Remote Sensing of Environment xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Principles of development and production

F. Baret [a,*], M. Weiss [a], R. Lacaze [b], F. Camacho [c], H. Makhmara [d], P. Pacholcyzk [d], B. Smets [e]

[a] *INRA-EMMAH UMR 1114, Avignon, France*
[b] *HYGEOS, Toulouse, France*
[c] *EOLAB, Valencia, Spain*
[d] *CNES, Toulouse, France*
[e] *VITO, Mol, Belgium*

ARTICLE INFO

Article history:
Received 19 October 2011
Received in revised form 27 December 2012
Accepted 29 December 2012
Available online xxxx

Keywords:
Data fusion
FAPAR
FCOVER
Global products
LAI
MODIS
Uncertainty
VEGETATION

ABSTRACT

Essential climate variables such as *LAI* or *FAPAR* are required for the monitoring, understanding and modeling of land surfaces at the global scale. While several products were already developed from the current medium resolution sensors, the few validation exercises currently achieved highlighted significant discrepancies and inconsistencies. The objective of this study is to develop improved global estimates of *LAI*, *FAPAR* and *FCOVER* variables by capitalizing on the development and validation of already existing products. In a first step, the performances of the MODIS, CYCLOPES, GLOBCARBON and JRC-FAPAR products were reviewed. The MODIS and CYCLOPES products were then selected since they provide higher level of consistency. These products were fused to generate the improved *LAI*, *FAPAR* and *FCOVER* values that were later scaled to closely match their expected range of variation. Finally, neural networks were trained to estimate these fused and scaled products from SPOT-VEGETATION top of canopy directionally normalized reflectance values. The resulting GEOV1 products are associated to quality control flags as well as quantitative estimates of uncertainties. Performances of the GEOV1 products are finally evaluated in a companion paper. The GEOV1 products are freely available to the community at www.geoland2.eu from 1999 up to present, globally at 1/112° spatial sampling grid at the dekadal time step.

1. Introduction

The importance of continuously monitoring the Earth's surface was recently recognized by Global Climate Observing System (GCOS) (GCOS, 2006). Essential climate variables (ECVs) related to land surfaces such as *LAI* (leaf area index) and *FAPAR* (fraction of absorbed photosynthetic active radiation) may be derived from observations in the reflective solar domain. These vegetation biophysical variables play a key role in several surface processes, including photosynthesis, respiration and transpiration. *LAI* is defined as half the total developed area of green elements per unit horizontal ground area (Chen and Black, 1992). *FAPAR* is defined as the fraction of radiation absorbed by the green vegetation elements in the 400–700 nm spectral domain under specified illumination conditions. *FAPAR* is one of the main inputs in light use efficiency models (McCallum et al., 2009). In addition to *LAI* and *FAPAR* variables, *FCOVER*, the fraction of green vegetation as seen from nadir, is requested by some users for vegetation monitoring (Lacaze et al., 2009) as well as for partitioning contributions between soil and vegetation within specific models for numerical weather prediction, regional and global climate modeling, and global change monitoring (Avissar and Pielke, 1989). *FCOVER* is independent from the illumination conditions as opposed to *FAPAR* while showing sensitivity to vegetation amount intermediate between *FAPAR* and *LAI*.

Few global *LAI*, *FAPAR* and *FCOVER* products have already been generated from VEGETATION, SEAWIFS, MODIS and MERIS sensors with a spatial sampling distance close to 1 km. Improved atmospheric correction, radiometric calibration and model formulation have incrementally enhanced the retrieval accuracy of the successive reprocessing. Recent validation activities have shown however that significant discrepancies were existing between these global products as well as with ground measurements (Garrigues et al., 2008; McCallum et al., 2010; Weiss et al., 2007), calling thus for the development of new products that would reconcile these differences.

The FP7 Geoland2 project (http://www.gmes-geoland.info) intends to implement the Global Monitoring for Environment and Security (GMES) Land Monitoring Core Service that corresponds to the European contribution to Group of Earth Observation System of Systems (GEOSS). An operational system is developed to provide biophysical products that meet the users' needs (Lacaze et al., 2009) for monitoring natural ecosystems and managed lands. The main requirements correspond to

* Corresponding author. Tel.: +33 432722363.
E-mail address: baret@avignon.inra.fr (F. Baret).

0034-4257/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2012.12.027

1 km spatial sampling, 10 days (dekadal) frequency with products generated in near real time (less than 1 week), for a time period as long as possible (Ganguly et al., 2008b), and associated with quantitative uncertainties. These requirements are partly answering the recently updated GCOS ones for *LAI* and *FAPAR* ECVs with 250 m spatial resolution relaxed to 2 km for near term products, 2-weekly temporal frequency and accuracy better than max(20%, 0.5) for *LAI*, max(10%, 0.05) for *FAPAR* and a stability better than max(10%, 0.25) for *LAI* and max(3%, 0.02) for *FAPAR* (GCOS-138). The resulting products will thus eventually contribute to fulfill one of the GCOS tasks dedicated to the operationalization of the generation of *LAI* ECVs.

The objective of this paper is to describe the first version of Geoland2 *LAI*, *FAPAR* and *FCOVER* products called GEOV1. The principles used to derive the products will first be presented. Then, the algorithm development will be described. A companion paper (Camacho et al., in press) will finally report the validation results derived according to the CEOS/LPV guidelines (Baret et al., 2009; Morisette et al., 2006).

2. Algorithm development

2.1. General principles

Since significant efforts have already been made to develop and validate biophysical products as stated in the Introduction, it is thus proposed to capitalize on the existing products and associated validation results for the development of the GEOV1 products. Previous works (Verger et al., 2008) have demonstrated the capacity of learning machines such as neural networks to estimate biophysical products including *LAI* from several sensors. This property will be exploited here, allowing to use several sensors in order to build a long time series of products. The algorithm is made of three main steps as sketched in Fig. 1:

(1) Generation of the training dataset: Already existing products are first combined to provide the 'best estimates' of the biophysical variables that will constitute the training dataset.
(2) Neural network calibration: A neural network is trained to estimate these 'best estimates' from the input reflectance values as observed by specific sensors and the associated geometrical configuration. Quality flags and quantitative uncertainties are also derived.
(3) Application of the network: Once the network is calibrated, it is run to provide estimates of the biophysical variables for each of the sensors considered, along with the quality flags and quantitative uncertainties.

Note that it would have been possible to follow more formally the scheme proposed by Verger et al. (2008) and later developed in Verger et al. (2011). However this would need to use concurrently and in real time two (or more) sensors. This was not compatible with the available processing capacity for GEOV1. Further, the use of a single product in the learning database as proposed in Verger et al. (2011) such as MODIS collection 5 would not allow improvement of the biases sometimes observed, but would mainly decrease the frequency of missing data and smooth the temporal series. Finally, the proposed GEOV1 algorithm is designed to work with no prior information on the vegetation type since Verger et al. (2008) showed that a single training across all biomes was performing similarly as multiple specific training for each biome type.

2.2. Generation of training dataset

The way the training dataset is generated from already existing products is sketched in Fig. 1, top box. Four main steps are identified: (1) selecting the most relevant products, (2) setting the products on consistent spatial and temporal supports, (3) fusing the products and (4) eventually scaling the fused products. Details of each of these steps are given in the following.

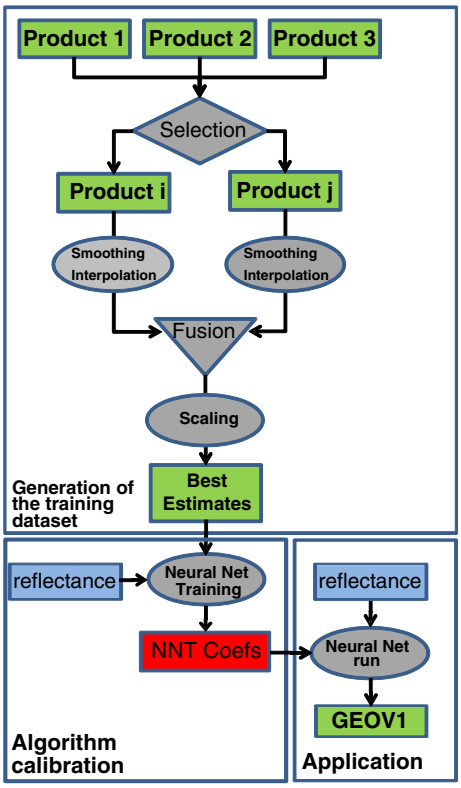

Fig. 1. Schematic description of the principle used to develop the GEOV1 product.

2.2.1. Selection of products

The available candidate global products listed in Table 1 are first evaluated to select the most relevant ones. For this purpose, the results from previous validation exercises are used here.

