



**HAL**  
open science

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

Frédéric Baret, M Weiss, R Lacaze, F Camacho, H Makhmara, P Pacholczyk,  
B Smets

## **► To cite this version:**

Frédéric Baret, M Weiss, R Lacaze, F Camacho, H Makhmara, et al.. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: Principles of development and production. *Remote Sensing of Environment*, 2013, 137, pp.299-309. 10.1016/j.rse.2012.12.027 . hal-01321330

**HAL Id: hal-01321330**

**<https://hal.science/hal-01321330v1>**

Submitted on 29 May 2020

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

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

1 **GEOV1: LAI, FAPAR Essential Climate Variables and FCOVER global time series**  
2 **capitalizing over existing products. Part1: Principles of development and production.**

3

4 F. Baret<sup>1</sup>, M. Weiss<sup>1</sup>, R. Lacaze<sup>2</sup>, F. Camacho<sup>3</sup>, H. Makhmara<sup>4</sup>, P. Pacholczyk<sup>4</sup> and B. Smets<sup>5</sup>

5 <sup>1</sup> INRA-EMMAH UMR 1114, Avignon, France

6 <sup>2</sup> HYGEOS, Toulouse, France

7 <sup>3</sup> EOLAB, Valencia, Spain

8 <sup>4</sup> CNES, Toulouse, France

9 <sup>5</sup> VITO, Mol, Belgium

10 Corresponding author: baret@avignon.inra.fr

11 **ABSTRACT**

12 Essential Climate variables such as *LAI* or *FAPAR* are required for the monitoring,  
13 understanding and modeling of land surfaces at the global scale. While several products were  
14 already developed from the current medium resolution sensors, the few validation exercises  
15 currently achieved highlighted significant discrepancies and inconsistencies. The objective of  
16 this study is to develop improved global estimates of *LAI*, *FAPAR* and *FCOVER* variables by  
17 capitalizing on the development and validation of already existing products. In a first step, the  
18 performances of the MODIS, CYCLOPES, GLOBCARBON and JRC-FAPAR products were  
19 reviewed. The MODIS and CYCLOPES products were then selected since they provide  
20 higher level of consistency. These products were fused to generate the improved *LAI*, *FAPAR*  
21 and *FCOVER* values that were later scaled to closely match their expected range of variation.  
22 Finally, neural networks were trained to estimate these fused and scaled products from SPOT-  
23 VEGETATION top of canopy directionally normalized reflectance values. The resulting  
24 GEOV1 products are associated to quality control flags as well as quantitative estimates of  
25 uncertainties. Performances of the GEOV1 products are finally evaluated in a companion  
26 paper. The GEOV1 products are freely available to the community at [www.geoland2.eu](http://www.geoland2.eu) from  
27 1999 up to present, globally at 1/112° spatial sampling grid at the dekadal time step.

28

29 **1 Introduction**

30 The importance of continuously monitoring the Earth's surface was recently recognized by  
31 GCOS (Global Climate Observing System) (GCOS 2006). Essential Climate Variables (ECV)  
32 related to land surfaces such as *LAI* (Leaf Area Index) and (Fraction of Absorbed  
33 Photosynthetic Active Radiation) may be derived from observations in the reflective solar  
34 domain. These vegetation biophysical variables play a key role in several surface processes,  
35 including photosynthesis, respiration and transpiration. *LAI* is defined as half the total  
36 developed area of green elements per unit horizontal ground area (Chen and Black 1992).  
37 *FAPAR* is defined as the fraction of radiation absorbed by the green vegetation elements in the  
38 400 - 700 nm spectral domain under specified illumination conditions. *FAPAR* is one of the  
39 main inputs in light use efficiency models (McCallum et al. 2009). In addition to *LAI* and  
40 *FAPAR*, *FCOVER*, the fraction of green vegetation as seen from nadir, is requested by some

41 users for vegetation monitoring (Lacaze et al. 2009) as well as for partitioning contributions  
 42 between soil and vegetation within specific models for Numerical Weather Prediction, regional  
 43 and global climate modeling, and global change monitoring (Avisar and Pielke 1989).  
 44 FCOVER is independent from the illumination conditions as opposed to FAPAR while  
 45 showing sensitivity to vegetation amount intermediate between *FAPAR* and *LAI*.

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

Products	Sensors				Spatial sampling distance at equator	Temporal sampling (days)	Time period	Reference
		LAI	FAPAR	FCOVER				
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 <sup>(1)</sup>		✓		2.17 km	1	1997-2006	(Gobron et al. 2006)

48 <sup>(1)</sup> JRC-FAPAR may be derived from several sensors including SEAWIFS, MERIS, MODIS  
 49 and VEGETATION but produced globally over a long time period only from SEAWIFS.

