Algorithm Theoretical Basis Document for Vegetation parameters (VEGA)

PRODUCTS: LSA-421 (MDFVC), LSA-422 (MTFVC), LSA-423 (MDLAI), LSA-424 (MTLAI), LSA-425 (MDFAPAR), LSA-426 (MTFAPAR), LSA-450 (MTFVC-R), LSA-451 (MTLAI-R), LSA-452 (MTFAPAR-R)

Reference Number: SAF/LAND/UV/VR_VEGA/2.0
Issue/Revision Index: Issue 2.0
Last Change: 24/02/2016

DOCUMENT SIGNATURE TABLE

	Name	Date	Signature
Prepared by :	F.J. García-Haro, F. Camacho		
Approved by :	Land SAF Project Manager (IPMA)		

DOCUMENTATION CHANGE RECORD

Issue / Revision	Date	Description :
Version 1.0	8/11/2013	Version prepared for the ORR
Version 2.0	24/02/2016	Version prepared for the reprocessing chain, including changes made to version 2 of FVC (LSA-421 and LSA-422), LAI (LSA-423 and LSA-424), and FAPAR (LSA-425 and LSA-426) products.

DISTRIBUTION LIST

Internal Consortium Distribution		
Organisation	**Name**	**No. Copies**
IPMA	Isabel Trigo	
IPMA	Sandra Coelho Freitas	
IPMA	Isabel Monteiro	
IPMA	Carla Barroso	
IPMA	João Paulo Martins	
IPMA	Pedro Diegues	
IPMA	Benvinda Barbosa	
IPMA	Ana Veloso	
IDL	Carlos da Camara	
IDL	Teresa Calado	
IDL	Sofia Ermida	
KIT	Folke-S. Olesen	
KIT	Frank Goettsche	
MF	Jean-Louis Roujean	
MF	Dominique Carrer	
MF	Gregoire Jacob	
RMI	Françoise Meulenberghs	
RMI	Arboleda Alirio	
RMI	Nicolas Ghilain	
UV	F. Javier García Haro	
UV/EOLAB	Fernando Camacho	
UV	Beatriz Martinez	
UV	María Amparo Gilabert	
UV	Manuel Campos Taberner	
UV/EOLAB	Jorge Sánchez	
External Distribution		
Organisation	**Name**	**No. Copies**
EUMETSAT	Frédéric Gasiglia	
EUMETSAT	Dominique Faucher	
EUMETSAT	Lorenzo Sarlo	
EUMETSAT	Lothar Schueller	
EDISOFT	Tiago Sepúlveda	
EDISOFT	Joana Rosa	
EDISOFT	Joaquim Araújo	
GMV	Mauro Lima	

Steering Group Distribution		
Nominated by:	**Name**	**No. Copies**
IPMA	Pedro Viterbo	
EUMETSAT	Lorenzo Sarlo	
EUMETSAT	Lothar Schueller	
EUMETSAT	Christopher Hanson	
EUMETSAT	Harald Rothfuss	
STG/AFG (USAM)	Francesco Zauli	
MF	Jean-François Mahfouf	
RMI	Rafiq Hamdi	
KIT	Johannes Orphal	
VITO	Bart Deronde	

TABLE OF CONTENTS

List of Figures

List of Tables

1 Introduction

The EUMETSAT Satellite Application Facility on Land Surface Analysis (Land-SAF) generates, on an operational basis, Vegetation Parameters (VEGA) from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard Meteosat Second Generation (MSG) satellites (Schmetz et al., 2002; Trigo et al., 2010). The VEGA products are the Fractional Vegetation Cover (FVC), the Leaf Area Index (LAI) and the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR). This document details the algorithm used for the retrieval of these products. There are two version of each product, the daily products MDFVC (LSA-401), MDLAI (LSA-404) and MDFAPAR (LSA-407) and the 10-day products MTFVC (LSA-402), MTLAI (LSA-405), MTFAPAR (LSA-408). The methodology detailed here applies also to the re-processed VEGA dataset – LSA-450 (MTFVC-R), LSA-451 (MTLAI-R), LSA-452 (MTFAPAR-R).
The products are computed from level 1.5 SEVIRI data corresponding the short-wave channels at 0.6 µm (VIS1), 0.8 µm (VIS2) and 1.6 µm (NIR). The VEGA products are generated at full spatial resolution (3 km/pixel sampling distance at nadir), for 4 different geographical areas within the MSG disk (i.e., Europe, Northern Africa, Southern Africa and South America, see Figure 1) as well as in a single MSG-Disk product covering the full the Meteosat disk.

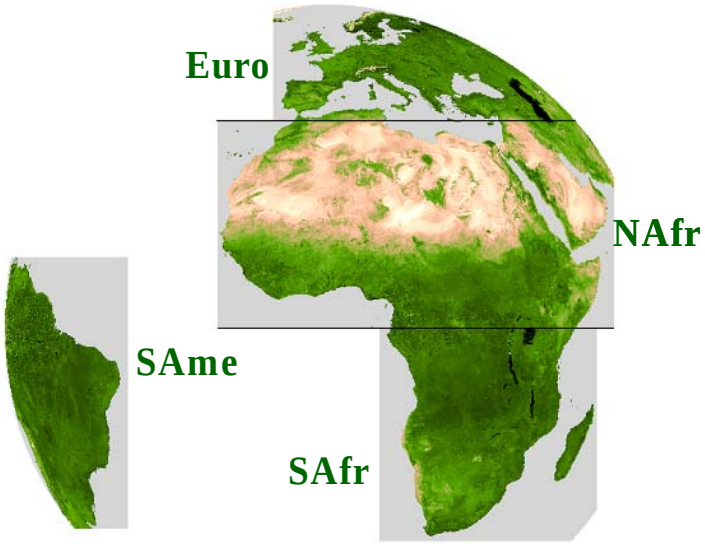


Figure 1 - The LSA SAF geographical areas.

The monitoring of earth surface dynamic processes requires global observations of the structure and the functioning of vegetation. Among the several state variables, the FVC, the LAI and the FAPAR are key variables for a wide range of land biosphere applications. The FVC represents the fraction of green vegetation covering a unit area of horizontal soil. The FVC determines the partition between soil and vegetation contributions for further estimates of total emissivity and temperature. LAI is a quantitative measure of the amount of live green leaf material present in the canopy per unit ground surface. The FVC and LAI are important structural properties of land

surface areas occupied by plant canopies, which yield complementary information to describe the three-dimensional structure of the vegetation attributes. FAPAR is a measure of how large a fraction of the sunlight leaves absorb in the 0.4–0.7 μm spectrum. FAPAR expresses a canopy's energy absorption capacity and is thus a key variable in models assessing vegetation primary productivity and, more generally, in carbon cycle models (e.g. Sellers et al., 1997; Gobron and Verstraete, 2009).

The present document is one of the product manuals dedicated to LSA SAF users. The algorithm theoretical basis of the vegetation parameters of the daily (MDFVC, MDLAI and MDFAPAR) and ten-day (MTFVC, MTLAI and MTFAPAR) Vegetation Products generated by the LSA SAF system are described in the following sections. The characteristics of SEVIRI based VEGA products provided by the LSA SAF are described in Table 1. Further details on the LSA SAF product requirements may be found in the Product Requirements Document (PRD) and the Product User Manual (PUM) which are available on the LSA SAF website http://landsaf.meteo.pt).

Table 1.- Product Requirements for MSG VEGA products, in terms of area coverage, resolution and accuracy.

			Resolution		Accuracy		
Product	**Identifier**	**Coverage**	**Temporal**	**Spatial**	**Threshold**	**Target**	**Optimal**
MDFVC	LSA-401	MSG disk	1-day	MSG pixel	20%	15%	10%
MTFVC	LSA-402	MSG disk	10-days	MSG pixel	20%	15%	10%
MDLAI	LSA-404	MSG disk	1-day	MSG pixel	1,5	1	0,5
MTLAI	LSA-405	MSG disk	10-days	MSG pixel	1,5	1	0,5
MDFAPAR	LSA-407	MSG disk	1-day	MSG pixel	20%	15%	10%
MTFAPAR	LSA-408	MSG disk	10-days	MSG pixel	20%	15%	10%

2 Theoretical Framework

2.1 Introduction

The FVC and LAI are important structural properties of land surface areas occupied by plant canopies, which yield complementary information to describe the three-dimensional structure of the vegetation attributes. For fully and healthy developed canopies, LAI indicates the amount of green vegetation that absorbs or scatters the solar radiation in determining the characteristics of the remote sensing signal. In other words, it represents the interface between the soil background and the atmosphere for the energy and mass exchanges. The scaling effect is quite important for LAI and only marginal for FVC because this latter is quasi-linearly related to the reflectance (Malingreau and Belward 1992, Weiss et al. 2000). FAPAR is generally well correlated with the LAI, the more for healthy fully developed vegetation canopies. FAPAR depends both on canopy structure, leaf and soil optical properties and irradiance conditions (Baret and Guyot, 1991).

The FVC determines the partition between soil and vegetation contributions for further estimates of total emissivity and temperature. For such, an accurate assessment of FVC is mandatory for a thorough description of land surface processes in the surface parameterisation schemes implemented in the climate and weather forecasting models. FVC is generally close to FAPAR

with the advantage of being defined independently of illuminations conditions making it an intrinsic canopy attribute. Besides, the FVC is relevant for a wide range of Land Biosphere Applications such as agriculture and forestry, environmental management and land use, hydrology, natural hazards monitoring and management, vegetation-soil dynamics monitoring, drought conditions and fire scar extent. The LAI is a key input of Numerical Weather Prediction (NWP) models, regional and global climate modelling, weather forecasting and global change monitoring. Besides, the LAI is relevant for Land Biosphere Applications such as agriculture and forestry, environmental management and land use, hydrology, natural hazards monitoring and management, vegetation-soil dynamics monitoring and drought conditions. FAPAR has been recognized as one of the fundamental terrestrial state variables in the context of the global change sciences (Steering Committee for GCOS, 2003; Gobron et al., 2006). It is a key variable in models assessing vegetation primary productivity and, more generally, in carbon cycle models implementing up-to-date land surfaces process schemes (e.g., Sellers et al., 1997). Besides, it is an indicator of the health of vegetation.

FVC, LAI and FAPAR are used extensively to represent vegetation abundance and canopy structure and reflect changes in vegetation from global to local scales because they echo to rapid changes in climatic conditions or environmental stress factors. For effective use in coarse scale models, these variables must be collected over a long period of time and for all ecosystems of the terrestrial surface. To resolve rapid changes of vegetation status and amount under both the influence of climate and human activities, relatively high frequency observations are required, currently provided by the SEVIRI instrument.
Many available global maps used in climate and NWP still depend on land cover classifications or are based on correlation between vegetation properties and simple spectral indices (e.g. ECOCLIMAP). However, in the last few years, methodologies have been developed to estimate land surface bio-physical parameters from large scale optical sensors in an operational way. In addition to remove atmospheric effects from the signal, algorithms for retrieving vegetation parameters from wide FOV sensors should deal properly with the anisotropic behavior of surface's reflectances. These algorithms have been implemented in operational processing lines to provide advanced bio-physical products from POLDER (Leroy et al. 1997, Roujean and Lacaze 2002), MODIS and MISR (Knyazikhin et al. 1999), MERIS (Gobron et al. 1999, Bacour et al. 2006, Baret et al. 2007), SEAWIFS (Gobron et al. 2001), VEGETATION (Baret et al. 2007, Bartholomé et al. 2006) and GLOBCARBON (Plummer et al. 2006, Deng et al. 2006).

These operational methods can be roughly divided in two large groups: (*i*) Methods based on determining relationships (e.g. transfer functions) between vegetation indices and biophysical parameters and (*ii*) methods based on the inversion of physical models. The transfer function algorithm is associated with uncertainties originating from the calibration of the semi-empirical model, and the applicability of the algorithm to a range of vegetation types, angular configurations, seasons and locations.

Radiance measured by sensors inevitably consists of radiance from multiple ground cover because of surface heterogeneity at sub-pixel scale and the contribution of the neighbouring pixels due to the sensor's PSF. Taking also into consideration the influence of soil background in traditional vegetation indices as well as the influence of sun/view configuration in spectral reflectance, it is thus clear that we need to define more general statistical relationships based on simulations. Gobron et al (1999) proposed an optimisation of vegetation indices for retrieving FAPAR, based on simulations of a radiative transfer canopy model. Operational processing lines

for retrieving FAPAR from MERIS (Gobron et al. 1999) and SeaWiFS (Gobron et al., 2002) are implemented using this method.

