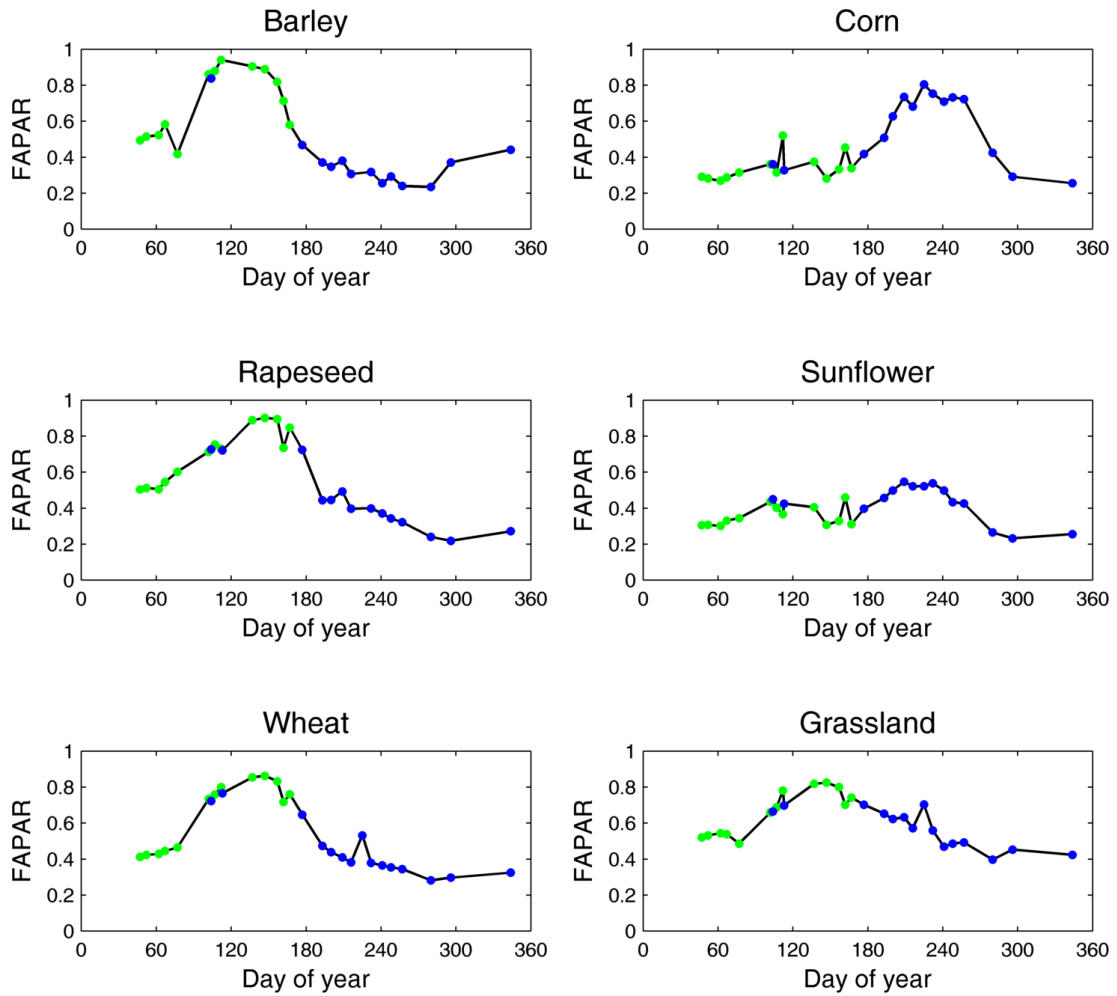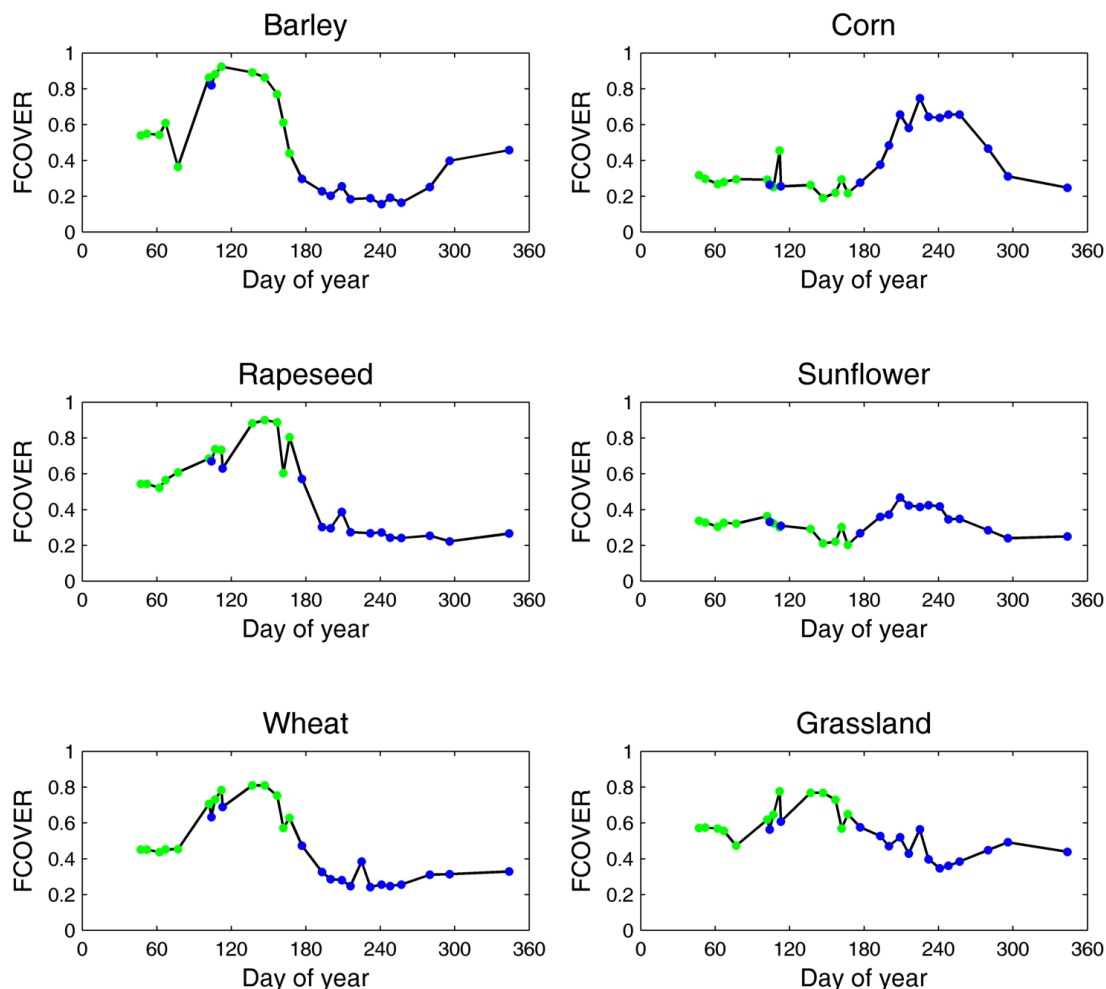
Figure 16: The temporal variation of the FCOVER estimated from Landsat-8 and SPOT4, 2013. The green and blue dots represent the estimations from SPOT4 and Landsat-8, respectively.