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

58 The FP7 Geoland2 project (<http://www.gmes-geoland.info>) intends to implement the  
 59 GMES (Global Monitoring for Environment and Security) Land Monitoring Core Service that  
 60 corresponds to the European contribution to GEOSS (Group of Earth Observation System of  
 61 Systems). An operational system is developed to provide biophysical products that meet the  
 62 users' needs (Lacaze et al. 2009) for monitoring natural ecosystems and managed lands. The  
 63 main requirements correspond to 1 km spatial sampling, 10 days (dekadal) frequency with  
 64 products generated in near real time (less than one week), for a time period as long as possible  
 65 (Ganguly et al. 2008b), and associated with quantitative uncertainties. These requirements are  
 66 partly answering the recently updated GCOS ones for LAI and FAPAR ECVs with 250m  
 67 spatial resolution relaxed to 2 km for near term products, 2-weekly temporal frequency and  
 68 accuracy better than max(20%, 0.5) for LAI, max(10%, 0.05) for FAPAR and a stability better  
 69 than max(10%,0.25) for LAI and max(3%, 0.02) for FAPAR (GCOS-138). The resulting  
 70 products will thus eventually contribute to fulfill one of the GCOS task dedicated to the  
 71 operationalization of the generation of and *LAI* Essential Climate Variables.

72 The objective of this paper is to describe the first version of Geoland2 *LAI*, and *FCOVER*  
 73 products called GEOV1. The principles used to derive the products will first be presented.  
 74 Then, the algorithm development will be described. A companion paper (Camacho et al.  
 75 2012) will finally report the validation results derived according to the CEOS/LPV guidelines  
 76 (Baret et al. 2009; Morisette et al. 2006).

## 77 2 ALGORITHM DEVELOPMENT

### 78 2.1 General principles

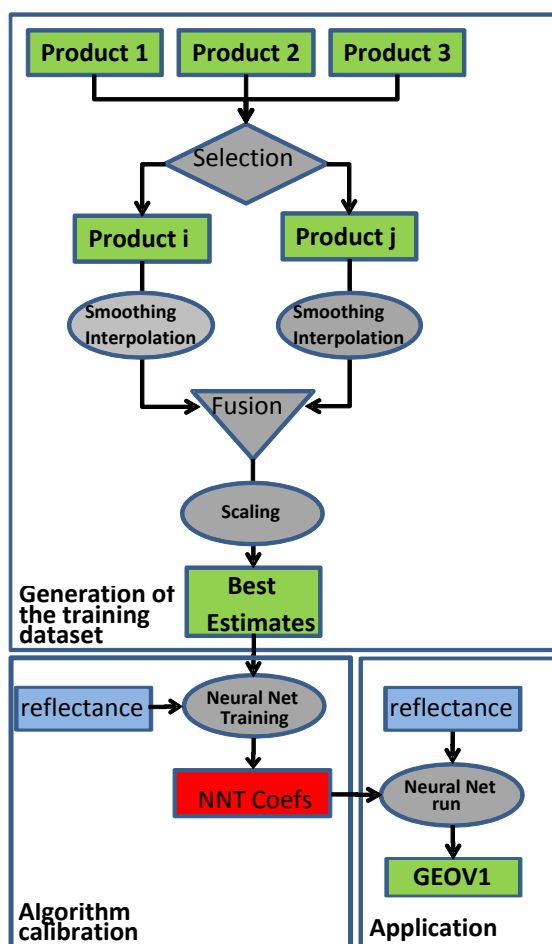
79 Since significant efforts have already been made to develop and validate biophysical products  
80 as stated in the introduction, it is thus proposed to capitalize on the existing products and  
81 associated validation results for the development of the GEOV1 products. Previous works  
82 (Verger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

86 **1-Generation of the training dataset:** already existing products are first combined to provide  
87 the ‘best estimates’ of the biophysical variables that will constitute the training dataset.

88 **2-Neural network calibration:** a neural network is trained to estimate these ‘best estimates’  
89 from the input reflectance values as observed by specific sensors and the associated  
90 geometrical configuration. Quality flags and quantitative uncertainties are also derived.

91 **3-Application of the network:** once the network is calibrated, it is run to provide estimates of  
92 the biophysical variables for each of the sensors considered, along with the quality flags and  
93 quantitative uncertainties.