Methods based on the inversion of physically based models are theoretically powerful and they can potentially be applied to varying surface types. The primary limitation of this approach is the difficulty in determining all the input parameters of the model properly, along with the fact that a different set of parameters may yield to very similar spectro-directional signatures leading to unstable solutions (Weis and Baret, 1999). Many inversion methods utilize one-dimensional or turbid models, thereby assuming surface homogeneity within an image pixel. Hence the sub-pixel mixing of land cover types, which is especially common at coarse resolutions satellites due to the complex spatial patterns of sub-pixel vegetation cover, is usually ignored. One practical solution to this problem is the linearization approach, which moreover permits modeling of spatial heterogeneity explicitly.

The process of identifying the sub-pixel proportions of the constituent components is called spectral mixture analysis (SMA). SMA approaches are methods especially adequate for global studies, since the spatial variability within pixel is high. It is assumed that the random variables associated to each ground cover component are statistically independent, implying that there is no significant multiple scattering between them, i.e. all photons reaching the sensor have interacted with just one cover type. The use of SMA is widely accepted because they yield an understandable solution to the complex problem of mixed reflectance. A variety of studies have supported this assumption on a number of space-borne sensors, including TM, AVHRR, VEGETATION and MODIS (e.g., Defries et al. 2000, Camacho-de Coca et al. 2004; García-Haro et al. 2005a; Fisher and Mustard, 2007).

2.2 Algorithm description

The LAI is estimated from a FVC using a semi-empirical approach as in Roujean and Lacaze (2002). This method relies on a tractable physical model for interception of solar irradiance by vegetative canopies. The vegetation canopy is represented by a semi-infinite plane-parallel horizontally homogeneous layer. The medium is characterized by the total leaf area index and the mean leaf inclination function. A clumping index is introduced to account for non-randomness in the leaf spatial distribution. Its use allows for improved estimation of radiation interception and for unbiased estimates of true LAI in highly clumped canopies such as conifer and tropical forests.

Thus the algorithm for retrieving FVC and LAI relies on optimised SMA methods. This approach has consistently proven to improve in the accuracy of the estimates compared to traditional remote sensing techniques that use unadjusted spectral vegetation indices (Gilabert et al. 2000; Peddle *et al*., 2001; Hu *et al.*, 2004). The suitability of using this algorithm was demonstrated in García-Haro et al. (2004). Traditionally, spectral vegetation indices were developed using information obtained only in the red and near infrared wavelengths. The incorporation of information from the middle infrared, MIR (i.e. channel 1.6 of MSG) bands tends to improve the estimation because channels in the MIR have been observed to exhibit the greatest degree of sensitivity to changes in LAI (Chen et al. 2002).

A statistical approach is proposed for retrieving daily FAPAR from BRDF data, corrected of surface's reflectance anisotropy and minimising the effect of soil reflectance (Roujean and Bréon, 1995). The principle of the algorithm is based on simulations of visible and near infrared spectral reflectance values in optimal angular geometries identified based on numerical experiments (simulations of a radiative transfer code). A vegetation index, called RDVI (Renormlized Difference Vegetation Index), is introduced, which shows to be less sensitive to background reflectance variability. A pre-established relationship is then applied between RDVI computed in an optimal angular geometry and daily FAPAR.

FAPAR algorithm was successfully applied and validated using different coarse spatial resolution sensors, including POLDER/ADEOS (Roujean and Lacaze, 2002), and POLDER/PARASOL showing a good consistency with ground truth (Baret and Pavageau, 2006). In order to evaluate the performance of this algorithm against other methods, the algorithm was applied to VGT (VEGETATION/SPOT) CYCLOPES L3A data, and compared with the CYCLOPES L3B FAPAR product (Baret et al., 2007) and MODIS C4.1 products. The inter-comparison was carried out over one-year of data for the European continent. This exercise demonstrated the good spatial and temporal performance as compared with equivalent satellite products of the Roujean and Bréon model (Verger et al., 2006). The suitability of using this algorithm for retrieving FAPAR from SEVIRI data was investigated in Camacho-de Coca (2007).

The application of LSA SAF methodology to VGT data representing an optimal BRDF sampling (García-Haro et al. 2005b) outperforms the results from other SMA literature methods such as MESMA (Roberts et al. 1998), VESMA (Camacho-de Coca et al. 2004) and VMESMA (García-Haro et al 2005a). The LSA SAF methodology was applied to VGT data and the results were compared with ground truth based maps distributed on Europe and representing different continental biome types and conditions, obtaining excellent results (RMSE values of 0.08 for FVC, 0.22 for LAI and 0.07 for FAPAR) although over a limited number of sites (Verger et al. 2007). In addition, the prototype of LSA SAF LAI product with VGT and MODIS data over Europe was compared with CYCLOPES LAI (Baret et al., 2007) and MODIS LAI C5 (Knyazikhin et al., 1999) products showing the reliability of the LSA SAF LAI algorithm (Verger et al., 2009).

2.3 Algorithm inputs

LSA SAF algorithms to retrieve FVC, LAI and FAPAR rely on the use of optimal geometries, which reduces the uncertainties resulting from different canopy types and background reflectance.

The algorithms use as input the directional coefficients of the BRDF model for the different spectral channels resulted from simulating the BRDF following the general expression presented by Roujean et al. (1992):

$$R(\theta_v, \theta_s, \phi) = k_0 + k_1 f_1(\theta_v, \theta_s, \phi) + k_2 f_2(\theta_v, \theta_s, \phi) \quad (1)$$

where θ_v, θ_s, ϕ stand for the sun zenith, view zenith and relative azimuth angles, respectively, and f_1, f_2 stand for the geometric and volume scattering kernels, respectively. The normalisation of the image to a common geometry solves the major source of uncertainty on wide FOV sensor's data, which is introduced by the anisotropy of surface's reflectance. The negative impact due to view/sun

angles variations in surface reflectance are thus minimized because the products are derived using the same geometry for the whole SEVIRI disk.

While the FVC and LAI products rely only on the k0 BRDF parameter, the FAPAR relies on the use of the three BRDF parameters. Inputs for retrieving FVC are thus atmospherically corrected cloud-cleared TOC k_0 parameters (Daily, 10-Day) in the three relevant SEVIRI spectral channels: red (VIS-0.6), near-infrared, NIR (VIS-0.8) and middle-infrared, MIR (IR-1.6). Physically the k_0 parameters correspond to isotropic reflectance, i.e. reflectance factor values directionally normalized to reference illumination and observation zenith angles of 0°. This geometry leads to a minimum contribution of the shadow proportion (hotspot geometry) and a physically correct estimation of FVC (Roujean and Lacaze 2002), coinciding with the complement to unity of the gap fraction at nadir direction. Estimating the FVC with increased values of the sun zenith would lead to an overestimation of FVC. At this geometry, however, the contribution of illuminated soil background is significant, constituting thus a source of "noise" that has long been recognized as major problem in remote sensing of vegetation (e.g. Huete *et al.* 1988).

Inputs for retrieving FAPAR are atmospherically corrected cloud-screened TOC k_i parameters **(**Daily, 10-Day**)** in two SEVIRI channels: red (VIS-0.6) and near-infrared, NIR (VIS-0.8). The quality of the SEVIRI BRDF parameters has been addressed in a related document (SAF/LAND/UV/VEGA_VR/2.1). Problematic areas with large BRDF uncertainty values correspond to high latitudes over Europe and in South America. k2 product presents generally the largest uncertainties as well as shaky temporal profiles (only for daily products) in other regions, particularly in Western Africa and regions in the south hemisphere. The quality of the FAPAR is directly related to the BRDF quality, whereas FVC and LAI uncertainties are associated to the quality of k0.

2.4 FVC Algorithm description

The algorithm to estimate the FVC assumes that observed variability in the mixed pixels arises from the intraclass variability of mixed pure classes representing the target (vegetation) and the soil background. In general, target and background are non Gaussian with unknown probability density function. In practice, this involves that large number of pixels representing each class is available or that the parametric modeling of the distribution of each class is available. The algorithm relies on a probabilistic SMA method in which endmember signatures are no longer treated as constants, but they are represented by multi-modal probability density functions. The use of standardized SMA improves understanding of the impact of endmember variability on the derivation of subpixel vegetation fractions at a global scale.

Notationally, let **r** be the spectrum of each mixture pixel, i.e. a column vector $(r_1,r_2,...,r_n)$, where n is the total number of bands. The traditional SMA assumes that **r** can be approximated by a linear mixture of endmember reflectances **E** weighted by their corresponding fractional proportions **f**:

$$\mathbf{r} = \mathbf{E}\,\mathbf{f} + \boldsymbol{\varepsilon} \qquad (2)$$

where **E** [$n \times c$] is the matrix of endmembers, **f** is a vector with the c unknown proportions in the mixture, and **ε** is the residual vector. In the classical SMA methods, the within-class variability of the endmembers is usually regarded as noise (fluctuations around the expected values) attributed to variations of physical factors such as humidity, soil types, topography. This variability is pooled into a residual term **ε**, that would ideally be a multivariate normal distribution with mean zero. The

mixing equation is accompanied by two constraints: (1) the normalisation constraint says that a pixel is well defined by its components, whose proportions or abundances should therefore add up to unity, and (2) the positivity constraint says that no component of a mixed pixel can make a negative contribution.

The observed reflectances are usually correlated and moreover affected by measurement errors and therefore it is convenient to search for the best solution in a statistical sense and to quantify the uncertainties of the retrieved parameter estimates. The least-square principle establishes that the unknown parameters are those that minimise the Mahalanobis distance between the pixel **r** and point **E f**:

$$\chi^2 = (\mathbf{r} - \mathbf{E}\,\mathbf{f})^{T}\,\mathbf{V}(\mathbf{r})^{-1}\,(\mathbf{r} - \mathbf{E}\,\mathbf{f}) \qquad (3)$$

where $\mathbf{V}(\mathbf{r})$ denotes the error matrix of the observations **r**.

The effectiveness of the linear mixture model is dependent on the degree of separation of the different signatures, as well as the level of noise which is present in the scene. This imposes clear limits on the number and similarity of the land cover classes that can be reliably distinguished. SMA methods usually use a few endmembers to represent the entire scene, which is a significant simplification since they do not fully identify the less prevalent materials in a multispectral scene. Besides, many of these methods assume that EMs cannot form linear expressions of one another (i.e. they may not have a non-singular matrix).

The intra-species spectral variability, lighting and topographic effects, uncertainty related to apparent surface reflectance retrievals, and noise in field or image endmembers are sources of uncertainty that compromise the ability to retrieve relevant information using SMA. Since vegetation in different ecosystems are spectrally dissimilar to typical green vegetation, SMA using only green vegetation spectra will lead to significant errors of FVC. It is an essential dilemma in SMA that the continuous variation of endmember signatures demands more endmembers to represent them, while the limited dimensionality of coarse resolution remotely sensed data necessitates fewer endmembers (Song, 2005). Several efforts have been recently conducted in the SMA literature to account for the natural variability in material spectra with a relatively high degree of success (Bosdogianni et al.1996, Roberts et al. 1998; e.g. Bateson et al. 2000; García-Haro et al 2005a; Song 2005). The algorithm adapted a variable multiple endmember spectral mixture analysis (García-Haro et al. 2005), in which soil and vegetation components are represented by a multi-modal probability density function.

Figure 2 illustrates a schematic representation of the problem in a two-dimensional (e.g. the red-NIR) space. The proposed method assumes that at a SEVIRI spatial resolution, any point in the vegetation region can combine with another point in the non-vegetation region to produce a mixture signal. Both vegetation and non-vegetation classes are represented by a multi-modal distribution attributable to differences due to biophysical and biochemical composition. Rather than a single signature to represent a certain vegetation species or soil type, its entire variability is accounted for by using Gaussian probability densities. This addresses more reliably the pixel deviations from its expected value due to the natural variability of the scene materials.