The validation exercise achieved by Garrigues et al. (2008) for *LAI* products was showing few missing values for the GLOBCARBON *LAI* product, in relation to its monthly temporal resolution. However, many artifacts were observed, including unexpected outliers and shifts in the phenology due to its coarse temporal resolution and low sampling frequency. Garrigues et al. (2008) and Weiss et al. (2007) demonstrated that CYCLOPES was providing very smooth temporal course while showing a saturation for *LAI* values larger than 4. For MODIS collection 5 *LAI* products, only few local validation activities were reported (De Kauwe et al., 2011; Kraus, 2008; Sprintsin et al., 2009) while Yuan et al. (2011) presented some results of a global validation exercise focusing on the evaluation of a smoothed version of MODIS *LAI* to reduce the spatial and temporal inconsistencies observed at the local spatial or temporal scales. Ganguly et al. (2008a) validated an adaptation of MODIS *LAI* products for AVHRR (Advanced Very High Resolution Radiometer) data, showing fair consistency with CYCLOPES *LAI* product over the sites considered. However, Verger et al. (2011) demonstrated that MODIS was showing more shaky temporal *LAI* profiles as compared to the CYCLOPES ones.

Among the global *FAPAR* products available, MODIS and GLOBCARBON correspond roughly to the same definition, i.e. black-sky values at the time of the satellite overpass, i.e. around 10:35 for MODIS aboard Terra and 10:30 for VEGETATION. However, if JRC-FAPAR corresponds also to black-sky *FAPAR* at the time of the satellite overpass, SEAWIFS is crossing the equator around 12:20, i.e. for significantly smaller sun zenith angles as compared to MODIS and VEGETATION. The CYCLOPES *FAPAR* product

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

Table 1
The currently *LAI*, *FAPAR* and *FCOVER* products available globally at approximately 1 km spatial sampling distance.

Products	Sensors	*LAI*	*FAPAR*	*FCOVERR*	Spatial sampling distance at equator	Temporal sampling (days)	Time period	Reference
MODIS C5	MODIS	✓	✓		1 km	8	2000–present	(Yang et al., 2006b)
CYCLOPES V3.1	VEGETATION	✓	✓	✓	1 km	10	1999–2007	(Baret et al., 2007)
GLOBCARBON	VEGETATION	✓	✓		1 km	30	1999–2007	(Deng et al., 2006)
JRC-FAPAR	SEAWIFS[a]		✓		2.17 km	1	1997–2006	(Gobron et al., 2006)

[a] JRC-FAPAR may be derived from several sensors including SEAWIFS, MERIS, MODIS and VEGETATION but produced globally over a long time period only from SEAWIFS.

corresponds to black-sky *FAPAR* at 10:00 illumination conditions which is a close approximation of the daily integrated black-sky *FAPAR* value (Baret et al., 2005, 2007). Therefore, MODIS aboard Terra, GLOBCARBON and CYCLOPES *FAPAR* products share very similar definitions, while JRC-FAPAR derived from SEAWIFS is expected to show lower values because of the smaller sun zenith angles experienced close to solar noon. Further, it is limited to sun zenith angles lower than 50° which may pose a problem for the higher latitudes and/or part of the season. Comparison between MODIS collection 5 and CYCLOPES *FAPAR* products was showing a good agreement, with however some overestimation of MODIS for the low *FAPAR* values (McCallum et al., 2010). GLOCARBON and JRC-FAPAR were found much lower in magnitude, while a good consistency in spatial and temporal trends was found between CYCLOPES and JRC-FAPAR (McCallum et al., 2010). GLOBCARBON products were showing the largest discrepancies with all the other products.

The MODIS and CYCLOPES *FAPAR* products were therefore selected ensuring a good consistency between *LAI* and *FAPAR* values since both MODIS and CYCLOPES provide concurrently *LAI* and *FAPAR* products. However, CYCLOPES and MODIS products are based on different assumptions and inversion techniques. CYCLOPES considers canopies as a turbid medium for all the biome types, while allowing pixels to be a mixture of pure bare soil and pure vegetation patches, i.e. including some possible clumping at the landscape level. The inversion of the radiative transfer model is achieved using a neural network. Conversely, MODIS algorithm is biome dependent and assumes some clumping at the plant level for some biomes including savanna and forests. The retrieval of *LAI* is achieved using a look-up-table inversion technique.

Apart from CYCLOPES *FCOVER* products, no other global *FCOVER* product is currently available apart from the SAF-LAND products covering the METEOSAT disk (Camacho-de Coca et al., 2006). However, several studies have pointed out that *NDVI* (Normalized Difference Vegetation Index) could be a good proxy for *FCOVER* (Baret et al., 1995; Carlson and Ripley, 1997; Gutman, 1991). Camacho-de Coca et al. (2006) compared several regional *FCOVER* products over Africa and showed that the CYCLOPES *FCOVER* product was very consistent with other products although a significant and systematic bias was observed. This was also confirmed by Verger et al. (2009). It is therefore proposed to select the CYCLOPES *FCOVER* original product while rescaling it to provide values more consistent with ground measurements as detailed later.

2.2.2. Spatial and temporal sampling for the training dataset

The training dataset is generated over a sample of sites representative of the global distribution of vegetation types and conditions. For this purpose, the BELMANIP2 set of sites was used. It corresponds to 420 sites located in relatively flat and homogeneous areas (at a kilometric resolution over 10 × 10 km^2 domains). BELMANIP2 differs from the original BELMANIP set of sites (Baret et al., 2006) by improving the global representativeness and homogeneity. Fig. 2 shows the distribution of sites and their coordinates are available at WWW1. Years 2003 and 2004 were selected to represent the whole seasonality as well as significant inter-annual variability.

The same spatial and temporal supports are required to allow combining several products together. The MODIS *LAI* and *FAPAR* products selected were thus re-projected onto the cylindrical projection system with 1/112° sampling grid used as a reference for the VEGETATION, CYCLOPES and GEOV1 products. Because of the point spread function of the several products considered as well as the possible geometrical uncertainties on pixel localization, a 3 × 3 pixels spatial support was used. Note that the algorithm is trained over 3 × 3 pixels although it will ultimately apply to single pixels. This scale discrepancy is acceptable because of the homogeneity of the BELMANIP2 sites.

The temporal sampling used to fuse MODIS and CYCLOPES products will be that of the CYCLOPES original products, i.e. dekadal (10 days). It will allow using directly the normalized reflectance values derived from VEGETATION based on the CYCLOPES preprocessing algorithm (Baret et al., 2007) that will also constitute the GEOV1 temporal sampling. The values to be fused at a given GEOV1 dekad were computed as follows: For CYCLOPES, the product values

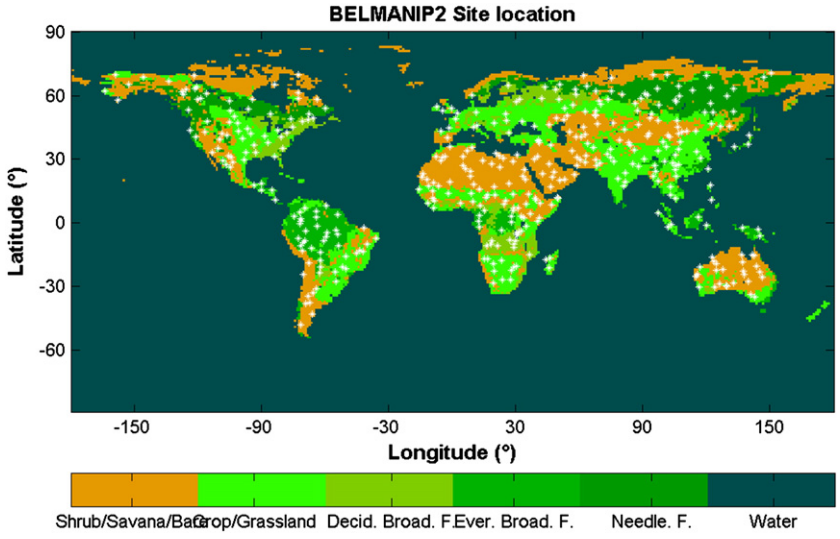

Fig. 2. The 420 BELMANIP2 sites used to sample vegetation types and conditions.