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



104

105 **Figure 1. Schematic description of the principle used to develop the GEOV1 product.**

106 **2.2 Generation of training dataset**

107 The way the training dataset is generated from already existing products is sketched in Figure  
 108 1, top box. Four main steps are identified: (1) selection of the most relevant products, (2)  
 109 setting the products on consistent spatial and temporal supports, (3) fusing the products and  
 110 (4) eventually scaling the fused products. Details of each of these steps are given in the  
 111 following.

112 **2.2.1 Selection of products**

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

115 The validation exercise achieved by Garrigues et al. (2008) for *LAI* products was showing few  
 116 missing values for the GLOBCARBON *LAI* product, in relation to its monthly temporal  
 117 resolution. However, many artifacts were observed, including unexpected outliers and shifts  
 118 in the phenology due to its coarse temporal resolution and low sampling frequency. Garrigues  
 119 et al. (2008) and Weiss et al. (2007) demonstrated that CYCLOPES was providing very  
 120 smooth temporal course while showing a saturation for *LAI* values larger than 4. For MODIS  
 121 collection 5 *LAI* products, only few local validation activities were reported (De Kauwe et al.  
 122 2011; Kraus 2008; Sprintsin et al. 2009) while Yuan et al.(2011) presents some results of a  
 123 global validation exercise focusing on the evaluation of a smoothed version of MODIS *LAI* to

124 reduce the spatial and temporal inconsistencies observed at the local spatial or temporal  
125 scales. Ganguly et al. (2008a) validated an adaptation of MODIS *LAI* products for AVHRR  
126 data, showing fair consistency with CYCLOPES *LAI* product over the sites considered.  
127 However, Verger et al. (2011) demonstrated that MODIS was showing more shaky temporal  
128 *LAI* profiles as compared to the CYCLOPES ones.

129 Among the global *FAPAR* products available, MODIS and GLOBCARBON correspond  
130 roughly to the same definition, i.e. black-sky values at the time of the satellite overpass, i.e.  
131 around 10:35 for MODIS aboard Terra and 10:30 for VEGETATION. However, if JRC-  
132 *FAPAR* corresponds also to black-sky *FAPAR* at the time of the satellite overpass, SEAWIFS  
133 is crossing the equator around 12:20, i.e. for significantly smaller sun zenith angles as  
134 compared to MODIS and VEGETATION. The CYCLOPES *FAPAR* product corresponds to  
135 black-sky *FAPAR* at 10:00 illumination conditions which is a close approximation of the daily  
136 integrated black-sky *FAPAR* value (Baret et al. 2005; Baret et al. 2007). Therefore, MODIS  
137 aboard Terra, GLOBCARBON and CYCLOPES *FAPAR* products share very similar  
138 definitions, while JRC-*FAPAR* derived from SEAWIFS is expected to show lower values  
139 because of the smaller sun zenith angles experienced close to solar noon. Further, it is limited  
140 to sun zenith angles lower than 50° which may pose a problem for the higher latitudes and/or  
141 part of the season. Comparison between MODIS collection 5 and CYCLOPES *FAPAR*  
142 products was showing a good agreement, with however some overestimation of MODIS for  
143 the low *FAPAR* values (McCallum et al. 2010). GLOBCARBON and JRC-*FAPAR* were found  
144 much lower in magnitude, while a good consistency in spatial and temporal trends were found  
145 between CYCLOPES and JRC-*FAPAR* (McCallum et al. 2010). GLOBCARBON products  
146 were showing the largest discrepancies with all the other products.

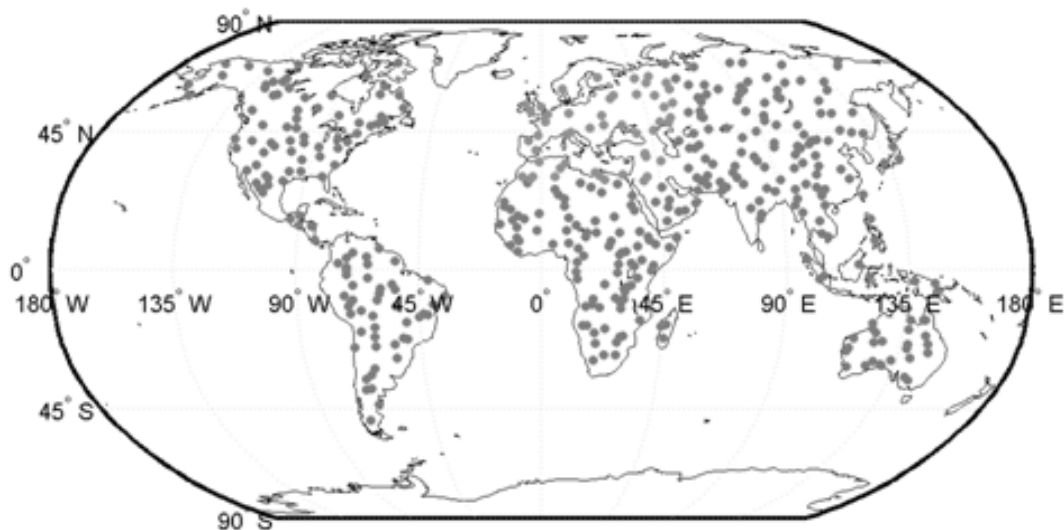
147 The MODIS and CYCLOPES *FAPAR* products were therefore selected ensuring a good  
148 consistency between *LAI* and *FAPAR* values since both MODIS and CYCLOPES provide  
149 concurrently *LAI* and *FAPAR* products. However, CYCLOPES and MODIS products are  
150 based on different assumptions and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