4.4. COMPARISON WITH GROUND MEASUREMENTS

Ground measurements data over three main land cover types (wheat, maize and sunflower) were collected in 2013 using the downward-looking digital hemispherical cameras. The protocol of the measurements was based on the one proposed in VALERI [*Baret et al.*, 2003]. LAI, black-sky FAPAR and FCOVER were calculated from these photos using CAN_EYE [*Weiss and Baret*, 2010]. Results show that the derived products are correlated well with the field measurements (R>0.84), although with a slight bias of 0.3 for LAI, 0.05 for FAPAR and 0.09 for FCOVER (Figure 17). The accuracy of LAI and FAPAR products are within the absolute uncertainty requirements (LAI: ±0.5, FAPAR: ±0.05) set by the GCOS [*GCOS,* 2011].
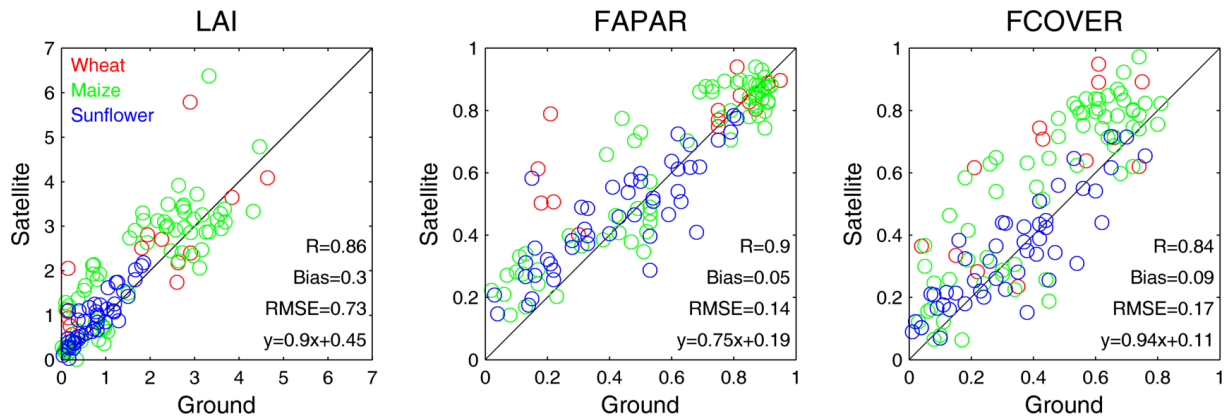
Figure 17: Validation of Landsat-8 LAI, FAPAR and FCOVER with field measured values over three biomes in the study area.

5. CONCLUSIONS

This ATBD provides an algorithm to generate biophysical variables from Landsat-8 images. It applies to all vegetation types and no prior knowledge of land cover types is used. Note that the radiative transfer model embedded in the algorithm assumes the canopy as homogeneous. Therefore, there is no consideration of clumping index and only effective LAI is derived. The validation of this algorithm will be performed on sites with field measurements. The algorithm will be fine-tuned according to these validation results.

6. RISK OF FAILURE AND MITIGATION MEASURES

The use of this algorithm requires the production of top of canopy reflectance data from Landsat-8. In this study, the top of canopy reflectance data from THEIA (http://www.theia-land.fr/en) are used. Other sources, such as the USGS Landsat Surface Reflectance Climate Data Record website (http://landsat.usgs.gov/CDR_LSR.php) also provide global Landsat-8 top of canopy reflectance data. The users could downloaded the reflectance data from these website or generate the data by performing atmospheric correction themselves.

7. REFERENCES

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

Baret, F., M. Weiss, R. Lacaze, F. Camacho, H. Makhmara, P. Pacholcyzk, and B. Smets (2013), GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production, Remote Sensing of Environment, 137(0), 299-309.

Baret, F., M. Weiss, P. Bicheron, and B. Bethelot (2009), S2PAD - Sentinel-2 MSI - Level 2B Products Algorithm Theoretical Basis DocumentRep., Vega, GMBH, .

Baret, F., et al. (2007), LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION Part 1: Principles of the algorithm, *Remote Sensing of Environment*, *110*, 275-286.

Baret, F., C. Bacour, M. Weiss, K. Pavageau, D. Béal, V. Bruniquel, P. Regner, J. Moreno, C. Gonzalez, and J. Chen (2004), Canopy biophysical variables estimation from MERIS observations based on neural networks and radiative transfer modelling: principles and validation, in *ENVISAT workshop*, edited, Salzburg (Austria).