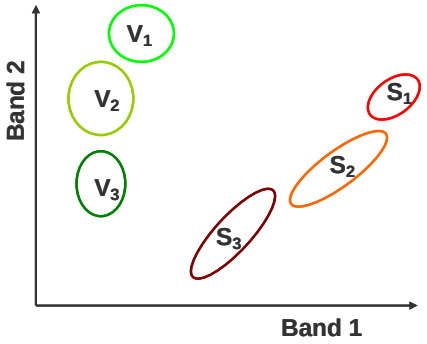


Figure 2 - Illustration of the probabilistic mixing model concept for 2 band case. Drawn are isoprobability contours (Gaussian clusters) associated with pure endmembers, i.e. feasibility regions on the spectra of the constituents due to the natural variability in the material itself and other effect such as atmospheric/illumination conditions and data noise.

The coarse spatial resolution of SEVIRI data poses a significant challenge for endmember selection in traditional SMA. The algorithm assumes that each unknown pixel can be modelled by those candidate models (i.e. pairs of vegetation and soil subclasses) which are compatible with the SMA assumptions (i.e. they should lie on the "extended convex hull" defined by the model). The design of the algorithm is based on a five-step procedure (see figure 3). A brief summary of the main algorithm steps is now given.

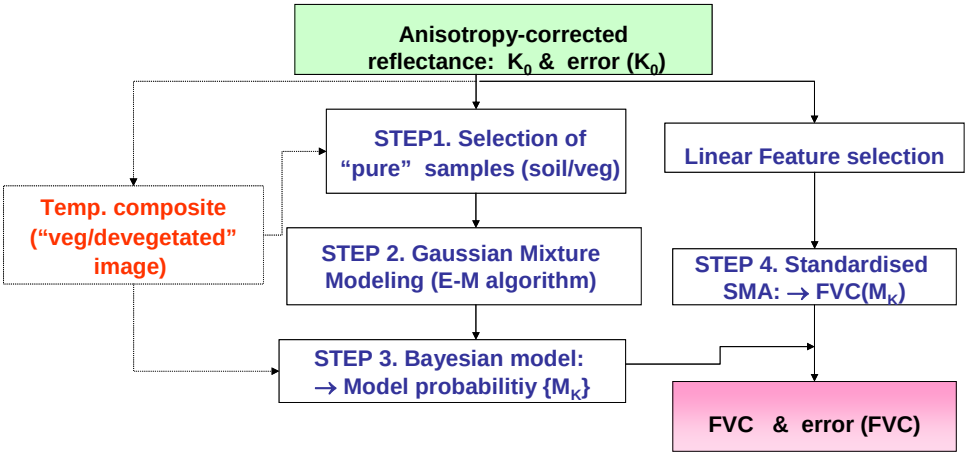


Figure 3 - Flow chart of the algorithm for FVC determination.

An important improvement of the VEGA version 3.0 to guarantee the success of modelling approach is the accumulation of temporal information. Two composite k_0 images were generated, a vegetated one corresponding to the peak of season and a devegetated one corresponding to the minimum canopy closure. The goal was to select the observation, on a per pixel basis, from all high quality observations (cloud- and snow-free) over a one year period to construct a composite k_0 image reflecting the period which present the maximum/minimun vegetation activity, based on the FVC annual cycle. Different vegetated/devegetated k_0 images were computed for several years in order to assess the possible influence of the intra-annual variability and select the most appropriate year. These composite images have critically contributed to the success of step 1 (to more reliably identify

bare soils and very densely vegetated areas) and step 3 (allowing to improve the identification of soil/vegetation models for each pixel).

Step1. Definition of an exhaustive training data set

The aim is to provide an adequate characterization of the variability of the pure components. Soil (bare soil, rock or human-built surfaces) and vegetation (which is prevalent in crops, herbaceous or forest ecosystems) classes are represented by a large number of training set of pixels. A large robust database allows to minimize the effect of noisy observations, inaccurate atmospheric correction and adverse environmental conditions.

The training process makes use of the very detailed information over the SEVIRI geographical areas coming from land cover classifications and other validated biophysical products. In the current (v3.0) version, the selection of vegetation/soil pure pixels used the vegetated/devegetated k_0 images and relied on a multi-criterion approach which considers different data sources.

(i) The training samples were chosen to lie in the borders of the convex hull of the red (0.6 μm), NIR (0.8 μm) and MIR (1.6 μm) SEVIRI channels, for the two k0 composite (vegetated/devegetated) images

(ii) The samples were chosen to be homogeneous (ideally covered by a single surface type at the coarse SEVIRI resolution) based on GLC2000 classification. Some selected pixels were further examined based on fine resolution imagery (e.g. CORINE-100m over Europe, MOD13Q1 reflectance 250-m, Landsat reflectance 30-m).

(iii) Some samples were also added based on ancillary data and exploiting the availability of homogeneous Belmanip sites, the GOFC-GOLD Reference Data Portal and crowdsourcing data such as Geo-Wiki (Fritz et al. 2009).

(iv) Temporal curves of FVC (version v2.1) were to examine seasonal profiles. The >8% percentile for soil and >92% perecentile for vegetation were considered to exclude non pure pixels. Training areas of non-vegetated class included desert areas, with practically no trace of vegetation along the year. Other land cover classes such as sparsely vegetated and shrubland land cover types were selected to represent the underlying soil background in most vegetated areas.

(v) Endmembers are not merely represented by the most spectrally pure targets but they should span over a wider confidence region. A growing region algorithm was used to increase the number of pixels within a 5x5 pixel neighbourhood. The variability of the pixels that represent the local pure classes is then responsible for the uncertainty with which the mixture proportions can be identified.

Samples were further verified using two different purity methods, the convex hull and the unmixing abundances computed in an iterative process, and the possible outliers were filtered out. Pixels having less than 90-95% of soil/vegetation were thus excluded. Finally, a cross-checking with validated LAI MODIS and Copernicus Global Land products was used to confirm that no outliers were present (samples were within the <15% percentile of LAI values for soil and >85% percentile for vegetation).

Step 2. Gaussian Mixture Model of the vegetation/soil components

The analysis method is probabilistic, based on the assumption that the observations in the *k*-th class are generated by a probability specific to that class ϕ_k. The usual assumption is that Gaussian distribution function may accommodate the variability of the different soil types and vegetation

varieties that can be found in the scene. For a vector $\boldsymbol{x} \in \Re^n$ the class conditional distribution is given by

$$\phi_k(\boldsymbol{x} \mid \mu_k, \Sigma_k) = (2\pi)^{-n/2} \left|\Sigma_k\right|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu_k)^T \Sigma_k^{-1}(\boldsymbol{x}-\mu_k)\right) \quad (4)$$

Each Gaussian (cluster) is described by its mean vector μ_k and covariance matrix Σ_k. Geometric features (shape, volume, orientation) of the clusters are determined by the covariances Σ_k, which may also be parameterized to impose cross-cluster constraints. In general, it corresponds to a long, thin ellipsoid for soil components and spherical shapes for vegetation components (see figure 4a). The model thus incorporates the noise and variability of the sample data into statistical limits around each cluster centroid.

However, the samples selected include a variety of vegetation species (i.e. they do not comprise of pure stands of a single species). Although the variability of each class can be assessed using a nonparametric representation of the constituent pure classes (Petrou and Foschi 1999), a parametric flexible model is preferred to fit the training data. A mixture model weighted sum of Gaussian distributions (or clusters) f_k has been selected, which consists of a linear combination of normal distributions (Bishop 1995):

$$f_j(\mathbf{x} \mid \mu_{jk}, \Sigma_{jk}) = \sum_{k=1}^{G_j} \tau_k \phi_k(\mathbf{x} \mid \mu_{jk}, \Sigma_{jk}) \qquad \text{j = s, v} \quad (5)$$

where τ_k is the probability that an observation belongs to the *k*-th component ($\tau_k \geq 0$; $\sum_{k=1}^{G_j} \tau_k = 1$), j represents the soil (s) or vegetation (v) component, and G_j the number of Gaussian components. This idea has been suggested a number of times in the literature (eg. Scott 1992) and is the basis of mixture discriminant analysis (MDA). This reflects a common situation where different types of background or vegetation varieties are present in the same pixel.

The algorithm uses the Expectation-Maximization (E-M) approach (McLachlan and Krishnan, 1977, Bishop 1995) to estimate the means μ_{jk} and covariances Σ_{jk} the individual Gaussian components, which shows to be an efficient method when the number of Gaussians is defined beforehand. This iterative approach can be shown to converge if the data conform reasonably well to the model and the iteration is started at reasonable values. The goal is to maximize the likelihood of observed data. The *K*-means algorithm was used to initialise the E-M parameters (μ_{jk}, Σ_{jk}). Multiple initializations were made to avoid numerical problems local maxima.

A necessary input is the number and type of distributions. There are trade-offs between the choice of the number of clusters and that of the clustering model. Our requirement for this choice is that the number and type of distributions must provide a faithful representation of the data. If a simpler model is used, then more clusters may be needed to provide a good representation of the data. If a more complex model is used, then fewer clusters may suffice. As a simple example, consider the situation with a single Gaussian cluster whose covariance matrix corresponds to a long, thin ellipsoid. If a model with equal-volume spherical components (i.e. the model underlying the *k*-means method) were used to fit this data, then more than one hyperspherical cluster would be needed to approximate to the single elongated ellipsoid. A typical number of 4-5 Gaussians has been used to represent the variability of the soil and vegetation components over the different SEVIRI

geographical areas. Although the selection of the database and tuning of EM parameters was made separately for each SEVIRI region, pixels from other regions presenting similar characteristics were added. This may have favoured that no discontinuity in the retrievals is observed in the borders between adjacent regions.

Step 3. Model selection

Let us define a model M_k as a pair of class-conditional distributions for vegetation-soil, i.e. $M_K \equiv (f_{veg(k)}, f_{soil(k')})$. The algorithm computes all possible models by taking all possible pairs $\{M_1, M_2, M_N\}$ since at SEVIRI spatial resolution, any vegetation variety could combine with any non-vegetation type to form a mixture signal. Let $\pi(M_K)$ be the *a priori* probability of having the model M_k at a particular pixel. A Bayesian approach, which is based on minimization of the Bayes risk most often seen in classification problems, is used for the model selection. The basic idea is that the posterior probability or membership of model M_K given pixel data **r**, namely $p(M_K \mid \mathbf{r})$, is proportional to the probability of the data given model M_k, namely $p(\mathbf{r} \mid M_K)$, times the model's prior probability $\pi(M_K)$:

$$p(M_K \mid \mathbf{r}) = \frac{p(\mathbf{r} \mid M_K) \cdot \pi(M_K)}{\sum_{i=1}^{N} p(\mathbf{r} \mid M_i) \cdot \pi(M_i)} \tag{6}$$

This approach assigns a weight to the models M_k according its posterior probability in the mixture pixel **r**, $p(M_K \mid \mathbf{r})$. Although *a priori* probabilities are often unknown and taken to be equal, they provide a means to inject prior information in the algorithm. *A priori* probabilities $\pi(M_K)$ were used to reduce the possibility of the misidentification of the models based upon prior-knowledge about the statistical distribution of the reflectance for the main land cover classes.

The class-specific probability of each individual pixel, $p(\mathbf{r} \mid M_k)$, can be obtained as follows:

$$p(\mathbf{r} \mid M_K) = \int_x \int_{x'} \phi_{veg(k)}(\mathbf{x} \mid \mu_k, \Sigma_k)_{veg(k)}(x) \cdot \phi_{soil(k')}(\mathbf{x'} \mid \mu_{k'}, \Sigma_{k'}) \Gamma(\mathbf{x}, \mathbf{x'}, \mathbf{r})\, d\mathbf{x'}\, d\mathbf{x} \tag{7}$$

where $\Gamma(\mathbf{x}, \mathbf{x'}, \mathbf{r})$ is an "accept-reject" function that is the unity when the line joining **x** (vegetation) and **x'** (soil) intercepts the region in the feature space centred at the mixture **r**, and is zero otherwise.

Solving the integral (7) basically consists in counting the number of model samples from each possible mixture combination that can give rise to the mixture pixel **r**. We considered around **r** an envelope volume **V**(**r**) given by the radiometric uncertainties attached to the input, i.e. the covariance structure of the k_0 product. This integral can be solved using either quadrature rules or Monte Carlo methods (i.e. from endmember spectra drawn from multivariate normal distribution).