157 Apart from CYCLOPES *FCOVER* products, no other global *FCOVER* product is currently  
158 available apart from the SAF-LAND products covering the METEOSAT disk (Camacho-de  
159 Coca et al. 2006). However, several studies have pointed out that NDVI could be a good  
160 proxy for *FCOVER* (Baret et al. 1995; Carlson and Ripley 1997; Gutman 1991). Camacho-de  
161 Coca et al. (2006) compared several regional *FCOVER* products over Africa and showed that  
162 the CYCLOPES *FCOVER* product was very consistent with other products although a  
163 significant and systematic bias was observed. This was also confirmed by Verger et al.  
164 (2009). It is therefore proposed to select the CYCLOPES *FCOVER* original product while  
165 rescaling it to provide values more consistent with ground measurements as detailed later.

### 166 2.2.2 Spatial and temporal sampling for the training dataset

167 The training dataset is generated over a sample of sites representative of the global  
168 distribution of vegetation types and conditions. For this purpose, the BELMANIP2 set of sites  
169 was used. It corresponds to 420 sites located in relatively flat and homogeneous areas (at a

170 kilometeric resolution over 10 x 10 km<sup>2</sup> domains). BELMANIP2 differs from the original  
 171 BELMANIP set of sites (Baret et al. 2006) by improving the global representativeness and  
 172 homogeneity. Figure 2 shows the distribution of sites and their coordinates are available at  
 173 WWW1. Years 2003 and 2004 were selected to represent the whole seasonality as well as  
 174 significant inter-annual variability.



175

176 **Figure 2. The 420 BELMANIP2 sites used to sample vegetation types and conditions.**

177 The same spatial and temporal supports are required to allow combining several products  
 178 together. The MODIS *LAI* and *FAPAR* products selected were thus re-projected onto the  
 179 cylindrical projection system with 1/112° sampling grid used as a reference for the  
 180 VEGETATION, CYCLOPES and GEOV1 products. Because of the point spread function of  
 181 the several products considered as well as the possible geometrical uncertainties on pixel  
 182 localization, a 3x3 pixels spatial support was used. Note that the algorithm is trained over 3x3  
 183 pixels although it will ultimately apply to single pixels. This scale discrepancy is acceptable  
 184 because of the homogeneity of the BELMANIP2 sites.

185 The temporal sampling used to fuse MODIS and CYCLOPES products will be that of the  
 186 CYCLOPES original products, i.e. dekadal (10 days). It will allow using directly the  
 187 normalized reflectance values derived from VEGETATION based on the CYCLOPES  
 188 preprocessing algorithm (Baret et al. 2007) that will also constitute the GEOV1 temporal  
 189 sampling. The values to be fused at a given GEOV1 dekad were computed as follows: for  
 190 CYCLOPES, the product values corresponding to the GEOV1 dekad are considered; for  
 191 MODIS 8 days products, all the values available within  $\pm 10$  days around the GEOV1 dekad  
 192 are first gathered resulting in a maximum of 27 values (3 MODIS dates times 9 pixels). Note  
 193 that the CYCLOPES 30 days temporal resolution (with Gaussian weighing) is still larger than  
 194 that of the 16 days (2 times 8 days) of the MODIS selected values. However, widening the  
 195 temporal window for MODIS resulted in artifacts on the seasonality. Then the 3x3 aggregated  
 196 values for each GEOV1 dekad are computed only if at least 5 valid individual values (over the  
 197 9 for CYCLOPES and 27 for MODIS) are available. Valid pixels are defined by the main and  
 198 main + saturation QC (Quality Control) for MODIS, and valid input reflectance QC for  
 199 CYCLOPES. For MODIS, because of the relatively large variability observed over time and  
 200 space, values were further filtered using the difference  $\Delta = LAI_{f=90\%} - LAI_{f=70\%}$  between the  
 201  $f=90\%$  and  $f=70\%$  percentiles ( $f$ ) computed over the valid *LAI* values: when  $\Delta > 0.2$