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

corresponding to the GEOV1 dekad are considered; for MODIS 8 days products, all the values available within ±10 days around the GEOV1 dekad are first gathered resulting in a maximum of 27 values (3 MODIS dates times 9 pixels). Note that the CYCLOPES 30 days temporal resolution (with Gaussian weighing) is still larger than that of the 16 days (2 times 8 days) of the MODIS selected values. However, widening the temporal window for MODIS resulted in artifacts on the seasonality. Then the 3 × 3 aggregated values for each GEOV1 dekad are computed only if at least 5 valid individual values (over the 9 for CYCLOPES and 27 for MODIS) are available. Valid pixels are defined by the main and main + saturation quality control (QC) for MODIS, and valid input reflectance QC for CYCLOPES. For MODIS, because of the relatively large variability observed over time and space, values were further filtered using the difference $\Delta = LAI_{f=90\%} - LAI_{f=70\%}$ between the $f = 90\%$ and $f = 70\%$ percentiles (f) computed over the valid *LAI* values: When $\Delta > 0.2$ corresponding to an unexpectedly large scattering of the data over homogeneous sites and during a short time period, the cases were rejected. Finally, for the GEOV1 dates fulfilling the above criterions both for CYCLOPES and MODIS products, the *LAI* and *FAPAR* values corresponding to the 70% percentile were computed for CYCLOPES and MODIS. This allows minimizing the influence of possible residual cloud contamination and atmospheric effects that negatively biased the product values (Chen et al., 2006). However, because of the homogeneity of the sites and the short time period considered, the values once filtered as described earlier, should be closely distributed around the median, i.e. *LAI* or *FAPAR* values at 50% and 70% frequencies should be very close together for a given date and site.

2.2.3. Fusing the products

No fusion process was applied for *FCOVER* since GEOV1 will derive only from CYCLOPES *FCOVER* products. Conversely, for *LAI* and *FAPAR*, MODIS and CYCLOPES products were fused to benefit from their complementarities. An optimal fusion scheme would be a weighted average between the two products, with weights, w_{FAPAR} and w_{LAI} respectively for *FAPAR* and *LAI*, being driven by the uncertainties associated to each product:

$$\begin{cases} FAPAR_{fused} = FAPAR_{MOD} w_{FAPAR} + FAPAR_{CYC}(1-w_{FAPAR}) \\ LAI_{fused} = LAI_{MOD} w_{LAI} + LAI_{CYC}(1-w_{LAI}) \end{cases} \quad (1)$$

where the subscripts *fused*, *MOD*, and *CYC* correspond respectively to the fused, MODIS and CYCLOPES products. However, the uncertainties attached to the CYCLOPES and MODIS products only refer to the theoretical performances, and model assumptions as well as the structure of the uncertainties were not accounted for. Uncertainties may be also derived from the comparison with ground measurements as already achieved for MODIS and CYCLOPES. However, these ground measurements are not very numerous (Camacho et al., in press; Garrigues et al., 2008). Further, it is not advisable to use the validation data to calibrate an algorithm in order to preserve the required independency between the calibration and the validation processes. For these reasons, the weight used in the fusion between MODIS and CYCLOPES was based on heuristic arguments.

Garrigues et al. (2008) and Weiss et al. (2007) reported that CYCLOPES *LAI* was showing some saturation for *LAI* values around 4. Conversely, MODIS *LAI* and *FAPAR* values were generally higher than expected for the very low vegetation amounts (Fig. 4). Further, the MODIS algorithm assigns zero values for *LAI* and *FAPAR* over pixels classified as bare, which may pose problems in case of misclassification. It was thus proposed to fuse the products by reducing the contribution of MODIS products for low *LAI* and *FAPAR* values and enhancing the MODIS contribution for the larger *LAI* and *FAPAR* values as sketched in Fig. 3 with $w_{LAI} = w_{FAPAR} = \min\left(1, \frac{LAI_{CYC}}{4}\right)$. The weight, w, is driven by LAI_{CYC} (Fig. 3) since LAI_{CYC} appears more stable as compared to

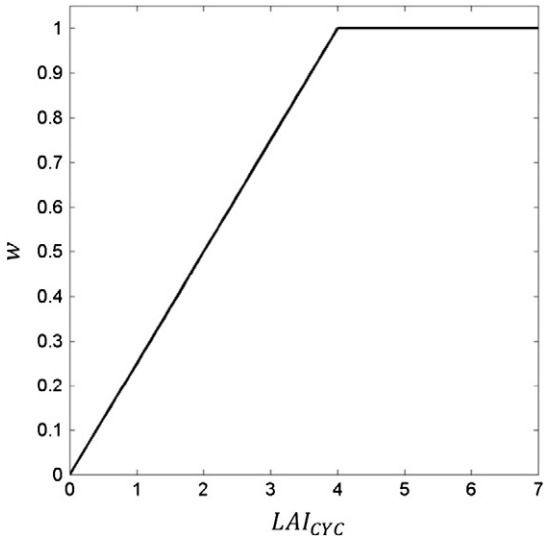

Fig. 3. Variation of the weight (w) as a function of the LAI_{CYC} value.

MODIS *LAI* (Verger et al., 2011). The threshold of $LAI_{CYC} = 4$ corresponds to the value when LAI_{CYC} starts to saturate. The parallel processing applied to both *LAI* and *FAPAR* (Eq. 1) is expected to keep a good consistency between these two variables.

2.2.4. Scaling the fused products

The fused *FAPAR* products showed that the maximum values (at 99% percentile) are around 0.90 (Fig. 4b, dashed black line) although the maximum values are expected to be close to 0.94 (Baret and Guyot, 1991) corresponding to full cover dense vegetation with albedo in the PAR domain close to 0.06. Therefore, the fused values were scaled according to

$$FAPAR_{scaled} = \frac{0.94}{0.90} FAPAR_{fused} \quad (2)$$

The highest $FCOVER_{CYC}$ value was approximated by the 99% percentile value, i.e. $FCOVER_{CYC}(99\%) = 0.69$ (Fig. 4c). This is in agreement with results obtained by Verger (2008) showing that CYCLOPES *FCOVER* product was underestimating actual values, although being strongly linearly correlated with other regional products including the SAF-Land one. It is thus proposed to correct for this systematic underestimation by applying a scaling factor. This factor was computed considering the highest $FCOVER_{CYC}$ values observed that should correspond to full coverage ($FCOVER = 1$). The 'scaled' *FCOVER* product, $FCOVER_{scaled}$ used later to train the neural network will thus be computed according to

$$FCOVER_{scaled} = \frac{1}{0.69} FCOVER_{CYC} \quad (3)$$

Note that oppositely to *FCOVER* and *FAPAR* products, no specific theoretical upper limit exists for *LAI* since values larger than 10 are often reported at least at the local scale (Scurlock et al., 2001). Regarding the theoretical lower bound ($LAI = FAPAR = FCOVER = 0$) corresponding to bare soil, the cumulated frequencies displayed in Fig. 4b do not show particular problems for the fused product. However MODIS products show an offset of 0.25 for *LAI* (Fig. 4a) and 0.18 (Fig. 4b) for *FAPAR* values, confirming the previous observations of McCallum et al. (2010). This bias was corrected by the fusion process used for deriving GEOV1 where CYCLOPES *LAI* and *FAPAR* products contribute the most for these low vegetation amounts.

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

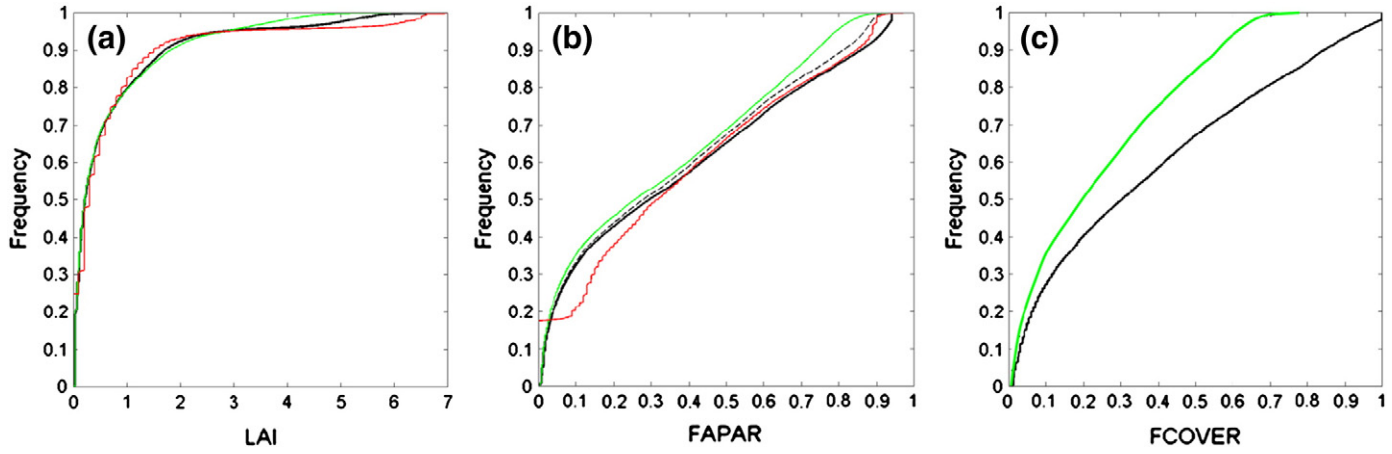

Fig. 4. Cumulated frequency of CYCLOPES (green solid line), MODIS (red solid line) fused (dashed black line for *FAPAR*) and scaled (black solid line) *LAI* (a), *FAPAR* (b) and *FCOVER* (c) products as observed over the 420 BELMANIP2 sites during years 2003–2004.