Although the analysis can be applied using as input a single date (as in version v2.1), the optimal results can be achieved if multiple observations over a seasonal period are available. An important improvement was made in version v3.0 to more reliably determine the probability of soil/vegetation models by considering simultaneously the dates of maximum/minimum canopy closure (i.e the vegetated/devegetated k_0 images). For example, the devegetated image would correspond to periods

that do not have significant vegetation cover during the dry season (e.g. harvested crops) allowing thus improve identifying the spectral characteristics of underlying soil background.

The mixing space concept is illustrated in figure 4, which corresponding to the training soil and vegetation samples identified in Africa. Bare areas should be predominantly found in certain bare, sparsely vegetated and open shrublands (GLC2000 classes 19, 14 and 12) whereas purely vegetated areas should be mostly found in close forest classes, herbaceous and croplands (GLC2000 classes 1, 2, 13 and 16). The best suited number of mixture components (gaussians) was 4 for vegetation and 5 for soil.

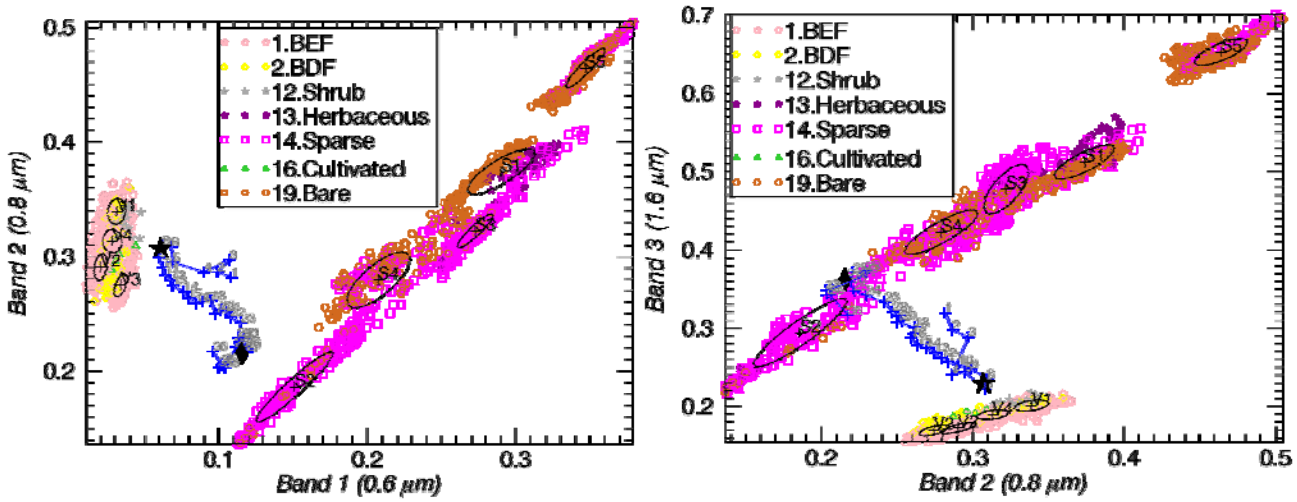


Figure 4 - Illustration of the probabilistic mixing model concept over Africa regions in the k0 space of the red (0.6 µm), NIR (0.8 µm) and MIR (1.6 µm) SEVIRI channels. Drawn are isoprobability contours (Gaussian clusters) associated with pure endmembers (V1-V4 for vegetation typs and S1-S5 for soils). BED and BDF refer to broadleaved evergreen and broadleaved deciduous forest, respectively.

We can observe that there is a high variability in soil background reflectance. In general, desert areas present abnormally high values in the red channel, with practically no contrast between red and NIR and, therefore. However, the spectral properties of these bare areas does not fit well the reflectance of underlying bareground in the vegetated areas, which in general presented darker reflectance.

The use of temporal information allows for improved identification of the spectral characteristics of vegetation and underlying soil background. This is illustrated in figure 4 for an open broadleaved deciduous pixel. Blue symbols correspond to the migration of this pixel during an entire year (at a decade time step and numbered from 1 to 36). The greater the amount of vegetation present, the greater the NIR reflectance and the lower the red reflectance. The period of drier conditions (decade 8) correspond to the devegetated period, in which the pixel situates relatively close to the soil line. After the start of development phase, it departs from the soil line, eventually reaching a high canopy closure. The maximum (vegetated) and minimum (devegetated) FVC values were marked as solid diamond and start, respectively, in the k_0 feature space. The use of reflectance in these two periods, allowed identifying the most likely models for the vegetation (V1 and V4) and soil (S2), i.e. the darkest soil type.

Step 4. Estimation of FVC

A feature extraction step consists of mapping the input vector of observations $\mathbf{r} \in \Re^n$ (in our case, n=3) onto a new feature description $\mathbf{u} \in \Re^m$. Our purpose is allowing the classes to be separated, and remove, as far as possible, noise and other errors of the inputs. In previous VEGA algorithm (v2.1 version) the spectral dimensionality of the data was augmented, by adding an extra equation to the SMA. This is built by substracting the NIR and red channels: DVI= $(k_0)_{NIR}$ - $(k_0)_{VIS}$. The difference vegetation index (DVI) is well related to the FVC (Roujean and Lacaze, 2002) and, as opposed to nonlinear transformations such as NDVI, preserves the problem linearity. However, in the current version (v3.0), no use of DVI is made. Instead, the strategy has consisted in decreasing the relative influence of MIR channel, since the vegetation retrieval is specially hampered by the influence of undesired soil variability in the MIR waveband. By promoting information in the red and NIR channels retrievals were less affected by the unmixing method used, reducing the retrieval error. This strategy has shown to be more necessary in low-vegetated areas since it mitigates the underestimation of FVC in areas with abnormally high MIR background reflectance and the overestimation of FVC over dark soils.

A standardisation transform is applied, which transforms the data to a set of variations about the mean value with a mean value of zero and a standard deviation of one.

$$\hat{\mathbf{r}} = \frac{\mathbf{r} - \mu_r}{\sigma_r} \tag{9}$$

where $\hat{\mathbf{r}}$ is the standardized vector associated to the pixel vector $\mathbf{r}$, with mean μ_r and standard deviation σ_r. Using the standardised endmembers, $\hat{\mathbf{E}}_i$ (i =1,…,c), the unmixing is formulated as follows:

$$\hat{\mathbf{r}} = \sum_{i=1}^{c} \hat{\mathbf{E}}_i \hat{f}_i + \hat{\varepsilon} \tag{10}$$

where $\hat{f}_i$ is the proportion of such EM in the standardised coordinates, and $\hat{\varepsilon}$ is the residual vector expressed in standardised units. Although the solution is similar to the conventional SMA, the sum-to-one condition is now expressed as

$$\sum_{i=1}^{c} \frac{\hat{f}_i}{\sigma_{Ei}} = \frac{1}{\sigma_r} \tag{11}$$

We refer to this approach as standardized unmixing (García-Haro et al. 2005a). Through this standardization, SMA is less sensitive the brightness variability within each vegetation-soil component, at reducing the influence of external factors such as shading, brightness differences due to variability of surface roughness and terrain illumination. For example, spectra of very dense green vegetation corresponding to different canopy structures such as agricultural crops and forests, which are spectrally very different, become very similar after applying the standardization.

SMA based on standardised signatures is performed for all possible models. The algorithm performs a standardisation on both the endmember and the image spectra as a previous step before applying the SMA. The fractions of soil and vegetation are calculated using the algorithm of García-Haro et al. (1996), which provides a unique and unbiased solution that is computationally fast. Let FVC(M_K) be the fraction of vegetation obtained using the M_K model. FVC is estimated as a linear combination of single-model estimates:

$$FVC = \sum_{k=1}^{N} p(M_k \mid \mathbf{r}) \cdot FVC(M_k) \quad (12)$$

In this sum, the contribution of each model is weighted by its Bayesian *a posteriori* probability $p(M_K \mid \mathbf{r})$. Although this approach is similar to other literature methods (e.g. Asner and Heidebrecht 2002), the method is different since it decomposes the soil/vegetation into a number of subclasses and does not assume an equal model probability for the different subclasses. Rather, the Bayesian theory is applied which combines both, *posterior* and *prior* probabilities, of each class.

The use of a temporally invariant model for soil substrate and vegetation types -based on the calculated memberships- offers a stable frame to assess the evolution of vegetation dynamics for long-term monitor and reduces errors due to an undesired over-stratification of the scene.

Step 5. FVC uncertainty estimation

The statistical confidence intervals of FVC predictions, Err(FVC), is assessed taking into account two different sources of error, namely ε_{SMA} and ε_{model}, i.e.:

$$Err(FVC) = \sqrt{\varepsilon_{SMA}^2 + \varepsilon_{model}^2} \quad (13)$$

(i) Uncertainty due to the propagation of the input errors (ε_{SMA})

The impact of input errors, Err(k_0), on the prediction of FVC is assessed statistically taking into account from the usual error propagation laws as in the classical Spectral Mixture Analysis method. The inputs are treated as independent random variables with finite variances given by the covariance matrix of k_0 BRDF product. It is assumed that they are uncorrelated. The analytical solution for the standard deviation of estimating FVC (ε_{SMA}) can be found in García-Haro et al. (1996). In addition to covariance matrix of k_0 (i.e. the AL-C_0-CK product), ε_{SMA} is also influenced by the zone-dependent separability between vegetation and soil optical components (García-Haro et al. 2004), and is generally higher for dark surfaces. Hence the quality of retrievals is generally larger over herbaceous and cultivated areas whereas less accurate estimates are expected over needle leaf forests.

(ii) Uncertainty due to the model selection (ε_{model})

The second source of error, ε_{model}, is due to the dispersion (variance) of the possible solutions. This error recognizes that each SEVIRI pixel may have various mixture models of different soil types and vegetation varieties, and is expressed as follows:

$$\varepsilon_{model} = \sqrt{\sum_{K} p(M_K \mid \mathbf{r}) \cdot (FVC(M_K) - FVC)^2} \quad (14)$$

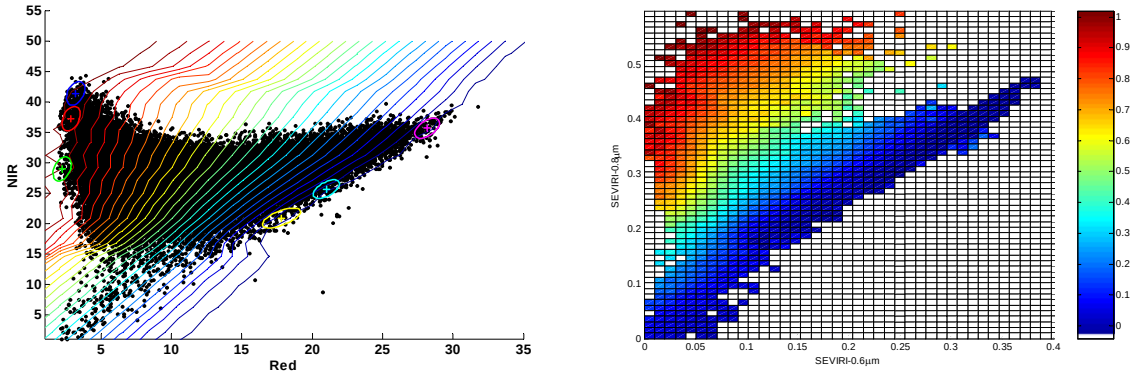
In general, the higher the number of candidate mixture classes contributing to the mixed pixel the higher the ε_{model} error. This error also penalizes pixels outside of the convex hull defined by the soil and vegetation densities (e.g. outliers, dark pixels along the coastline, etc.). The ε_{model} errors have been partly reduced thanks to the aid of priori knowledge extracted from a land cover classification.

While step 1 of FVC algorithm requires scientist guidance and oversight to inject information into the algorithm, steps 2-5 are fully automated. The first steps of the analysis (1-3) are computed beforehand and provide a set of auxiliary information, e.g. $p(M_K \mid \mathbf{r})$. This increases the robustness of the algorithm and speeds up considerably the determination of the vegetation products. Further, for bare surfaces such as deserts, FVC is not computed but it is set to a zero value. Unvegetated areas are assessed by the agreement of the land cover classifications.