202 corresponding to an unexpectedly large scattering of the data over homogeneous sites and  
203 during a short time period, the cases were rejected. Finally, for the GEOV1 dates fulfilling the  
204 above criteria both for CYCLOPES and MODIS products, the *LAI* and *FAPAR* values  
205 corresponding to the 70% percentile was computed for CYCLOPES and MODIS. This allows  
206 minimizing the influence of possible residual cloud contamination and atmospheric effects  
207 that negatively biased the product values (Chen et al. 2006). However, because of the  
208 homogeneity of the sites and the short time period considered, the values once filtered as  
209 described above, should be closely distributed around the median, i.e. *LAI* or *FAPAR* values at  
210 50% and 70% frequencies should be very close together for a given date and site.

### 211 2.2.3 Fusing the products

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

$$217 \begin{cases} FAPAR_{fused} = FAPAR_{MOD}w_{FAPAR} + FAPAR_{CYC}(1 - w_{FAPAR}) \\ LAI_{fused} = LAI_{MOD}w_{LAI} + LAI_{CYC}(1 - w_{LAI}) \end{cases} \quad [1]$$

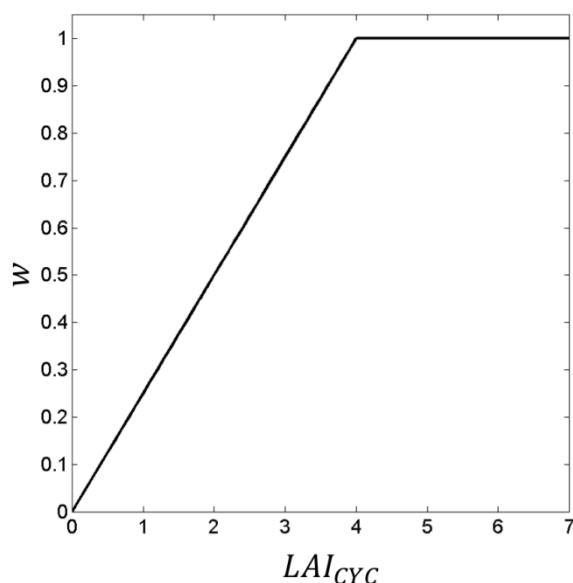
218 Where the subscripts ‘*fused*’, ‘*MOD*’, ‘*CYC*’ correspond respectively to the fused, MODIS and  
219 CYCLOPES products. However, the uncertainties attached to the CYCLOPES and MODIS  
220 products only refer to the theoretical performances, and model assumptions as well as the  
221 structure of the uncertainties were not accounted for. Uncertainties may be also derived from  
222 the comparison with ground measurements as already achieved for MODIS and CYCLOPES.  
223 However, these ground measurements are not very numerous (Camacho et al. 2012; Garrigues  
224 et al. 2008). Further, it is not advisable to use the validation data to calibrate an algorithm in  
225 order to preserve the required independency between the calibration and the validation  
226 processes. For these reasons, the weight used in the fusion between MODIS and CYCLOPES  
227 were based on heuristic arguments.

228 Garrigues et al. (2008) and Weiss et al. (2007) reported that CYCLOPES *LAI* was showing  
229 some saturation for *LAI* values around 4. Conversely, MODIS *LAI* and *FAPAR* values were  
230 generally higher than expected for the very low vegetation amounts (Figure 4). Further, the  
231 MODIS algorithm assigns zero values for *LAI* and *FAPAR* over pixels classified as bare,  
232 which may pose problems in case of misclassification. It was thus proposed to fuse the  
233 products by reducing the contribution of MODIS products for low *LAI* and *FAPAR* values and  
234 enhancing the MODIS contribution for the larger *LAI* and *FAPAR* values as sketched in

235 Figure 3 with  $w_{LAI} = w_{FAPAR} = \min(1, \frac{LAI_{CYC}}{4})$ . The weight,  $w$ , is driven by  $LAI_{CYC}$  (Figure 3)

236 since  $LAI_{CYC}$  appears more stable as compared to MODIS *LAI* (Verger et al. 2011). The  
237 threshold of  $LAI_{CYC}=4$  corresponds to the value when  $LAI_{CYC}$  starts to saturate. The parallel  
238 processing applied to both *LAI* and *FAPAR* (equation 1) is expected to keep a good  
239 consistency between these two variables.