2.3. Training the neural networks

The inputs and outputs of the training dataset and the neural network architecture and learning process are described in this section.

2.3.1. Inputs

The inputs of the neural network correspond to the bidirectional reflectance factor (BRF) as measured by VEGETATION aboard SPOT. They correspond to directionally normalized top of canopy reflectance in the red, NIR and SWIR bands as derived from the CYCLOPES L3a products. The preprocessing steps include cloud screening, atmospheric correction based on a climatology of aerosols, and BRDF (Bi–directional Reflectance Distribution Function) normalization using a robust fit of Roujean's model (Hagolle et al., 2004; Roujean et al., 1992). Details about the processing from L0 (raw signal) to L3a can be found in the literature (Baret et al., 2007). However, the original CYCLOPES L3a products were reprocessed to benefit from updated radiometric calibration coefficients for the VEGETATION sensors. The blue band was not considered here since it brings only little extra information on the surface as compared to the red, NIR and SWIR bands, while being very sensitive to errors in the atmospheric correction (Bacour et al., 2006). To match the spatial support used for the 'best estimates' described earlier, the median was computed over the valid pixels within the $3 \times 3 = 9$ potential ones. The median for each band was preferred here, since the use of the 70% percentile for each band would result in possible spectral inconsistencies because the sensitivity of reflectance to canopy variables is different between bands. Note however that the CYCLOPES L3a products show a good stability over the BELMANIP2 sites which are relatively homogeneous as earlier demonstrated by Verger et al. (2008).

In addition to the BRF in the red, NIR and SWIR bands, the algorithm used the cosine of the median value of the sun zenith angle (θ_s) corresponding to the valid observations during the 30 days window over which the directionally normalized top of canopy reflectance values were composited.

2.3.2. Outputs

The outputs correspond to the targeted *LAI*, *FAPAR* and *FCOVER* variables. To further verify that the resulting 'best estimates' were consistent, they were plotted against $NDVI = (BRF_{NIR} - BRF_{Red}) / (BRF_{NIR} + BRF_{Red})$ values computed from the CYCLOPES BRF products in the red (BRF_{Red}) and in the near infrared (BRF_{NIR}). Cases with $NDVI < 0.05$ were rejected since these values are not expected over bare soil or vegetation pixels (results derived from *NDVI* values computed for VEGETATION sensor bands based on a large database of soil reflectance available at WWW2 and described in Liu et al. (2002). Then, for each class of *NDVI* values (20 classes over the [0, 1] domain of variation), the cases with *FAPAR* or *LAI* values lower (respectively higher) than the 5% percentile (respectively 95%) were rejected. This allowed to further improve the consistency of outputs with input reflectance values through *NDVI* as attested by Fig. 5. Relatively few cases were observed for *LAI* between 3 and 4. This is explained by the fact that forests that are the more likely to show such median to high *LAI* values (Weiss et al., 2007) represent only 25% of the global land surfaces. Further, many forest sites are frequently covered by clouds in addition to snow cover and poor illumination conditions that are frequently observed in winter for the high northern latitudes. Although this situation is not ideal, the neural networks should be able to interpolate efficiently between the cases available on both sides of this area.

The consistency of the output variables was further evaluated. As expected, the relationship between *LAI* and *FAPAR* was keeping very consistent as compared to the original CYCLOPES and MODIS products (Fig. 6). The same consistency is observed between *LAI* and *FCOVER* as well as between *FAPAR* and *FCOVER* (results not presented for the sake of brevity).

2.3.3. Neural network architecture and learning process

The previously described dataset is finally made of 14,200 cases where consistent top of canopy directionally normalized reflectance values are paired with 'best estimates' of *LAI*, *FAPAR* and *FCOVER* values. This represents roughly 47% of the total 30,200 data potentially existing over the 420 sites during the 2003–2004 period (72 dekads). The available data were randomly split into a training dataset made of 90% of the data available, and a test dataset (10% of the data) used for testing the hyper-specialization of the training process and evaluating the theoretical performances. The large fraction of data used for the training process allows getting a better representativeness of surface types and conditions considering the limited time period (2 years) and sites considered (420 sites) and the large fraction of missing data. The inputs and outputs are normalized to prevent possible numerical problems during the training process. Normalization is achieved by scaling between −1 and +1 the range of variation of input and output values according to

$$x_{norm} = 2\frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1 \quad (4)$$

where x represents either the inputs or outputs, x_{min} and x_{max} are respectively the minimum and maximum values of x and x_{norm} is the corresponding normalized value.

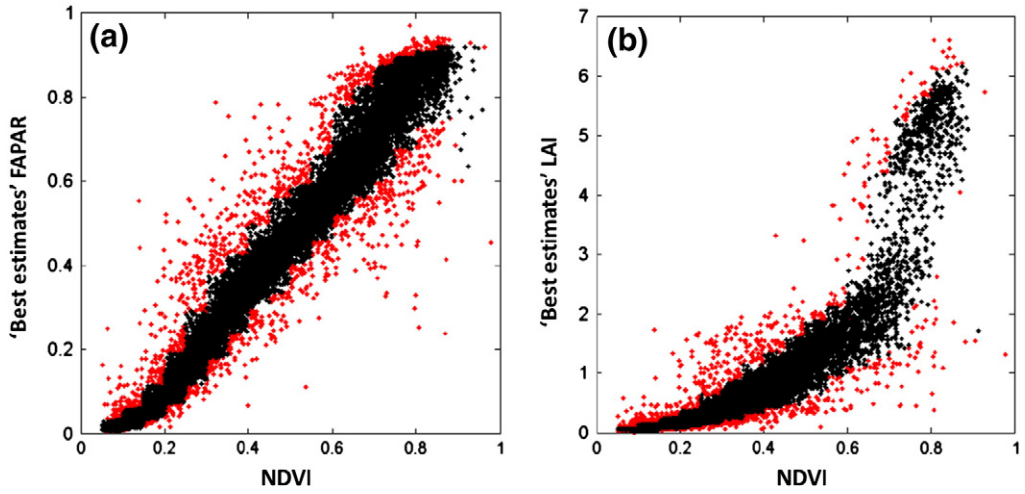

Fig. 5. Relationships between *NDVI* and both *FAPAR* (a) and *LAI* (b) 'best estimates'. The red points correspond to the cases rejected that are outside the [5% 95%] percentile range.

A back-propagation network architecture has been used since it proves very efficient in similar problems (Baret et al., 2007; Verger et al., 2010a). It is made of one hidden layer of 5 tangent-sigmoidal neurons, and one output layer made of a single linear neuron. This architecture includes 31 coefficients to be adjusted (25 synaptic weights and 6 biases) providing more than 400 cases per coefficient to be adjusted. The Levenberg–Marquardt optimization algorithm (Ngia and Sjoberg, 2000) is used for adjusting the synaptic weights and neuron bias to get the best agreement between the output simulated by the network and the corresponding value of canopy biophysical variable in the training dataset. The initial values of the weights and biases were set to a random value between −1.0 and +1.0. Three networks were trained in parallel for each targeted variable with variation in the initial values. The one providing the best performances over the test dataset was selected. Results obtained during the training process showed that the three parallel networks were performing very similarly, indicating a robust training process. A dedicated neural network for each variable was preferred here to using a unique network for estimating concurrently the three output variables because the associated architecture is simpler, leading to an easier training process while still providing very good consistency between *LAI*, *FAPAR* and *FCOVER* estimated variables, similarly to what is observed on the training dataset (Fig. 7).

The theoretical performances were evaluated over the test dataset. It shows that the training was very efficient for the three variables. The dispersion around the 1:1 line is very small and no bias is observed over the whole range of variation of the three variables (Fig. 8). Note that *LAI* shows less accurate estimates for $3 < LAI < 4$ because of the slightly lower number of data available for this *LAI* range as seen in Fig. 5.

Further evaluation was achieved to check the specific performances of the network for the main great vegetation types. Results show (Fig. 9) that no biases are observed in the estimated *LAI*, except a slight underestimation ($LAI_{est} - LAI_{train} \approx 0.1$). The positive and negative residuals are always well balanced. The residuals are increasing in absolute value from the low *LAIs* (Non-Forest) up to the largest one (EBF). In case of EBF, most of the variability is imputed to the remaining cloud contamination.