Figure 5 allows understanding better the physical basis of the algorithm and provides details of its performance. The results correspond to an analysis performed over the European region using two different satellite scenes: VEGETATION/SPOT (figure 5a) and MSG/SEVIRI (figure 4b). The best possible prediction of FVC given a realized value of red and NIR, i.e. FVC=f($k_{0,red}$ - $k_{0,NIR}$), are obtained by averaging FVC values over pixels for which red and NIR fields fall in a sufficiently narrow interval. Although there is a one-to-one relationship between retrieved FVC and k_0 field in the red-NIR domain (the algorithm uses also additional information), the triangle envelope formed by the data points (i.e. mixed pixels) is the physical basis for SMA. Non-vegetation features distribute primarily along the so-called “soil line”. Pure vegetation is assumed to be on the top vertex, although it exhibits also a significant spectral variability.

Since green vegetation strongly absorbs solar radiation in the red spectral region, and strongly scatters it in the NIR, these two bands are widely used to characterize land surfaces from remote sensing data. The algorithm uses also MIR information, which provides useful information about the vegetation water content and mitigates the inaccuracy caused from saturation phenomena in estimating vegetation coverage. In the feature space of these three channels, the soil/vegetation components are spectrally distinct enough and the bands are sufficiently uncorrelated. The algorithm outperforms the common SMA approaches since:

1. Unlike common SMA approaches, this method shows to be effective in incorporating detailed information about the soil and vegetation conditions in the scene through the training data provided in step 1. It is easily obtained from a visual inspection that unknown pixels that are similar to those in the training set present similar FVC values. Therefore, the algorithm has the ability to generalize valuable information ingested in step 1.
2. The isolines capture the essential relationship between reflectance and FVC, showing a general agreement with literature studies (e.g. Huete et al. 1988).
3. There is a gradual change in FVC reflectance along the “reflectance triangle”. This reduces ε_{SMA} errors, since they are proportional to the local gradient of FVC in the k_0 feature space.



(a) (b)

Figure 5 - Projection of the FVC estimates, as derived from two different scenes over Europe, onto the spectral $k_{0,red}$ - $k_{0,NIR}$ feature space (a) VGT data for the 15th of June 2003 and (b) SEVIRI for the 19th of August 2005. In figure 5a, ellipses are projections of the ellipsoids defined by the covariances of the soil/vegetation Gaussian components (step 2). The isolines (lines connecting points of identical FVC) range from 0 to 1 FVC values at 0.04 spacing intervals.

In general, most literature methods do not respond to gradual change in reflectance in a smooth way. Although the probabilistic unmixing is applied on a per-pixel basis, the estimator is regularized and tends to mitigate the effects of noise and isolated mixed pixels. A sensitivity analysis has proven the accuracy and reliability of the proposed approach compared to other literature methods (e.g. García-Haro et al. 2005a). Estimated FVC is artefacts-free and spatially consistent (see examples in figure 8).

2.5 LAI Algorithm description

The solution of the radiative transfer problem can be reduced to the problem of diffuse transmission of the solar radiation by a medium of finite optical thickness. Roujean (1996) proposed a tractable physical model for interception of solar direct irradiance by leaf canopies. Assuming that leaves are flat with bi-Lambertian properties (reflectances and transmittance are isotropic), the direct transmittance which represents the fraction of direct incident beam above the canopy which reaches the soil background level is given by the classical expression (Nilson, 1971):

$$T_d(\theta_s) = \exp[-(G(\theta_s)/\mu_s)LAI] \quad (15)$$

where $\mu_s=\cos\theta_s$, being θ_s the solar zenith angle and $G(\theta_s)$ is the average extinction function (Ross, 1981). The terms of single and multiple scattering are expressed in a similar form to (15) (see Roujean (1996) for details) and for the total transmittance the following simplified form is proposed:

$$T(\theta_s) \approx \exp[-b(G(\theta_s)/\mu_s)LAI] \quad (16)$$

where the backscattered parameter, b, is assumed equal to 0.945 for all vegetation types (see in situ measurements of b in Roujean et al. (1997), table 5).

In this model, a random foliage distribution in the canopy is considered. It can be noticed that relative to a random leaf spatial distribution, clumpling enhances the probability of light penetration,

leading to a large underestimate of LAI if randomness is assumed (Roujean, 1996). Hence the correct formulation of the model interception is

$$\mathrm{FIPAR}(\theta_s) = 1 - \mathrm{T}(\theta_s) = 1 - \exp[-b(\mathrm{G}(\theta_s)/\mu_s)\Omega \mathrm{LAI}] \quad (17)$$

Ω is the clumping index (Nilson, 1971) which accounts for the degree of dependence of the vegetation stands position. $\Omega > 1$ for vegetation with a regular arrangement; $\Omega = 1$ when the foliage elements are randomly distributed; $\Omega < 1$ when the canopy is organized into clumps. The clumping index is an important input to global carbon and water cycle models (Chen et al., 2003).

When the sun and the observer are both at zenith, FVC is equivalent to the fraction of solar radiation intercepted by the vegetation (Smith et al, 1993). Therefore, considering $FVC = FIPAR(\theta_s = 0)$, equation (17) yields (Roujean and Lacaze, 2002)

$$\mathrm{FVC} = 1 - \exp(-b \cdot \mathrm{G}(\theta_s = 0) \cdot \Omega \cdot \mathrm{LAI}) \quad (18)$$

In equation (18), a value of 0.5 is adopted for the leaf projection factor $G(\theta_s)$ considering spherical orientation of the foliage. In order to avoid maximum LAI values in fully vegetated areas (i.e. when FVC $\rightarrow$1) exceeding a value about 6-7, a coefficient a_0 in the range (1.04-1.07) is introduced in (18):

$$\mathrm{FVC} = a_o\{1 - \exp(-0.5 \cdot b \cdot \Omega \cdot \mathrm{LAI})\} \quad (19)$$

The clumping is assumed for simplicity to be homogeneous within each vegetation cover type. A cover-dependent empirical clumping index to each of the GLC2000 classes has been adopted. The values for each biome of the GLC2000 database, correspond to the maximum values from all valid retrievals as calculated from global POLDER multiangular data for the period November 1996 to June 1997 (Chen et al. 2005) and are shown in table 2. This assumption leads to a conservative first-order correction of the clumping effect, which has been adopted to avoid a likely over-correction in erroneously classified biomes. Typical clumping values are 0.68 for evergreen forest, 0.77-0.79 for deciduous forest, and 0.83-0.85 for herbaceous, shrub and cultivated areas.

Table 2: Cover-dependent clumping index values for LAI algorithm based on the GLC2000 land cover classification based on values obtained in Chen et al. (2005).

GLC200 Class	Global Class name (according to LCCS terminology)	Clumping Index
1	Tree Cover, broadleaved, evergreen	0.68
2	Tree Cover, broadleaved, deciduous, closed	0.79
3	Tree Cover, broadleaved, deciduous, open	0.78
4	Tree Cover, needle-leaved, evergreen	0.68
5	Tree Cover, needle-leaved, deciduous	0.77
6	Tree Cover, mixed leaf type	0.79
7	Tree Cover, regularly flooded, fresh water	0.69
8	Tree Cover, regularly flooded, saline water,	0.79
9	Mosaic: Tree cover / Other natural vegetation	0.82
10	Tree Cover, burnt	0.86
11	Shrub Cover, closed-open, evergreen	0.80
12	Shrub Cover, closed-open, deciduous	0.80
13	Herbaceous Cover, closed-open	0.83
14	Sparse Herbaceous or sparse Shrub Cover	0.84
15	Regularly flooded Shrub and/or Herbaceous Cover	0.85
16	Cultivated and managed areas	0.83
17	Mosaic: Cropland / Tree Cover / Other natural vegetation	0.76
18	Mosaic: Cropland / Shrub or Grass Cover	0.81
19	Bare Areas	0.99
20	Water Bodies	---
21	Snow and Ice	---
22	Artificial surfaces and associated areas	---

The theoretical uncertainty of LAI, Err(LAI) , propagates uncertainties attached to the FVC estimate, Err(FVC), and is also associated with the uncertainties of parameters a_0, b and Ω, according to the following expression:

$$(\mathrm{Err(LAI)})^2 = \left(\frac{\mathrm{Err(FVC)}}{a_1(a_0 - \mathrm{FVC})}\right)^2 + \left(\frac{\mathrm{LAI}\cdot \mathrm{Err}(a_1)}{a_1}\right)^2 + \left(\frac{\mathrm{FVC}\cdot \mathrm{Err}(a_0)}{a_0 a_1 (a_0 - \mathrm{FVC})}\right)^2 \quad (21)$$

where

$$a_1 = b \cdot \Omega \tag{22}$$

Typical uncertainty values adopted for the model parameters are: Err(a_0)=0.03 and Err(a_1)=0.04.

2.6 FAPAR Algorithm description

For the retrieval of daily FAPAR from space data without any prior knowledge on the land cover, a statistical relationship general enough for global applications (Roujean and Bréon, 1995) is defined based on simulations using the homogeneous SAIL model (Verhoef, 1984). The FAPAR information is derived from the red and NIR spectral bands. The SAIL model provides the BRDF data as well as the amount of radiation absorbed by vegetation. Inputs of SAIL model are leaf inclination distribution (LIFD), LAI, leaf transmittance, leaf reflectance and soil spectral albedo. A large number of vegetation canopy radiative transfer scenarios are simulated. The simulation includes a total of 5184 cases, by combining three LIFD (erectophile, spherical, planophile), eight LAI values (0.1, 0.3, 0.5, 1, 1.5, 2, 3, 4), 6 soil spectra and 36 combinations of leaves optical properties in the red and NIR spectral bands, assuming lambertian soil properties and leaf reflectance equal to leaf transmittance. The diffuse fraction of incoming radiation is held constant and equal to 0.2, which represents clear sky conditions. For each scenario, red and NIR reflectances with variations of sun and view zenith angle from 0º to 75º at interval of 15º, and variations of sun-view relative azimuth from 0º (backscattering) to 180º (forward-scattering) each 45º are obtained. Finally, the daily FAPAR is computed by integration of the instantaneous FAPAR over the day:

$$FAPAR = \frac{\int_t^{t'} APAR \ dt}{\int_t^{t'} PAR \ dt} \tag{23}$$

Where t and t' are the time for sunrise and sunset. The FAPAR was integrated over solar angles corresponding to a target located at 45ºN latitude and at the equinox.

The relationship between NDVI and daily FAPAR is analysed for the different geometries. Largest dispersion occurs when the sun is at zenith and sensor at nadir (i.e., k0 BRDF parameter used for retrieving FVC and LAI). The dispersion decreases as the sun and view zenith angle increases. This could be partly explained due to the optical path through the vegetation increases with sun zenith angles, and thus the background relative effect decreases (Roujean and Bréon, 1995). Among the many geometries investigated, some cases perform significantly better with approximately linear NDVI-FAPAR relationship and small dispersion. An optimal geometry based on the criteria of linearity and minimum dispersion between NDVI and daily-integrated FAPAR is found in the solar principal plane (θ_s=45º, θ_v=60º, φ=0º). The relationship is especially good when only one canopy type is considered. However, some dispersion is found between NDVI and FAPAR for low LAI values, which is explained by the influence of the background reflectance. Therefore, a vegetation index is proposed to minimize the soil reflectance effects, called RDVI (Renormalized Difference Vegetation Index), which has the same properties as NDVI for large LAI but performs better for low vegetation coverage. The RDVI is defined as follows:

$$RDVI = (NDVI \cdot DVI)^{1/2} = \frac{NIR - R}{\sqrt{NIR + R}} \tag{24}$$

Finally, the RDVI-FAPAR relationship in the optimal geometry is given in Eq. (25). In order to apply this relationship to remotely sensed data, it is needed first to be able of characterising the BRDF in order to compute the reflectance and thus the RDVI in the optimal geometry

$$FAPAR=1.81*(RDVI)_{opt} - 0.21 \quad (25)$$

Where $(RDVI)_{opt}$ refers to the RDVI computed in the optimal geometry. The reflectance in the optimal geometry (θ_s=45º, θ_v=60º, φ=0º) for each spectral channel is estimated from the Roujean et al. (1992) as follows:

$$R_{opt}(\lambda)= k0(\lambda) - 0.240*k1(\lambda) + 0.202*k2(\lambda) \quad (26)$$

• FAPAR uncertainty estimation

The theoretical FAPAR uncertainty is assessed by mathematically differentiating Eq. (25) with respect to the theoretical input error. The error of the BRDF parameters correspond with the diagonal elements of the uncertainty covariance matrix (C00, C11 and C22 fields of the AL-C_i-CK product) obtained for the three model parameters (SAF/LAND/MF/PUM_AL/1.4). These values only quantify the uncertainties due to the non-correlated (random) part of the error structure; in addition correlated errors may occur owing to calibration uncertainties or systematic bias. Therefore, the theoretical FAPAR error can be estimated propagating input error through the model, which mainly depends in turn on the errors in the reflectance computed in the optimal geometry for red and NIR channels.