240

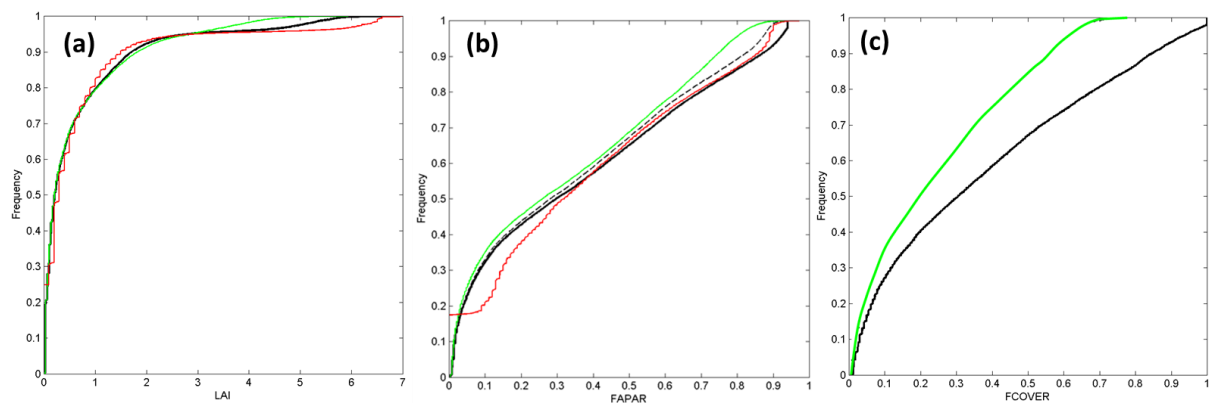
241 **Figure 3. Variation of the weight ( $w$ ) as a function of the  $LAI_{CYC}$  value.**242 **2.2.4 Scaling the fused products**

243 The fused  $FAPAR$  products showed that the maximum values (at 99% percentile) are around  
 244 0.90 (Figure 4b, dashed black line) although the maximum values are expected to be close to  
 245 0.94 (Baret and Guyot 1991) corresponding to full cover dense vegetation with albedo in the  
 246 PAR domain close to 0.06. Therefore, the fused values were scaled according to:

$$247 \quad FAPAR_{scaled} = \frac{0.94}{0.90} FAPAR_{fused} \quad [2]$$

248 The highest  $FCOVER_{CYC}$  value was approximated by the 99% percentile value, i.e.  
 249  $FCOVER_{CYC}(99\%)=0.69$  (Figure 4c). This is in agreement with results obtained by Verger  
 250 (2008) showing that CYCLOPES  $FCOVER$  product was underestimating actual values,  
 251 although being strongly linearly correlated with other regional products including the SAF-  
 252 Land one. It is thus proposed to correct for this systematic underestimation by applying a  
 253 scaling factor. This factor was computed considering the highest  $FCOVER_{CYC}$  values  
 254 observed that should correspond to full coverage ( $FCOVER=1$ ). The 'scaled'  $FCOVER$   
 255 product,  $FCOVER_{scaled}$  used later to train the neural network will thus be computed according  
 256 to:

$$257 \quad FCOVER_{scaled} = \frac{1}{0.69} FCOVER_{CYC} \quad [3]$$



258

259 **Figure 4. Cumulated frequency of CYCLOPES (green solid line), MODIS (red solid**  
 260 **line) fused (dashed black line for *FAPAR*) and scaled (black solid line) *LAI* (a), *FAPAR***  
 261 **(b) and *FCOVER* (c) products as observed over the 420 BELMANIP2 sites during years**  
 262 **2003-2004.**

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

## 272 2.3 Training the neural networks

273 The inputs and outputs of the training data set and the neural network architecture and  
 274 learning process are described in this section.