2.4. Associated uncertainties and quality assessment

All the QC flags associated to the top of canopy reflectance values are available along with the products. They describe the nature of the surface (land/sea), the presence of snow, the possible contamination by clouds or cloud shadow, the aerosol characteristics used for the atmospheric correction, and the possible saturation of the radiometric signal. Two additional qualitative assessment criterions more directly dedicated to the biophysical products are provided along with a quantitative estimation of the associated uncertainties. The way they are computed is described hereafter.

2.4.1. Input out of range

Since the algorithm is based on a learning machine approach, it is important to verify whether the inputs of a given observation keep

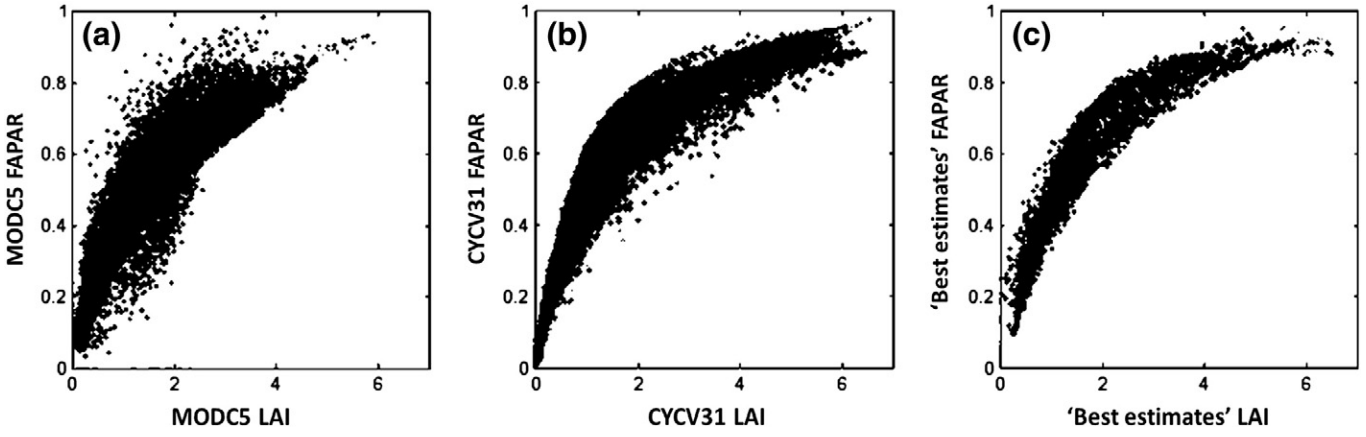

Fig. 6. Relationship between *LAI* and *FAPAR* for CYCLOPES (CYCV31 a), MODIS (MODC5 b) and 'best estimates' (c) as observed over the 420 BELMANIP2 sites during 2003–2004 period (14200 points).

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

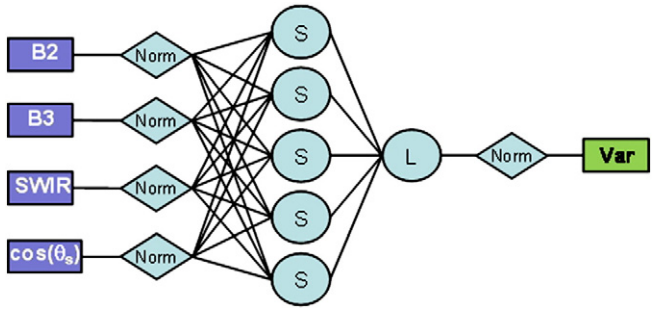

Fig. 7. Structure of the neural network used to derive *LAI*, *FAPAR* and *FCOVER* (the 'Var' box) from VEGETATION input reflectance in red (B2), NIR (B3) and SWIR as well as illumination geometry ($\cos(\theta_s)$).

within the range of variation of the training dataset called here the definition domain. If this condition is not fulfilled, the network will run in extrapolation mode, with no warranty about the realism of the outputs. The definition domain is limited by the convex hull formed in the BRF feature space by the cases used in the training process (Fig. 10). For the sake of simplicity and ease of implementation, the 3D feature space formed by B2, B3 and SWIR bands was gridded by dividing the range of variation of each band (Fig. 10 and Table 2) into 10 equally spaced classes. The ensemble of cells containing data used for the training form the definition domain. When the input BRF values are outside the definition domain, i.e. outside cells containing data used for the training, an 'input out of range' flag is raised. Note that the sun zenith angle was not included in the description of the definition domain. This would have induced increased complexity for marginal gain since the definition domains corresponding to several sun zenith angles are largely overlapping.

2.4.2. Output out of range

The physical limits for the three variables are described in Table 3. However, for *LAI*, the upper limit is not a physical limit, but a value just slightly higher than the maximum value that can be reached by the MODIS and CYCLOPES original products. Because of the several sources of uncertainties associated to the inputs, the algorithm calibration process including uncertainties attached to the original MODIS and CYCLOPES products, a tolerance is set for the extreme values: When the neural network provides biophysical variable estimates outside their definition range the value will be always set to the closest bound of the range, i.e. either the minimum or the maximum values (Table 3). The product uncertainty value will be also set to its maximum value. However, the output status flag is thus raised only when the output is outside the output range enlarged by the tolerance values [P^{tol}_{min} , P^{tol}_{max}] as defined in Table 3.

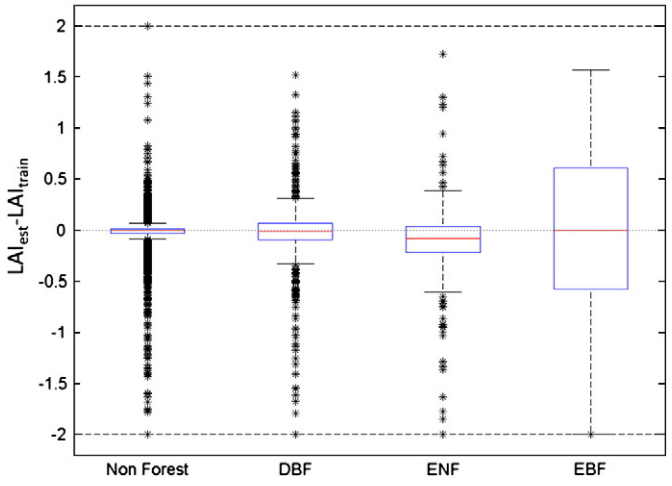

Fig. 9. Difference between the *LAI* estimated from the NNT (LAI_{est}) and the 'best estimates' *LAI* value used for the training (LAI_{train}) for the four main vegetation types: Evergreen Broadleaf Forest (EBF), Needle leaf Forest (ENF), Deciduous Broadleaf Forest (DBF) and Non-Forest. The red line, blue box and black whiskers represent respectively 50%, [25%,75%] and [5%-95%] percentiles. Outliers are indicated by stars.

2.4.3. Product uncertainties

The uncertainties associated to each biophysical variable are computed over the training dataset: For each case in the training dataset, an RMSE value is computed over the biophysical variables that have their corresponding input BRF values within the uncertainty domain (Fig. 11). Further, the sun zenith angles selected have also to be within $\pm 5°$ around the direction of the case considered. Uncertainties on reflectance are derived from Baret et al. (2007): The standard deviation, σ, is computed for each waveband according to the uncertainty model reported in Table 4:

$$\sigma = a + b \cdot \mathrm{BRF} \tag{5}$$

where *a* and *b* are coefficients.

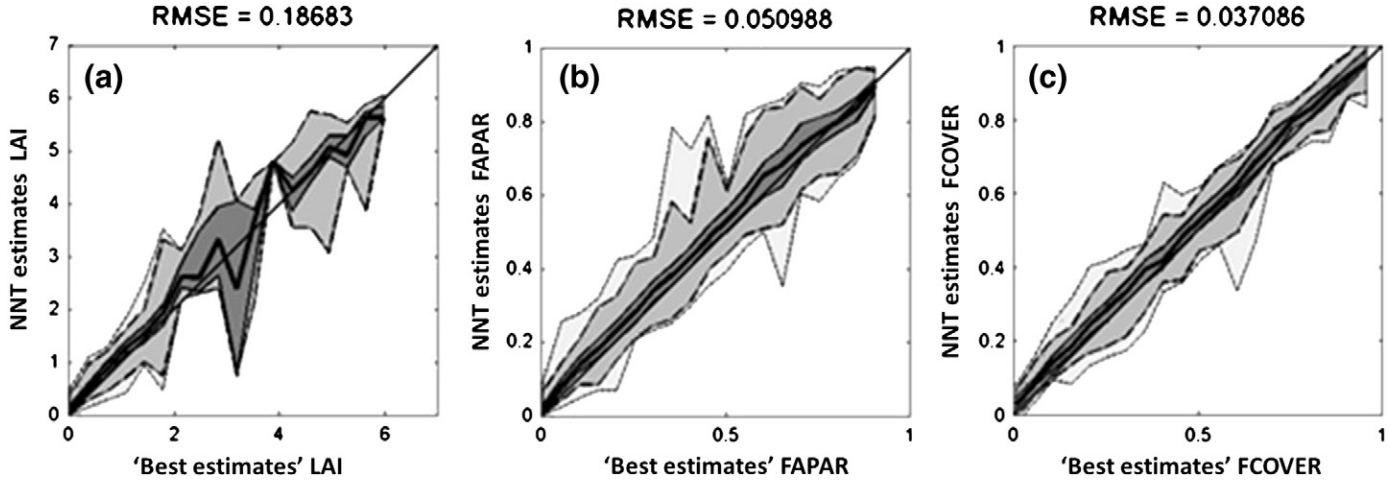

Fig. 8. Comparison between the 'best estimates' of *LAI* (a), *FAPAR* (b) and *FCOVER* (c) products and the values estimated from the trained neural networks (called NNT estimates). The solid thick line corresponds to the median value. Dark, medium and light gray areas correspond respectively to [25% 75%] [10% 90%] and [1% 99%] of the estimated cases for each class of actual values (20 classes are considered, from 0 to the maximum of actual values). The data correspond to the test dataset (n = 1420).