The error of the FAPAR is computed as:

$$Err(FAPAR) = 1.81 \cdot Err(RDVI) \quad (27)$$

The RDVI can be written as:

$$\begin{aligned} RDVI &= \frac{A}{B} \\ A &= R(C2) - R(C1) \\ B &= \sqrt{R(C1) + R(C2)} \end{aligned} \quad (28)$$

where R(Ci) is the reflectance in the optimal geometry in red (C1) and NIR(C2) channels. Thus, the error of the RDVI is written as:

$$Err(RDVI) = \frac{\partial(RDVI)}{\partial A} \cdot Err(A) + \frac{\partial(RDVI)}{\partial B} \cdot Err(B) = \frac{1}{B} \cdot Err(A) + \frac{A}{B^2} \cdot Err(B) \quad (29)$$

The derivative of A and B with respect to the reflectance is written as follows:

$$Err(A) = Err(R(C2)) + Err(R(C1)) \quad (30)$$

$$Err(B) = \frac{1}{2} \cdot (R(C1) + R(C2))^{-1/2} \cdot [Err(R(C2)) + Err(R(C1))] \quad (31)$$

Therefore,

$$Err(RDVI) = \frac{Err(R(C2)) + Err(R(C1))}{\sqrt{R(C1) + R(C2)}} + \frac{(R(C2) - R(C1))}{(R(C1) + R(C2))} \cdot \frac{1}{2} \cdot (R(C1) + R(C2))^{-1/2} \cdot [Err(R(C2)) + Err(R(C1))] \quad (32)$$

Which can be written as:

$$Err(RDVI) = [Err(R(C2)) + Err(R(C1))] \cdot \left[\frac{1}{\sqrt{R(C1) + R(C2)}} + \frac{1}{2} \frac{(R(C2) - R(C1))}{(R(C1) + R(C2))^{\frac{3}{2}}} \right] \quad (33)$$

Where Err(R(Ci)) in the optimal geometry used for retrieving the FAPAR is computed as:

$$Err(R) = Err(k_0) + 0.240 \cdot Err(k_1) + 0.202 \cdot Err(k_2) \quad (34)$$

And the theoretical uncertainty of the directional error is given in the C00, C11 and C22 fields of the HDF5_LSASAF_MSG_AL-C_i-CK product.

Figure 6 shows the FAPAR uncertainty computed as explained above with respect to the uncertainties of the k0 and k2 directional parameters for a fixed uncertainty of 0.01 in the k1 parameter, and for both low and high FAPAR values. The variations of the input errors are in agreement with the range of uncertainties found in the BRDF product (see SAF/LAND/UV/VR_VEGA/2.1-2 for more details). The largest uncertainties in the directional parameter are the largest the uncertainties in FAPAR are. A threshold based on the error of k2 equal to 0.25 has been used to blind unreliable FAPAR areas (with FAPAR error estimate larger than 0.2). The FAPAR error ranges typically between 0 and 0.2, with some cases showing an error estimate higher than 0.2 which are mainly located over Europe in wintertime. Pixels with errors ranging between 0.2 and 0.25 should be used with caution.

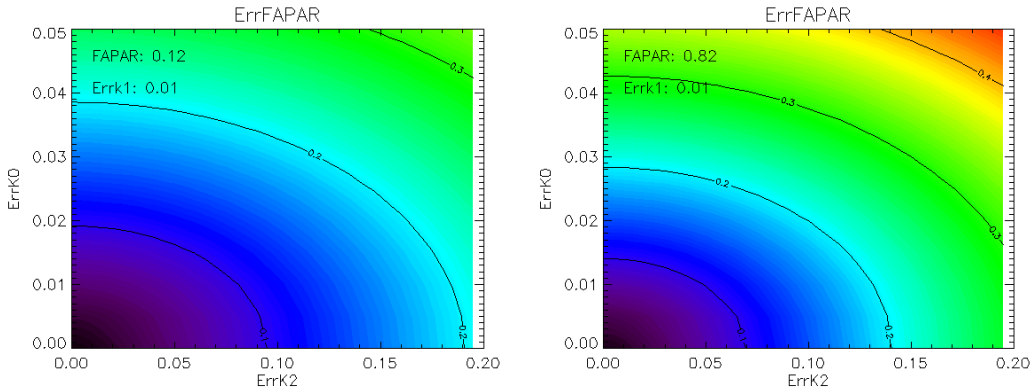


Figure 6 - Theoretical FAPAR error as a function of input k0 and k2 errors for a given k1 error of 0.01. Two different cases have been considered: Low FAPAR values (left panel) and high FAPAR values (right panel).

• Improvements regarding previous versions

FAPAR v2.1 (and later versions) introduces two main changes regarding the previous version implemented in the VEGA 2.0 code (see SAF/LAND/UV/ VR_VEGA/2.1-2 for a detailed description):

(1) FAPAR product is blinded over problematic areas with large BRDF errors. The user can identify easily these regions in the FAPAR v2.0 using the Error of FAPAR v2.0. **If ErrFAPAR v2.0 is larger than 0.4 then FAPAR v2.0 estimates are NOT reliable and should be discarded from any study.**

(2) ErrFAPAR is better estimated and more reliable than the ErrFAPAR v2.0. However, a linear relationship exists between v2.0 and v2.1 FAPAR uncertainties. Hence, the user can obtain the equivalent uncertainty of FAPAR v2.1 (and later versions) applying a factor of 0.75 to the ErrFAPAR v2.0, i.e.:

$$\textbf{ErrFAPAR = 0.75* ErrFAPAR v2.0} \qquad \textbf{(35)}$$

Applying these two conditions to the FAPAR v2.0 product (available since September 2006 to February 2008) the user can get consistent temporal series of FAPAR products, as the main algorithm does not changes. A systematic decrease in the magnitude of the FAPAR v2.0 between 2006 and 2007 observed in the south hemisphere should be explained due to changes in the AL2 code (see SAF/LAND/UV/ VR_VEGA/2.1-2)

2.7 Internal consistency between the LSA SAF Products

In order to analyse the consistency among the suite of LSA SAF vegetation products (ALBEDO), we first evaluate how the energy absorbed by the ground below vegetation can be estimated from two independent SEVIRI products, fraction of PAR absorbed by vegetation (FAPAR) and bi-hemispherical reflectance integrated over the photosynthetically active spectral region (BHRPAR).

From energy conservation, the fraction of PAR absorbed by the ground beneath the canopy is given by the formula (Diner et al., 2005):

$$FGROUND = 1 - BHRPAR - FAPAR \quad (36)$$

Figure 7 shows the relationship between MSG Daily LAI and FGROUND, as derived by combining two independent SEVIRI products (Eq. 36), for two different SEVIRI geographical regions, SAfr and Euro. The distribution can be well approximated by the exponential function, FGROUND= A·exp[−B·LAI], with A and B assumed to be constant. A similar relationship is observed for different periods of year 2007 and over different SEVIRI regions, which indicates a strong consistency between LAI and FGROUND at the resolution of SEVIRI products.

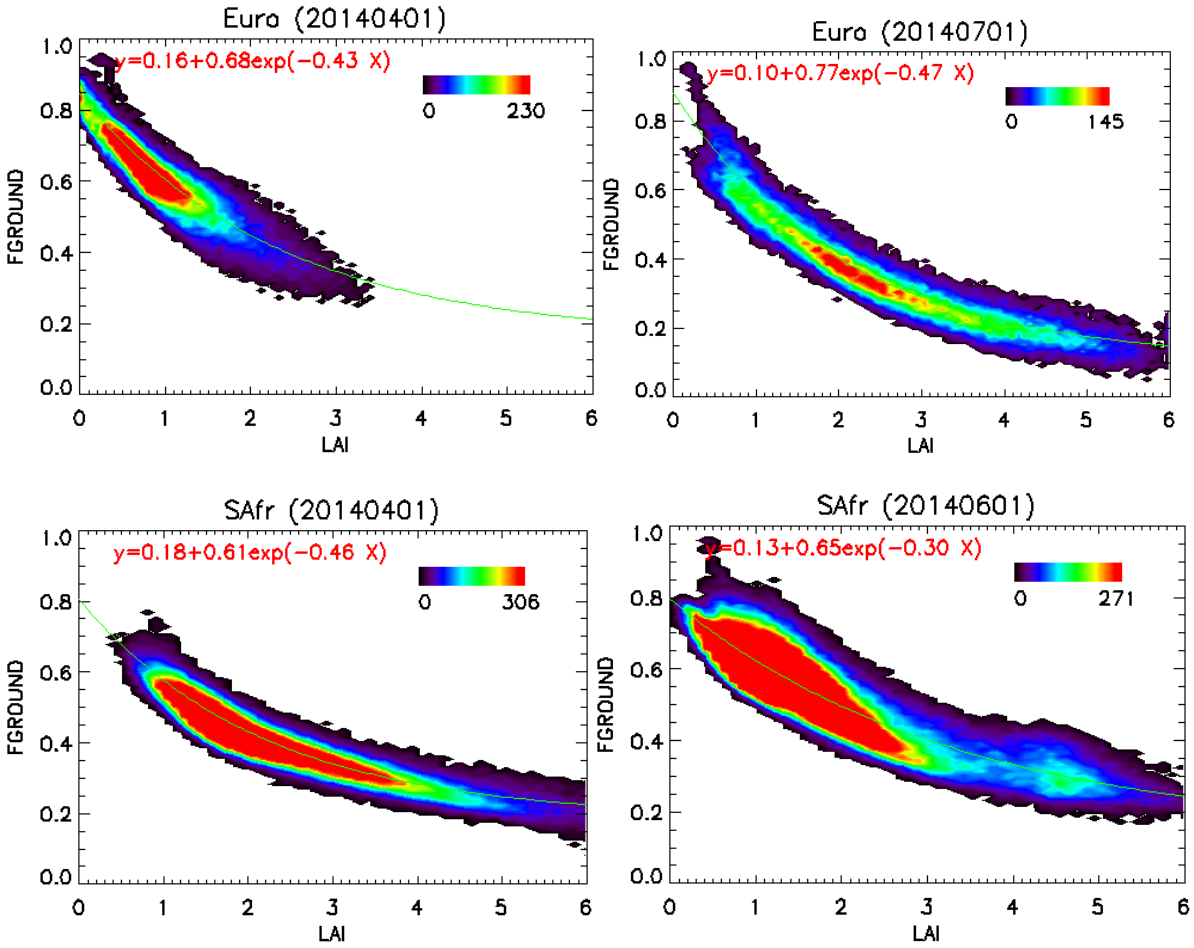


Figure 7 - Joint probability density plots between FGROUND and MSG Daily LAI at several different periods of the year. Top figures correspond to the Euro SEVIRI zone, whereas bottom figures correspond to the SAfr SEVIRI region.

These results can be interpreted as follows (Diner et al., 2005). The fraction of PAR absorbed by the ground is the downward PAR flux density, F, times the soil absorptance, 1−α, where α is the reflectance of the canopy background and F can be expressed via the Beer's law (Hu et al. 2007):

$$\mathrm{FGROUND} \approx \frac{1-\alpha}{1-\alpha\, r^{*}} \exp[-G(\theta_s)\cdot \mathrm{LAI}/\cos(\theta_s)] \quad (37)$$

where r* is the probability that photon entering through the lower canopy boundary will be reflected back by the vegetated layer (Knyazikhin & Marshak, 2000; Wang et al., 2003), θ_s is the solar zenith angle, and $G(\theta_s)$ is the average extinction function (Ross, 1981). This analysis evidenced how the use of satellite fields derived from SEVIRI instrument has enabled a realistic partitioning of incoming solar radiation between the canopy and the ground below the canopy. Further details are available in a related document (SAF/LAND/UV/VEGA_VR/2.1).