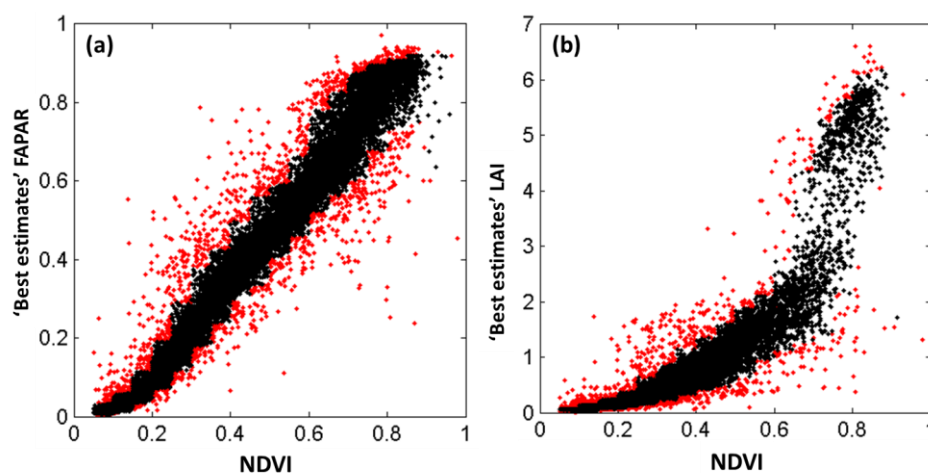
### 275 2.3.1 Inputs

276 The inputs of the neural network correspond to the bidirectional reflectance factor (BRF) as  
 277 measured by VEGETATION aboard SPOT. They correspond to directionally normalized top  
 278 of canopy reflectance in the red, NIR and SWIR bands as derived from the CYCLOPES L3a  
 279 products. The preprocessing steps include cloud screening, atmospheric correction based on a  
 280 climatology of aerosols, and BRDF normalization using a robust fit of Roujean's model  
 281 (Hagolle et al. 2004; Roujean et al. 1992). Details about the processing from L0 (raw signal)  
 282 to L3a can be found in (Baret et al. 2007). However, the original CYCLOPES L3a products  
 283 were reprocessed to benefit from updated radiometric calibration coefficients for the  
 284 VEGETATION sensors. The blue band was not considered here since it brings only little  
 285 extra information on the surface as compared to the red, NIR and SWIR bands, while being  
 286 very sensitive to errors in the atmospheric correction (Bacour et al. 2006). To match the  
 287 spatial support used for the 'best estimates' described earlier, the median was computed over  
 288 the valid pixels within the  $3 \times 3 = 9$  potential ones. The median for each band was preferred  
 289 here, since the use of the 70% percentile for each band would result in possible spectral  
 290 inconsistencies because the sensitivity of reflectance to canopy variables is different between  
 291 bands. Note however that the CYCLOPES L3a products show a good stability over the  
 292 BELMANIP2 sites which are relatively homogeneous as earlier demonstrated by Verger et al.  
 293 (2008).

294 In addition to the BRF in the red, NIR and SWIR bands, the algorithm used the cosine of the  
 295 median value of the sun zenith angle ( $\theta_s$ ) corresponding to the valid observations during the  
 296 30 days window over which the directionally normalized top of canopy reflectance values  
 297 were composited.

### 298 2.3.2 Outputs

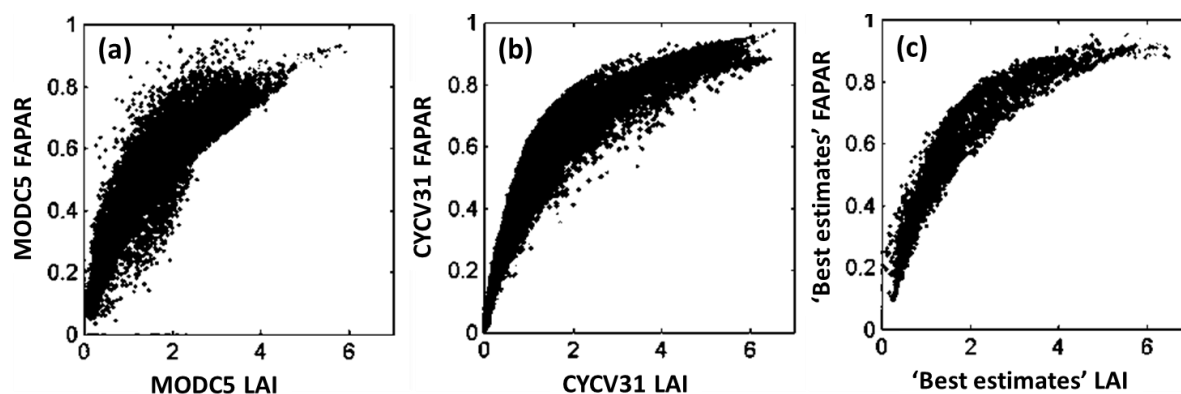
299 The outputs correspond to the targeted *LAI*, *FAPAR* and *FCOVER* variables. To further verify  
 300 that the resulting ‘best estimates’ were consistent, they were plotted against  $NDVI = (BRF_{NIR} -$   
 301  $BRF_{Red}) / (BRF_{NIR} + BRF_{Red})$  values computed from the CYCLOPES BRF products in the red  
 302 ( $BRF_{Red}$ ) and in the near infrared ( $BRF_{NIR}$ ). Cases with  $NDVI < 0.05$  were rejected since these  
 303 values are not expected over bare soil or vegetation pixels (results derived from *NDVI* values  
 304 computed for VEGETATION sensor bands based on a large data base of soil reflectance  
 305 available at WWW2 and described in Liu et al. (2002). Then, for each class of *NDVI* values  
 306 (20 classes over the [0, 1] domain of variation), the cases with *FAPAR* or *LAI* values lower  
 307 (respectively higher) than the 5% percentile (respectively 95%) were rejected. This allowed to  
 308 further improve the consistency of outputs with input reflectance values through *NDVI* as  
 309 attested by Figure 5. Relatively few cases were observed for *LAI* in between 3 and 4. This is  
 310 explained by the fact that forests that are the more likely to show such median to high *LAI*  
 311 values (Weiss et al. 2007) represent only 25% of the global land surfaces. Further, many  
 312 forest sites are frequently covered by clouds in addition to snow cover and poor illumination  
 313 conditions that are frequently observed in winter for the high northern latitudes. Although this  
 314 situation is not ideal, the neural networks should be able to interpolate efficiently between the  
 315 cases available on both sides of this area.