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

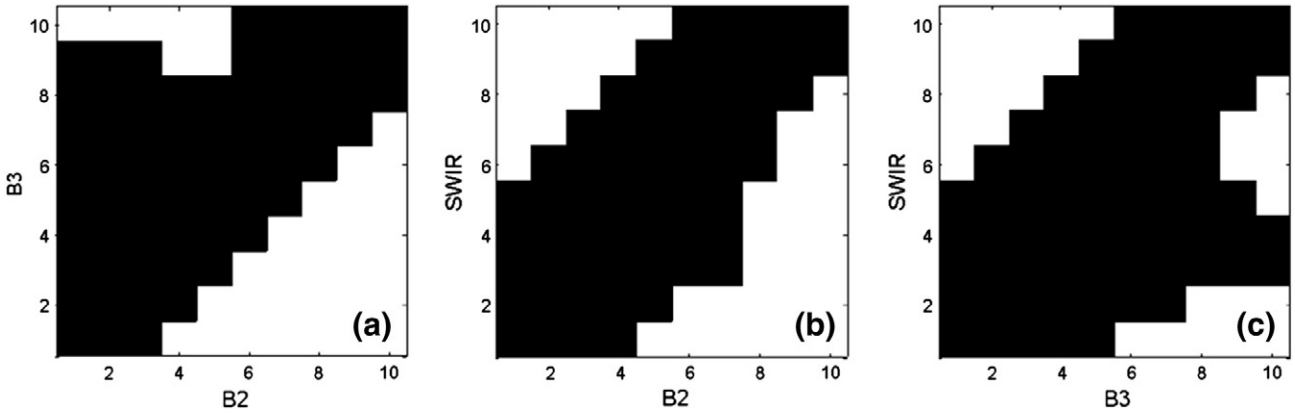

Fig. 10. Definition domain for input VEGETATION Top of canopy directionally normalized BRF values in B2, B3 and SWIR bands. Axes are scaled between 1 to 10, corresponding to the 10 classes of reflectance values ranging between the minimum value (class 1) and the maximum value (class 10). The cells containing data from the training dataset are represented in black. They form the definition domain.

A neural network is then trained to relate these computed RMSE values to the corresponding input variables. This learning process is similar to what was previously achieved for the product estimation itself and the network has the same architecture as the one used for the derivation of the products (Fig. 7). The same normalization as used for the variables is also applied for the inputs and outputs of the neural network dedicated to uncertainties. Results (Fig. 12) show that the model of uncertainties is relatively robust for *LAI*, *FAPAR* and *FCOVER*, with however some degradation of performances for the larger values of the biophysical variables.

3. Operational production and dissemination

The GEOV1 processing line was derived from the CYCLOPES V3.1 processing line which has been used to generate the 1999–2007 time series of CYCLOPES products (WWW3). It has been adapted to consider the specificities of the GEOV1 products in terms of algorithm, and output format requested by the users. It has also been consolidated to fit the software conventions (general coding rules, computer platform issues, filename and directory conventions) defined by the operational processing center. At different stages of the development process, reviews have been carried out to check that the several input specifications were answered in the design of the processing line. Further, a rigorous methodology of validation based upon unit and integration tests, as well as scientific analysis of the output using tools including visual control and statistical metrics has been applied.

Before its integration in the operational processing center, the processing line has been run to generate 2 years (2003–2004) of GEOV1 demonstration products used to perform the validation exercise presented in Camacho et al. (in press). They have also been supplied to Geoland2 users who checked the consistency with their requirements before starting the operational production of real time and the whole 12 years time series processing of observations. The processing line has been optimized to ensure operational management with permanent QC while benefiting from the variable computation resources available for parallel processing.

The GEOV1 products are generated in multi-band hdf5 format (the variable, its uncertainty, the quality flag, the number of input observations, and the land–sea mask) and in tiles of $10^\circ \times 10^\circ$ covering the land surfaces of the whole globe. They are available in open access through the Geoland2 web platform (WWW4) where users can browse the catalogue, order the products after registration, and subscribe to receive the products. The GEOV1 products are also disseminated via the Eumetcast system to African and South American users.

Table 2
Range of variation of BRF values (minimum and maximum) observed in the training dataset and used to compute the definition domain.

	Minimum	Maximum
B2	0.000	0.429
B3	0.036	0.547
SWIR	0.000	0.648

Table 3
Minimum, maximum, resolution and tolerance values used to raise the output out of range flag.

	Minimum	Maximum	Resolution	Tol_{min}	Tol_{max}
LAI	0	7.0	0.01	−0.2	7.20
FAPAR	0	1.0	0.01	−0.05	1.05
FCOVER	0	1.0	0.01	−0.05	1.05

4. Conclusion

The GEOV1 *LAI*, *FAPAR* and *FCOVER* products capitalize on the efforts undertaken this last decade in the development and validation of biophysical products from medium resolution observations. The pragmatic approach used here is based on the fusion of CYCLOPES

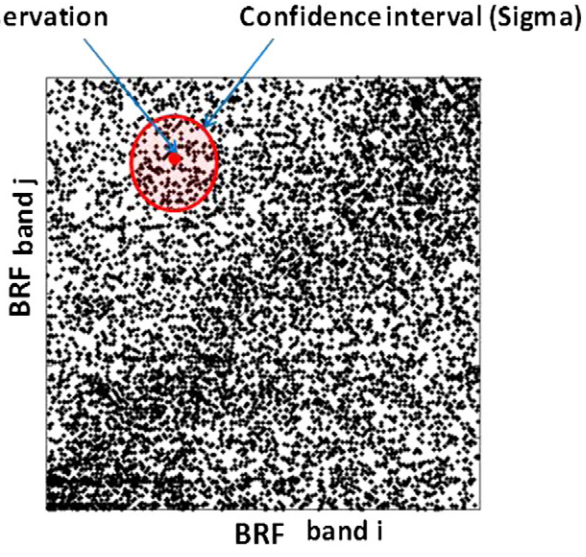

Fig. 11. Scheme showing how the uncertainties attached to the products were computed.

Table 4
Values of the uncertainty model used for the input BRFs.
From Baret et al. (2007).

	B2	B3	SWIR
a	0.005	0.003	0.005
b	0.05	0.03	0.03

and MODIS products that were demonstrated to perform the best. However, their deficiencies observed respectively for low *LAI* values for MODIS and high *LAI* values for CYCLOPES have been corrected in the fusion process. Further, the resulting *FAPAR* and *FCOVER* products have been scaled to reach the theoretical upper limit expected for high vegetation amounts. Finally, one of the main advantages of the GEOV1 algorithm is that it did not explicitly use a biome classification that introduces some spatial inconsistencies and sometimes temporal inconsistencies (when class assignment change unexpectedly) as outlined by Yang et al. (2006a) for MODIS.

The approach undertaken here was not calibrated using ground validation measurements, preserving the required independency between calibration and validation processes. However, this led to use a more pragmatic approach based on heuristic arguments that explains the sometimes subjective selection of criterions used in the development of the GEOV1 algorithm and several alternative solutions could have been proposed. Nevertheless the GEOV1 products resulting from the largely inductive approach undertaken yield robust, consistent and accurate estimates of these key biophysical variables as demonstrated in Camacho et al. (in press). The validation shown in the companion paper demonstrates that GEOV1 significantly improved the performances of currently existing products, both regarding accuracy and spatial as well as temporal consistency. It provides quantified estimates of uncertainties, although these are simply derived from the training dataset that reflects mostly the sensitivity of the product to input reflectance values. Further investigations should be directed towards a better quantification of the uncertainties, including 'model' assumptions since both MODIS and CYCLOPES are based on radiative transfer model inversion.