3 Product Description

The LSA SAF SEVIRI/MSG chain processes separately the whole SEVIRI disk. The projection and spatial resolution correspond to the characteristics of Level 1.5 MSG/SEVIRI instrument data. Information on geo-location and data distribution is available at the LSA SAF web-site: http://landsaf.ipma.pt.

The vegetation products (VEGA) are delivered at Daily (MDFVC, MDLAI, MDFAPAR) and ten-day (MTFVC, MTLAI, MTFAPAR) time step based on the cloud-free BRDF k0-parameter (albedo product). Comments regarding the temporal characteristics and spatial coverage of the albedo products (details are given in SAF/LAND/MF/PUM_AL/1.5) therefore also apply to the vegetation products. VEGA ten-day differs of VEGA daily products only in the BRDF input. The daily albedo product is computed using an iterative scheme with a characteristic time scale of five days. By contrast, the 10-day albedo product is currently a classic composite expanding over a 30-day period.

The v3.0 algorithm of MSG FAPAR product is identical to the one of the previous version (v2.1). However, the v3.0 version of the MSG FVC and LAI products has been developed to provide a more accurate identification of the vegetation and soil components. Therefore, the main limitation of the VEGA v2.1 is improved in the new version (v3.0), improving notably the consistency with Copernicus Global Land products based on SPOT/VEGETATION observations, mainly over sparse vegetation in southern regions in Africa and South America. Furthermore, a better filtering of pixels affecting by residual snow is applied.

An example of the LSA SAF VEGA (FVC, LAI and FAPAR) v3.0 daily products is shown in Figure 8. The outputs present practically no missing data except for areas which are usually covered by snow or frequent cloud cover. Large uncertainties are generally found in areas where the BRDF reliability is poor.

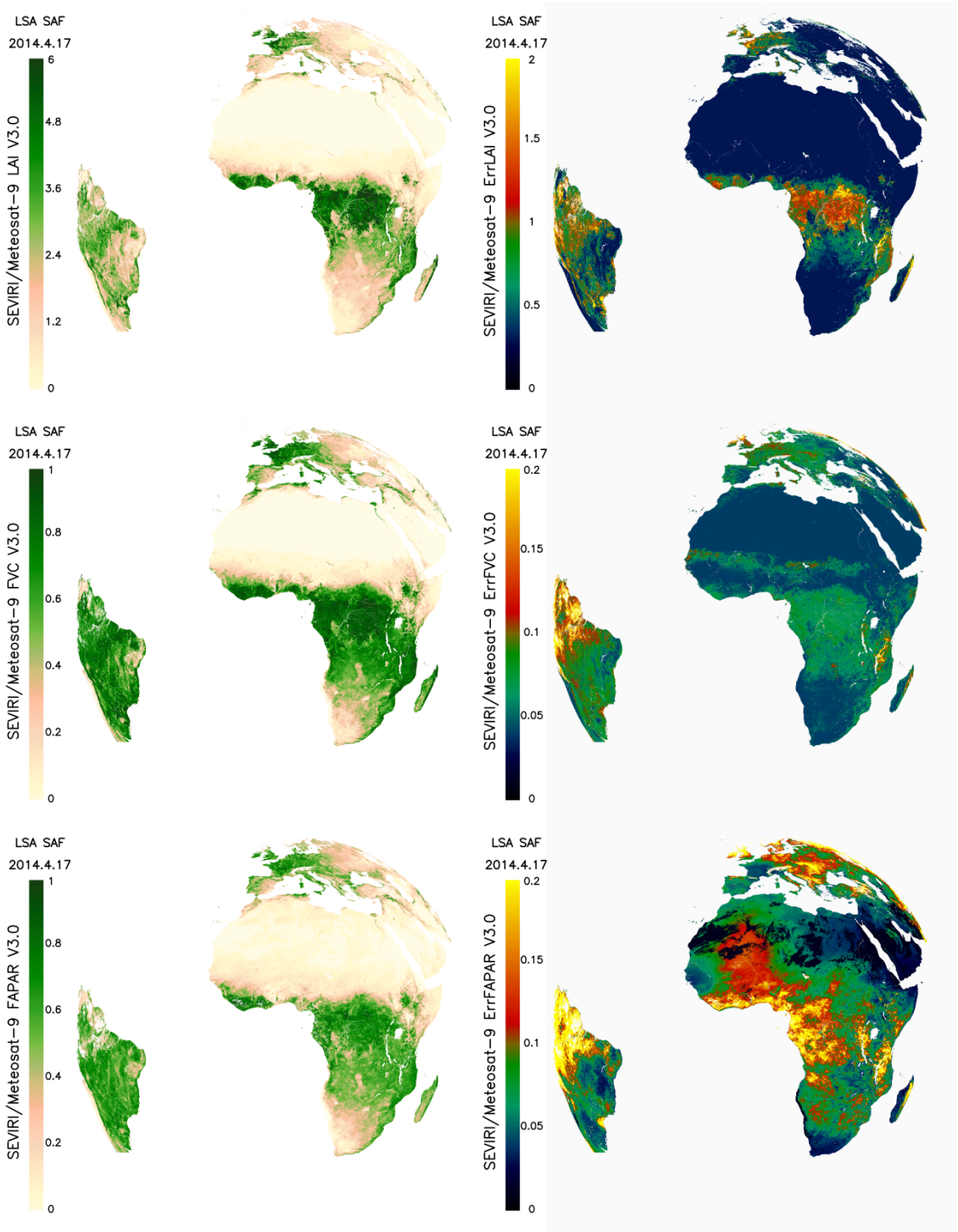


Figure 8 – MSG Daily LAI (top), FVC (middle) and FAPAR (bottom) LSA SAF VEGA (version v3.0) product composition of the four LSA SAF geographical areas corresponding to the 17th of April 2014: products (left panels) and their respective error estimates (right panels).

4 References

Asner, G. P., & Heidebrecht, K. B. (2002). Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations. International Journal of Remote Sensing, 23(19), 3939– 3958.

Bacour, C., Baret, F., Beal, D., Weiss, M., Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. Remote Sensing of Environment, 105, 313-325.

Baret, F. and G. Guyot (1991). Potentials and Limits of Vegetation Indices for LAI and APAR Assessment. Remote Sensing of Environment, 35: 161-173.

Baret, F and K. Pavageau, (2006). Validation of PARASOL land products. POSTEL / INRA. May 2006, 50 pp.

Baret, F., O. Hagolle, B. Geiger, P. Bicheron, B. Miras, M. Huc, B. Berthelot, f. Nino, M. Weiss, O. Samain, J.L. Roujean, and M. Leroy, (2007). LAI, FAPAR, and FCover CYCLOPES global products derived from Vegetation. Part 1: principles of the algorithm, Remote Sensing of Environment, 110:305-316.

Bartholomé E. et al. (2006). VGT4Africa User manual. First Edition. EUR 22344 EN. Luxembourg: Office for Official Publications of the European Communities, 268 pp.

Bateson, C. A., Asner, G. P., & Wessman, C. A. (2000). Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis. IEEE Transactions on Geoscience and Remote Sensing, 38, 1083–1094.

Bicheron P. & Leroy M., (1999), A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model. Remote Sensing of the Environment, 67, 251-266.

Bishop, C.M. (1995), Neural Networks for Pattern Recognition, Oxford:Oxford University Press.

Bonan, G. (1995). Land– atmosphere interactions for climate system models: coupling biophysical, biogeochemical, and ecosystem dynamical processes. Remote Sensing of Environment, 51, 57– 73.

Bosdogianni, P., M. Petrou, and J.Kittler. (1996) Robust mixed pixel classification in remote sensing. In *VIII ISPRS-Congress*, Vienna, Austria.

Camacho-de Coca, F., F.J. García-Haro, M.A. Gilabert and J. Meliá, (2004). Vegetation cover seasonal changes assessment from TM imagery in a semiarid landscape. International Journal of Remote Sensing, Vol. 25, No.17: 3451-3476.

Camacho-de Coca, F., F.J. García-Haro, R. Lacaze, M. Leroy, J.-L. Roujean and J. Meliá (2005). *Validation of the LSA SAF FVC and LAI prototype algorithm with POLDER/ADEOS data over Western Europe.* EUMETSAT Meteorological satellite conference, Dubrovnik, Croatia, 19-25 Sept 2005 (in press).

Camacho-de Coca, F. 2007. Evaluation of the Land-SAF FAPAR prototype along one year of MSG BRDF data: Algorithm, Product description, and inter-comparison against equivalent satellite products and ground-truth. SAF Visiting Scientist Report, 60 pp (available at http://landsaf.meteo.pt).

Chen, J., Pavlic, G., Brown, L, Cihlar, J., Leblanc, S.G., White, H.P., Hall, R.J., Peddle, D.R., King, D.J., Trofymow, J.A., Swift, E., Van der Sanden, J. & Pellikka, P.K.E. 2002. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. Remote Sensing of Environment 80: 165-184.

Chen, J. M., Liu, J., Leblanc, S. G., Lacaze, R., & Roujean, J.-L. (2003). Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. Remote Sensing of Environment, 84, 516–525.

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

Defries R. S., Hansen M. C. and Townshend J. R. G., 2000: Global continuous fields of vegetation characteristics: a linear mixture model applied to multi-year 8km AVHRR data. *Int. J. Remote Sensing*, 21(6&7), 1389-1414.

Deng, F., Chen, J. M., Plummer, S., Chen, M., and Pisek, J. (2006). Algorithm for global leaf area index retrieval using satellite imagery. *IEEE Transactions Geoscience and Remote Sensing*, 44:2025-2047.

Dickinson R.E., 1983: Land surface processes and climate – Surface albedos and energy balance, *Adv. Geophys.*, 25, 305-353.

Diner, D. J., Braswell, B. H., Davies, R., Gobron, N., Hu, J., Jin, Y., et al. (2005). The value of multiangle measurements for retrieving structurally and radiatively consistent properties of clouds, aerosols, and surfaces. Remote Sensing of Environment, 97, 495−518.

Ferranti, L. e P. Viterbo, (2006), The European Summer of 2003: Sensitivity of Soil Water Initial Conditions. *J. Climate*, 19, 3659-3680.

Fisher, J.I and J. F. Mustard, (2007), Cross-scalar satellite phenology from ground, Landsat, and MODIS data, Remote Sensing of Environment, 109, 261−273.

Fraley, C., & Raftery, A. E. (2002). Model based clustering, discriminant analysis and density estimation. Journal of the American Statistical Association, 97, 611 – 631.

Fritz, S., C. McCallum, C. Schill, R. Perger, F. Grillmayer, F. Achard, and M. Obersteiner, (2009), Geo-wiki. Org: the use of crowdsourcing to improve global land cover, Remote Sensing, 1: 345-354.

García-Haro, F.J., Gilabert, M.A. & Meliá, J., (1996), Linear spectral mixture modelling to estimate amount of vegetation from optical spectral data. *International Journal of Remote Sensing,* 17: 3373-3400.

García-Haro, F.J., F. Camacho-de Coca, J. Meliá (2004). Global mapping of vegetation parameters from SEVIRI/MSG data. *Remote Sensing for Agriculture, Ecosystem and Hydrology*, vol 5233, pp 30-41. Editorial: (SPIE: Bellingham, WA) edited by M. Owe, G. d'Urso, J. Moreno and A.Calera, ISBN 0-8194-5115-0, (Bellingham, WA).

García-Haro, F.J, S. Sommer, T. Kemper (2005a), Variable multiple endmember spectral mixture analysis (VMESMA), *International Journal of Remote Sensing*, 26:2135-2162.

García-Haro, F.J., F. Camacho-de Coca, J. Meliá, B. Martínez, Operational derivation of vegetation products in the framework of the LSA SAF project, (2005b), in Proceedings of 2005 EUMETSAT Meteorological Satellite Conference. Dubrovnik (Croatia). 19-23 September, (Eumetsat Publ.: Darmstad), ISBN 92-9110-073-0, ISSN 1011-3932, pp 247-254.

García-Haro, J, F. Camacho-de Coca y J. Meliá, (2006). A directional Spectral Mixture Analysis. IEEE Transactions on Geoscience and Remote Sensing, 44: 365-377.