Baret, F., M. Weiss, D. Allard, S. Garrigue, M. Leroy, and H. Jeanjean (2003), VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products, *Remote Sensing of Environment(submitted)*.

Camacho, F., J. Cernicharo, R. Lacaze, F. Baret, and M. Weiss (2013), GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products, Remote Sensing of Environment, 137(0), 310-329.

Campbell, G. S. (1986), Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution, *Agricultural and Forest Meteorology*, *36*, 317-321.

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

Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Mace, A. Pragnere, R. Myneni, Y. Knyazikhin, and L. Wang (2002), Retrieval of canopy biophysical variables from bidirectional reflectance Using prior information to solve the ill-posed inverse problem, *Remote Sensing of Environment*, *84*, 1-15.

Feuvrier, T., and C. Ruffel (2013), MACCS - Multi Mission Level 2 & Level 3 Images Production Software, Proceedings of the European Space Agency Living Planet symposium, Edinburgh.

Fourty, T., and F. Baret (1997), Amelioration de la precision des coefficients d'absorption specifique de la matiere seche et des pigments photosynthetiques, INRA, Avignon.

Garrigues, S., et al. (2008), Validation and Intercomparison of Global Leaf Area Index Products Derived From Remote Sensing Data, *Journal of Geophysical Research*, 113(G02028).

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

GCOS (2011), Systematic Observation Requirements for Satellite-Based Products for Climate, 2011 615 Update, Supplemental Details to the Satellite-Based Component of the Implementation Plan for the 616 Global Observing System for Climate in Support of the UNFCCC (2010 Update), Reference 617 Number GCOS-154, pp. 138, http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf.

Jacquemoud, S., and F. Baret (1990), PROSPECT: A Model of Leaf Optical Properties Spectra, *Remote Sensing of Environment*, *34*, 75-91.

Kuusk, A. (1991), Determination of Vegetation Canopy Parameters from Optical Measurements, *Remote Sensing of Environment*, *37*, 207-218.

McCallum, I., W. Wagner, C. Schmullius, A. Shvidenko, M. Obersteiner, S. Fritz, and S. Nilsson (2009), Satellite-based terrestrial production efficiency modeling, *Carbon Balance and Management*, *4*.

Newnham, G. J., and T. Burt (2001), Validation of a leaf reflectance and transmittance model for three agricultural crop species, *IEEE Transactions of Geoscience and Remote Sensing*, *7*, 2976-2978.

Prince, S. D. (1991), Amodel of regional primary production for use with course resolution satellite data, *International Journal of Remote Sensing*, *12*, 1313-1330.

Rummelhart, D. E., G. E. Hinton, and R. J. Williams (1986), Learning internal representations by error propagation, in *Parallel data processing*, edited, pp. 318-362, Cambridge, MA, USA M.I.T.

Verger, A., F. Baret, and M. Weiss (2014), Near Real-Time Vegetation Monitoring at Global Scale, JSTARS , IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 3473-3481.

Verger, A., Baret, F., Weiss, M. (2011), A multisensor fusion approach to improve LAI time series, *Remote Sensing of Environment*.

Verhoef, W. (1984), Light Scattering by Leaf Layers with Application to Canopy Reflectance Modeling: The SAIL Model, *Remote Sensing of Environment*, *16*, 125-141.

Weiss, M., and F. Baret (2010), CAN-EYE V6.1 user manual, *Instrument Manual Rep.*

Weiss, M., F. Baret , S. Garrigues, R. Lacaze, and P. Bicheron (2007), LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. part 2: Validation and comparison with MODIS Collection 4 products., Remote sensing of Environment, 110, 317-331.

Weiss, M., F. Baret, R. B. Myneni, A. Pragnère, and Y. Knyazikhin (2000), Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data, *Agronomie*, *20*, 3-22.

Zarco-Tejada, P. J., J. R. Miller, J. Harron, B. Hu, T. L. Noland, N. Goel, G. H. Mohammed, and P. Sampson (2004), Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies, *Remote Sensing of Environment*, *89*, 189-199.