A clear understanding of the actual definition of the retrieved variables and their consistency with application requirements is one of the key aspects to consider when deriving biophysical variables from remote sensing observations. For *FCOVER*, no particular question is raised since the *FCOVER* definition is simple and clear if restricted to the green vegetation elements including overstory and understory. For *FAPAR*, because of the good consistency between MODIS and CYCLOPES definitions, the output product is also well defined: black-sky *FAPAR* (green element including over and understory) at 10:15 (actually between 10:00 and 10:30). For *LAI*, the question is more complex since the original MODIS and CYCLOPES *LAI* products are defined differently, at least with regards to the assumptions embedded in the radiative transfer models used. For MODIS, clumping at the tree scale is accounted for, but not at the shoot, nor at the landscape scales. For CYCLOPES, clumping is accounted for at the landscape scale only. However, its effect will be significant mainly for the larger *LAI* values as discussed by Garrigues et al. (2006). Therefore, GEOV1 *LAI* product will marginally account for landscape clumping as demonstrated by the good consistency between MODIS and CYCLOPES *LAI* products for low to medium *LAI* values (Camacho et al., in press). However there is a special case of savannas with relatively low *LAI* values and significant clumping (Ryu et al., 2010). In this case, the radiation interaction between plants that are separated by significant distances is minimal, and clumping described at the plant level in MODIS may be approximated at the landscape scale as in CYCLOPES. Clumping at the shoot scale will not be accounted for in GEOV1 *LAI* product since neither MODIS, nor CYCLOPES account for it. Conversely, clumping at the plant scale was accounted mainly for forests and savannas in the original MODIS product but not in the CYCLOPES ones. However, clumping is mainly observed for the higher *LAI* values (overlap between leaves within the tree volume when the number of leaves increases) (Rochdi, 2003). Therefore, clumping at the plant scale as described in MODIS products should be partly preserved in the GEOV1 *LAI* product. The validation achieved in the companion paper (Camacho et al., in press) over a limited number of sites demonstrates that the GEOV1 *LAI* products are close to the values estimated from indirect techniques and that include the clumping effect.

These products may thus be used in a range of applications including those targeted for the ECVs and fulfill the needs for GMES and the GCOS task dedicated to the operationalization of the generation of *FAPAR* and *LAI* products. The GEOV1 products initiate a service whose sustainability is planned within the GMES Land Monitoring Core Service by adapting the algorithm, and the processing chain, to AVHRR/Metop, PROBA-V, and Sentinel-3 missions. Further, it will be completed backward using the AVHRR data as processed by Vermote et al. (2010) to get a long and consistent time series of more than 30 years.

Although the development of this series of biophysical products should represent an important step towards a more effective use of remote sensing observations, improvements are expected mainly through several aspects. Clouds constitute obviously the major limitation of optical systems that suffer from large areas and periods without data. Recent studies demonstrate that the fusion between several sensors improves data continuity (Hagolle et al., 2005; Verger et al., 2010b; Yang et al., 2006b). Alternative approaches based on enhanced

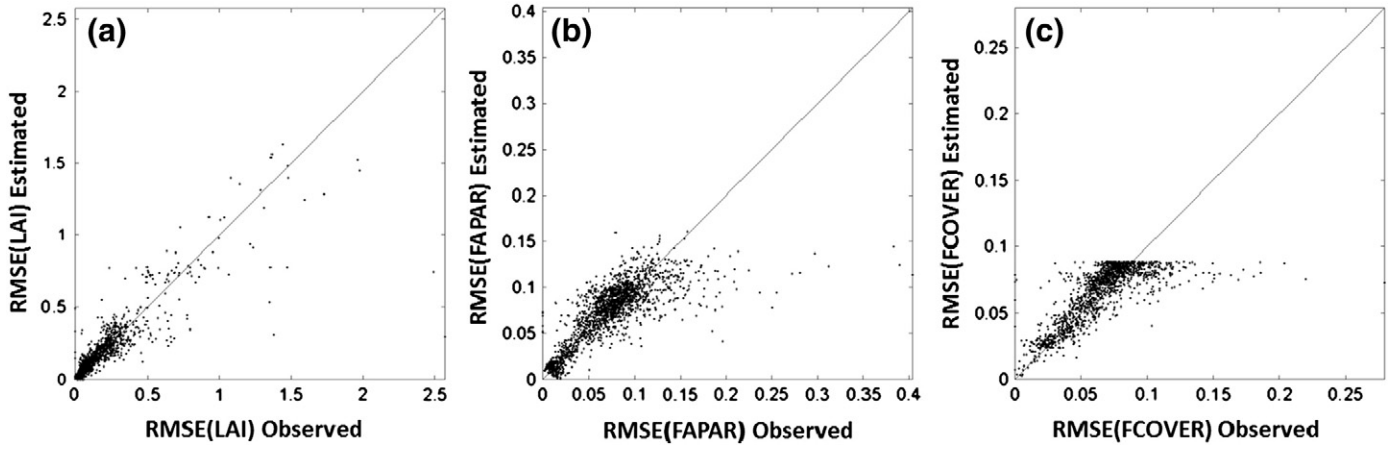

Fig. 12. Theoretical performances of the neural network model used to describe RMSE values for *LAI* (a) *FAPAR* (b) and *FCOVER* (c) from input reflectance values and geometry of observation.

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

time series processing may help removing outliers, filling gaps due to missing observations and smoothing the temporal profiles (Kandasamy et al., 2013; Verger et al., in press). However, the main limitation in such global products comes mainly from the little a priori information available and required to regularize the inversion process (Combal et al., 2002). The use of global classification is likely to be insufficient because of its limited accuracy (Defourny et al., 2009; Herold et al., 2008; Yang et al., 2006a) and because the variability in canopy architecture and optical properties within a given class is probably as large as between classes when the seasonal variability is considered. Further, the often mixed nature of kilometric pixels poses both a scaling issue and the difficulty to identify the several co-existing surface patches. These problems call for an improved spatial resolution that will allow resolving most of the vegetation patches and will authorize identifying the corresponding vegetation type from the past observations and use it as prior information. Such systems are currently being available with hectometric resolutions, such as the PROBA-V (300 m daily), Sentinel-3 (300 m every 2 days), and VIIRS (370 m daily). However, decametric systems such as Sentinel-2 or LDCM (Landsat Data Continuity Mission) in combination with the previous hectometric ones would probably provide the most efficient observation system.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no. 218795.

References

Avissar, R., & Pielke, R. A. (1989). A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Monthly Weather Review*, *117*, 2113–2136.

Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data: Principles and validation. *Remote Sensing of Environment*, *105*, 313–325.

Baret, F., Bacour, C., Béal, D., Weiss, M., Bruniquel, V., & Regner, P. (2005). Algorithm theoretical basis document for MERIS top of atmosphere land products (TOA_VEG). In (p. 32). Avignon: INRA-CSE.

Baret, F., Clevers, J. G. P. W., & Steven, M. D. (1995). The robustness of canopy gap fraction estimates from red and near infrared reflectances: A comparison of approaches. *Remote Sensing of Environment*, *54*, 141–151.

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, *35*, 161–173.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., et al. (2007). LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm. *Remote Sensing of Enviroment*, *110*, 275–286.

Baret, F., Morissette, J., Fernandes, R., Champeaux, J. L., Myneni, R., Chen, J., et al. (2006). Evaluation of the representativeness of networks of sites for the global validation and inter-comparison of land biophysical products. Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1794–1803.

Baret, F., Nightingale, J., Garrigues, S., & Nickeson, J. (2009). Report on the CEOS Land Product Validation sub-group meeting, Missoula, Montana, June 15th 2009. *The Earth Observer*, *21*, 26–30.

Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., & Weiss, M. (in press). GEOV1: LAI, FAPAR essential climate variables and FCover global time series capitalizing over existing products. Part 2: Validation and inter-comparison with reference products. *Remote Sensing of Environment*.

Camacho-de Coca, F., Jiménez-Muñoz, J. C., Martínez, B., Bicheron, P., Lacaze, R., & Leroy, M. (2006). Prototyping of the fCover product over Africa based on existing CYCLOPES and JRC products for VGT4Africa. In J. Sobrino (Ed.), *Proceedings of the RAQRS'II 2nd International Symposium on Recent Advances in Remote Sensing* (pp. 722–726). Valencia, Spain: Univ. Valencia.

Carlson, T. N., & Ripley, D. A. (1997). On the relationship between NDVI, fractional cover and leaf area index. *Remote Sensing of Enviroment*, *62*, 241–252.

Chen, J. M., & Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell & Environment*, *15*, 421–429.

Chen, J. M., Deng, F., & Chen, M. (2006). Locally adjusted cubic spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2230–2238.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al. (2002). Retrieval of canopy biophysical variables from bi-directional reflectance data. Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, *84*, 1–15.