Gilabert, M.A., García-Haro, F.J. & Meliá, J. (2000). A mixture modeling approach for estimate vegetation parameters in remote sensing. *Remote Sensing of Environment*, 72: 328-345.

Garrigues, S., D. Allard, F. Baret, 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 (2006) 286–298

Geiger, B. L. Franchisteguy, and J.L. Roujean, (2004), Land Surface Albedo from Meteosat Second Generation (MSG) Observations, IGARS, in press.

Geiger, B., Carrer, D., Franchistéguy, L., Roujean, J.L., Meurey, C., 2006, Product User Manual (PUM) Land Surface Albedo V1.4, SAF/LAND/MF/PUM_AL/1.4, [Online]. Available: http://landsaf.meteo.pt

Gobron N., Mélin F., Pinty B., Verstraete M. M., Widlowski J.-L. and Bucini G. (2001) 'A Global Vegetation Index for SeaWiFS: Design and Aplications', in Remote Sensing and Climate Modeling: Synergies and Limitations, Edited by M. Beniston and Verstraete M. M., Kluwer Academic Publishers, Dordrecht, 5-21.

Gobron N., B. Pinty, M. Verstraete, and Y. Govaerts (1999). The MERIS Global Vegetation Index (MGVI): description and preliminary application. *Int. J. Rem. Sensing*, 20:1917-1927.

Gobron, N., O. Aussedat, B. Pinty, M. Robustelli, M. Taberner and T. Lavergne (2006b). Thecnical Assistance for the validation of the ENVISAT MGVI geophysical product. EUR 22246 EN. European Communities, 2006.

Hu, B., Miller, J., Chen, J. & Hollinger, A. 2004. Retrieval of the Canopy LAI in the BOREAS Flux Tower Sites Using Linear Spectral Mixture Analysis. Remote Sensing of Environment 89: 176-188.

Hu, J. Y. Su, B. Tan, D. Huang, W. Yang, M. Schull, M.A. Bull, J. V. Martonchik, D. J. Diner, Y. Knyazikhin, R. B. Myneni, 2007, Analysis of the MISR LAI/FPAR product for spatial and temporal coverage, accuracy and consistency, Remote Sensing of Environment 107: 334–347

Huete, A.R. (1988). A Soil Adjusted Vegetation Index (SAVI). Remote Sens Env. 25:295-309.

Knyazikhin, Y., J. Glassy, J.L. Privette, Y. Tian, A. Lotsch, Y. Zhang, Y. Wang, J.T. Morisette, P. Votava, R.B. Myneni, R.R. Nemani, and S.W. Running, (1999), MODIS leaf area index (LAI) and fraction of potosynthetically active radiation aborbed by vegetation (FPAR), Product (MOD15) ATBD.

Knyazikhin, Y., & Marshak, A. (2000).Mathematical aspects of BRDF modeling: Adjoint problem and Green's function. Remote Sensing Reviews, 18, 263−280.

Lacaze, R, P. Richaume, O. Hautecoeur, T. Lalanne, A. Quesney, F. Maignan, P. Bicheron, M. Leroy, et F. M. Bréon, Advanced algorithms of the ADEOS2/POLDER2 land surface processing line: application to the ADEOS1/POLDER1 data, *Proceedings of the XXIIIth IGARSS Symposium, Toulouse, France,* July 2003.

Leroy, M., J.L. Deuzé, F.M. Bréon, O. Hautecoeur, M. Herman, J.C. Buriez, D. Tanré, S. Boufiès, P. Chazette et J.L. Roujean, Retrieval of atmospheric properties and surface bi-directional reflectances over the land from POLDER/ADEOS, *Journal of Geophysical Research, 102, 17,023-17,037*, 1997.

Liu, J., Chen, J., Cihlar, J., & Park, W. (1997). A process-based boreal ecosystem productivity simulator using remote sensing inputs. Remote Sensing of Environment, 62, 158– 175.

LSA SAF Product User Manual (PUM) of Land Surface Analysis Land Surface Albedo (SAF/LAND/MF/PUM_AL/1.4), December 2006, 48 pp.

LSA SAF Product User Manual (PUM) of Land Surface Analysis Vegetation products (SAF/LAND/UV/PUM_VEGA/2.0), February 2007, 46 pp.

Malingreau, J. P., and A. S. Belward. (1992). Scale considerations in vegetation monitoring using AVHRR data. International Journal of Remote Sensing. 13:2289-2307.

Martínez, B., F. Camacho-de Coca, F.J. García-Haro and A. Verger, 2005, *Early validation of MSG vegetation products over the Iberian Peninsula*. EUMETSAT Meteorological Conference. Dubrovnik. Croatia, 19-25 September 2005.

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

McLachlan, G. J., & Krishnan, T. (1997). The EM algorithm and extensions. New York: Wiley. McLachlan, G. J., & Peel, D. (2001). Finite mixture models. New York: Wiley.

Mitchell, K., et al., 2004: The multi-institution North American Land Data Assimilation System NLDAS: Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, J. Geophys. Res., 109, doi:10.1029/2003JD003823.

Myneni, R., Keeling, C., Tucker, C., Asrar, G., & Nemani, R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. Letters to nature. Nature, 386.

Nilson, T., A theoretical analysis of the frequency of gaps in plant stands, Agriculture and Meteorology, 8, 25-38, 1971.

Olthof, I., King, D., & Lautenschlager, R. (2003). Overstory and understory leaf area index as indicators of forest response to ice storm damage. Ecological Indicators, 3, 49– 64.

Peddle, D. Brunke, S. and Hall, F. 2001. A Comparison of Spectral Mixture Analysis and Ten Spectral Vegetation Indices for Estimating Boreal Forest Biophysical Information from Airborne Data. Canadian Journal of Remote Sensing 27 6:627-635.

Petrou, M. and Foschi, P.G., "Confidence in linear spectral unmixing of single pixels", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, issue 1, pp. 624-626, Jan. 1999.

Plummer, S., Arino, O., Simon, M., and Steffen, W. (2006). Establishing an earth observation product service for the terrestrial carbon community: the globcarbon initiative. *Mitigation and Adaptation Strategies for Global Change*, 11:97-111.

Ricard, M., Neuenschwander, A., Crawford, M., and Gibeaut, J.C., "Multisensor classification of wetland environments using airborne multispectral and SAR data," *Proc. 1997 International Geoscience and Remote Sensing Symposium,*Singapore, August 10-14, 1997.

Roberts, D. A., Gardner, M., & Church, R. (1998). Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. Remote Sens. Environ., 65, 267–279.

Ross, J.K. "The radiation regime and architecture of plants stands", Norwell, MA: Dr. W. Junk, 391 pp., 1981.

Roujean J.-L., M. Leroy and P.-Y. Deschamps, (1992), *A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data*, J. Geophys. Res., 97(D18), 20455-20468.

Roujean, J.-L. and F.-M. Bréon, (1995). Estimating PAR absorbed by vegetation from Bidirectional Reflectance Measurements. Remote Sensing of Environment, 51: 375-384.

Roujean, J. L. (1996), A tractable physical model of shortwave radiation interception by vegetation canopies,*J. Geophys. Res.*, vol. 101, pp. 9523–9532.

Roujean, J. L., Tanré, D., Bréon, F-M. and Deuzé, J-L. (1997) Retrieval of land surface parameters from airborne POLDER bi-directional reflectance distribution function during HAPEX-Sahel. Journal of Geophysical Research 102(D10), 11 201-11 218.

Roujean, J.L. and R. Lacaze, (2002). Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: A pragmatic method and its validation. J. Geophysical Res., 107D, 10129-10145.

Running, S. and Coughlan, J. (1988). A General Model of Forest Ecosystem Processes for Regional Applications. I. Hydrologic Balance, Canopy Gas Exchange and Primary Production Processes. Ecological Modeling 42: 125-154.

SAF for Land Surface Analysis (LSA SAF) Validation Report (SAF/LAND/IM/VR/1.5), January 2006, 128 pp.

SAF for Land Surface Analysis (LSA SAF) Validation Report (SAF/LAND/IM/VR/1.5), January 2006, 128 pp.

Schmetz, J., P. Pili, S. Tjemkes, D. Just, J. Kerkman, S. Rota, and A. Ratier (2002), An introduction to Meteosat Second Generation (MSG), *Bull. Amer. Meteor. Soc.*, *83*, 977-992.

Scott, D. W. 1992, Multivariate Density Estimation, New York: John Wiley & Sons.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher, 1997: BOREAS in 1997: Experiment overview, scientific results, and future directions. *J. Geophys. Res.,* 102, 28 731–28 769.

Song, C., 2005, Spectral mixture analysis for subpixel vegetation fractions in the urban environment: How to incorporate endmember variability?, Remote Sens. Environ. 95: 248–263.

Validation Report of Land Surface Analysis Vegetation products, (SAF/LAND/UV/VEGA_VR/2.0), January 2007, 66 pp.

Validation Report of Land Surface Analysis Vegetation products, (SAF/LAND/UV/VEGA_VR/2.1), 31 January 2007, 91 pp.

Verger, A., Camacho-de Coca, F., & Meliá, J. (2006), Inter-comparison of algorithms for retrieving operationally vegetation parameters at global scale: assessment over Europe along 2003. *Proceedings of the 2nd Recent Advances in Quantitative Remote Sensing*, Torrent, Sept. 2006 (Publ. Univ. València: València), Ed. J. Sobrino, ISBN: 84-3706554-X, 978-84-370-633-5, pp: 909-914.

Verger, A., Camacho-de Coca, F., García-Haro, J. and Meliá, J. (2007). Direct validation of FVC, LAI and FAPAR VEGETATION/SPOT derived products using LSA SAF methodology. In *Proceedings of IEEE International Geoscience and Remote Sensing symposium, 23-27 July 2007, Barcelona (Spain).*

Verger, A., F. Camacho, F. J. García-Haro y J. Meliá (2009). Prototyping of Land-SAF leaf area index algorithm with VEGETATION and MODIS data over Europe. Remote Sensing of Environment, doi:10.1016/j.rse.2009.06.009.

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

Wang, Y., Buermann,W., Stenberg, P., Voipio, P., Smolander, H., Häme, T., et al. (2003). A new parameterization of canopy spectral response to incident solar radiation: Case study with hyperspectral data from pine dominant forest. Remote Sensing of Environment, 85, 304−315.

Weiss, M., F. Baret, R. Myneni, A. Pragnère, and Y. Knyazikhin. 2000. Investigation of a model inversion technique for the estimation of crop characteristics from spectral and directional reflectance data. Agronomie 20:3-22.

Appendix A. Developers

The development and implementation have been carried out under the responsibility of the Universitat de València.

Authors: F. Javier García-Haro and Fernando Camacho

Appendix B. Glossary

AL:	Land Surface Albedo Product
AVHRR:	Advanced Very High Resolution Radiometer
BRDF:	Bi-directional Reflectance Distribution Function
CYCLOPES:	Carbon Cycle and Change in Land Observational Products from an Ensemble of Satellites
ECMWF:	European Centre for Medium-Range Weather Forecast
EPS:	EUMETSAT Polar System
EUMETSAT:	European Meteorological Satellite Organisation
FVC:	Fractional Vegetation Cover
GLC:	Global Land Cover
HDF:	Hierarchical Data Format
LAI	Leaf Area Index
NIR:	Near Infrared Radiation
MIR:	Middle Infrared Radiation
METEOSAT:	Geostationary Meteorological Satellite
METOP:	Meteorological Operational polar satellites of EUMETSAT
MERIS:	MEdium Resolution Image Spectrometer Instrument
MISR:	Multi-Angle Imaging Spectra-Radiometer
MF:	Météo-France
MODIS:	Moderate-Resolution Imaging Spectro-Radiometer
MSG:	Meteosat Second Generation
MTR:	Mid Term Review
NOAA:	National Oceanic and Atmospheric Administration (USA)
NWP:	Numerical Weather Prediction
POLDER:	POLarization and Directionality of Earth Reflectance
PSF:	Point Spread Function
SAF:	Satellite Application Facility
SEVIRI:	Spinning Enhanced Visible and Infrared Imager
SMA:	Spectral Mixture Analysis
SPOT:	Système Probatoire d'Observation de la Terre
SVPD:	Scientific Validation Plan Document
TOC:	Top of Canopy
TOA:	Top of Atmosphere
URD:	User Requirements Document