De Kauwe, M. G., Disney, M. I., Quaife, T., Lewis, P., & Williams, M. (2011). An assessment of the MODIS collection 5 leaf area index product for a region of mixed coniferous forest. *Remote Sensing of Environment*, *115*, 767.

Defourny, P., Schouten, L., Bartalev, S., Bontemps, S., Cacetta, P., de Wit, A. J. W., et al. (2009). Accuracy assessment of a 300 m global land cover map: The GlobCover experience. *Proceedings of the 33rd International Symposium on Remote Sensing of Environment* (Stresa (Italy)).

Deng, F., Chen, J. M., Chen, M., & Pisek, J. (2006). Algorithm for global leaf area index retrieval using satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2219–2229.

Ganguly, S., Samanta, A., Schull, M. A., Shabanov, N. V., Milesi, C., Nemani, R. R., et al. (2008). Generating vegetation leaf area index earth system data record from multiple sensors. Part 1: Theory. *Remote Sensing of Environment*, *112*, 4318–4332.

Ganguly, S., Schull, M. A., Samanta, A., Shabanov, N. V., Milesi, C., Nemani, R. R., et al. (2008). Generating vegetation leaf area index earth system data record from multiple sensors. Part 2: Evaluation. *Remote Sensing of Environment*, *112*, 4333.

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

Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., et al. (2008). Validation and intercomparison of global leaf area index products derived from remote sensing data. *Journal of Geophysical Research*, *113*.

GCOS (2006). GCOS-107. Supplemental details to the satellite based component of the "implementation plan for the global observing system for climate in support of the UNFCCC". In (p. 103). Geneve (Switzerland): GCOS/WMO.

Gobron, N., Pinty, B., Aussedat, O., Chen, J. M., Cohen, W. B., Fensholt, R., et al. (2006). Evaluation of fraction of absorbed photosynthetically active radiation products for different canopy radiation transfer regimes: Methodology and results using Joint Research Center products derived from SeaWiFS against ground-based estimations. *Journal of Geophysical Research-Atmospheres*, *111*.

Gutman, G. G. (1991). Vegetation indices from AVHRR data: An update and future prospects. *Remote Sensing of Environment*, *35*, 121–136.

Hagolle, O., Lobo, A., Maisongrande, P., Cabot, F., Duchemin, B., & De Pereyra, A. (2005). Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of VEGETATION 1 and 2 images. *Remote Sensing of Environment*, *94*, 172–186.

Hagolle, O., Nicolas, J. -M., Fougnie, B., Cabot, F., & Henry, P. (2004). Absolute calibration of VEGETATION derived from an interband method based on the sun glint over ocean. *IEEE Transactions on Geoscience and Remote Sensing*, *42*, 1472–1481.

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, *112*, 2538.

Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of smoothing and gap filling methods for time series of remote sensing observations. Application to MODIS LAI products. *Biogeosciences*, *9*, 17053–17097.

Kraus, T. (2008). Ground-based validation of the MODIS leaf area index product for east African rain forest ecosystems. In. Erlangen: Universitätsbibliothek der Universität Erlangen-Nürnberg.

Lacaze, R., Atzberger, C., Bartholomé, E., Combal, B., Calvet, J. C., Lefèvre, V., & Olsson, B. (2009). BioPar User Requirements. In. Toulouse.

Liu, G. R., Lin, T. H., & Kuo, T. H. (2002). Estimation of aerosol optical depth by applying the optimal distance number to NOAA AVHRR data. *Remote Sensing of Environment*, *81*, 247–252.

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., et al. (2009). Satellite-based terrestrial production efficiency modeling. *Carbon Balance and Management*, *4*, 8.

McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., et al. (2010). Comparison of four global FAPAR datasets over Northern Eurasia for the year 2000. *Remote Sensing of Environment*, *114*, 941.

Morisette, J., Baret, F., Privette, J. L., Myneni, R. B., Nickeson, J., Garrigues, S., et al. (2006). Validation of global moderate resolution LAI products: A framework proposed within the CEOS Land Product Validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1804–1817.

Ngia, L. S. H., & Sjoberg, J. (2000). Efficient training of neural nets for nonlinear adaptive filtering using a recursive Levenberg–Marquardt algorithm. *IEEE Transactions on Signal Processing*, *48*, 1915–1927.

Rochdi, N. (2003). *Un modèle générique d'agrégation des feuilles dans un couvert végétal: Application à la simulation du transfert radiatif* (pp. 169). Paris, France: Institut National Agronomique.

Roujean, J. L., Leroy, M., & Deschamps, P. Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, *97*, 20455–20468.

Ryu, Y., Sonnentag, O., Nilson, T., Vargas, R., Kobayashi, H., Wenk, R., et al. (2010). How to quantify tree leaf area index in an open savanna ecosystem: A multi-instrument and multi-model approach. *Agricultural and Forest Meteorology*, *150*, 63–76.

Scurlock, J. M. O., Asner, G. P., & Gower, S. T. (2001). Worldwide historical estimates and bibliography of leaf area index, 1932–2000. In. Oak Ridge, Tennessee, U.S.A.: Oak Ridge National Laboratory.

Sprintsin, M., Karnieli, A., Berliner, P., Rotenberg, E., Yakir, D., & Cohen, S. (2009). Evaluating the performance of the MODIS Leaf Area Index (LAI) product over a Mediterranean dryland planted forest. *International Journal of Remote Sensing*, *30*, 5061–5069.

Verger, A. (2008). Analisi comparativa d'algorismes operacionals d'estimacio de parametres biofisics de la coberta vegetal amb teledeteccio. *Departament de fisica de la terra i termodinamica* (pp. 277). Valencia, Spain: Universitat de Valencia.

Verger, A., Baret, F., García-Haro, F. J., Camacho, F., & Meliá, J. (2010). Consistency of vegetation estimates from SEVIRI/Meteosat observations and operational algorithms. In J. A. Sobrino (Ed.), *Third International Symposium on Recent Advances*

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027

in Quantitative Remote Sensing. Torrent, Spain: Publicacions de la Universitat de València.

Verger, A., Baret, F., & Weiss, M. (2008). Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment*, *112*, 2789–2803.

Verger, A., Baret, F., & Weiss, M. (2010). Fusion of MODIS and VEGETATION observations for improved consistency and continuity of LAI product time series. In J. A. Sobrino (Ed.), *Third International Symposium on Recent Advances in Quantitative Remote Sensing*. Torrent, Spain: Publicacions de la Universitat de València.

Verger, A., Baret, F., & Weiss, M. (2011). A multisensor fusion approach to improve LAI time series. *Remote Sensing of Enviroment*, *115*, 2460–2470.

Verger, A., Baret, F., Weiss, M., Kandasamy, S., & Vermote, E. (in press). The CACAO method for smoothing, gap filling and characterizing anomalies in satellite time series. *IEEE Transactions on Geoscience and Remote Sensing*. http://dx.doi.org/10.1109/TGRS.2012.2228653.

Verger, A., Martínez, B., Camacho-de Coca, F., & García-Haro, F. J. (2009). Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. *International Journal of Remote Sensing*, *30*, 2685–2704.

Vermote, E., Justice, C., Csiszar, I., Eidenshink, J., Mynemi, R., Baret, F., et al. (2010). A terrestrial surface climate data record for global change studies. In J. A. Sobrino (Ed.), *Third International Symposium on Recent Advances in Quantitative Remote Sensing*. Torrent, Spain: Publicacions de la Universitat de València.

Weiss, M., Baret, F., Garrigues, S., Lacaze, R., & Bicheron, P. (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.

Yang, W., Huang, D., Tan, B., Stroeve, J. C., Shabanov, N., Knyazikhin, Y., et al. (2006). Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the Terra MODIS sensor: 2000–2005. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1829–1842.

Yang, W., Shabanov, N. V., Huang, D., Wang, W., Dickinson, R. E., Nemani, R. R., et al. (2006). Analysis of leaf area index products from combination of MODIS Terra and Aqua data. *Remote Sensing of Environment*, *104*, 297–312.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., & Shangguan, W. (2011). Reprocessing the MODIS leaf area index products for land surface and climate modelling. *Remote Sensing of Environment*, *115*, 1171.

WWW references

WWW1: http://calvalportal.ceos.org/cvp/web/guest/olive
WWW2: https://www4.paca.inra.fr/emmah_eng
WWW3: http://postel.mediasfrance.org
WWW4: http://www.geoland2.eu

Please cite this article as: Baret, F., et al., GEOV1: *LAI* and *FAPAR* essential climate variables and *FCOVER* global time series capitalizing over existing products. Part1: Princip..., *Remote Sensing of Environment* (2013), http://dx.doi.org/10.1016/j.rse.2012.12.027