316

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

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



325

326 **Figure 6. Relationship between LAI and FAPAR for CYCLOPES (CYCV31 a), MODIS**  
 327 **(MODC5 b) and ‘best estimates’ (c) as observed over the 420 BELMANIP2 sites during**  
 328 **2003-2004 period (14200 points).**

329 **2.3.3 Neural network architecture and learning process**

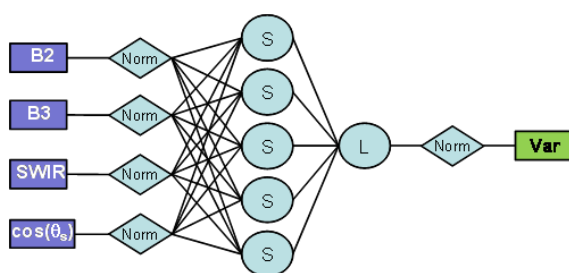
330 The previously described dataset is finally made of 14200 cases where consistent top of  
 331 canopy directionally normalized reflectance values are paired with ‘best estimates’ of LAI,  
 332 FAPAR and FCOVER values. This represents roughly 47% of the total 30200 data potentially  
 333 existing over the 420 sites during the 2003-2004 period (72 dekads). The available data was  
 334 randomly split into a training data set made of 90% of the data available, and a test data set  
 335 (10% of the data) used for testing the hyper-specialization of the training process and  
 336 evaluating the theoretical performances. The large fraction of data used for the training  
 337 process allows getting a better representativeness of surface types and conditions considering  
 338 the limited time period (2 years) and sites considered (420 sites) and the large fraction of  
 339 missing data. The inputs and outputs are normalized to prevent possible numerical problems  
 340 during the training process. Normalization is achieved by scaling between -1 and +1 the range  
 341 of variation of input and output values according to:

342 
$$x_{norm} = 2 \frac{x - x_{min}}{x_{max} - x_{min}} - 1 \quad [4]$$

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

345 A back-propagation network architecture has been used since it proves very efficient in  
 346 similar problems (Baret et al. 2007; Verger et al. 2010a). It is made of one hidden layer of 5  
 347 tangent-sigmoidal neurones, and one output layer made of a single linear neurone. This  
 348 architecture includes 31 coefficients to be adjusted (25 synaptic weights and 6 biases)  
 349 providing more than 400 cases per coefficient to be adjusted. The Levenberg–Marquardt  
 350 optimization algorithm (Ngia and Sjoberg 2000) is used for adjusting the synaptic weights and  
 351 neuron bias to get the best agreement between the output simulated by the network and the  
 352 corresponding value of canopy biophysical variable in the training dataset. The initial values  
 353 of the weights and biases were set to a random value between -1.0 and +1.0. Three networks  
 354 were trained in parallel for each targeted variable with variation in the initial values. The one  
 355 providing the best performances over the test dataset was selected. Results obtained during  
 356 the training process showed that the three parallel networks were performing very similarly,  
 357 indicating a robust training process. A dedicated neural network for each variable was  
 358 preferred here to using a unique network for estimating concurrently the three output variables

359 because the associated architecture is simpler, leading to an easier training process while still  
 360 providing very good consistency between *LAI*, *FAPAR* and *FCOVER* estimated variables,  
 361 similarly to what is observed on the training dataset (Figure 6).

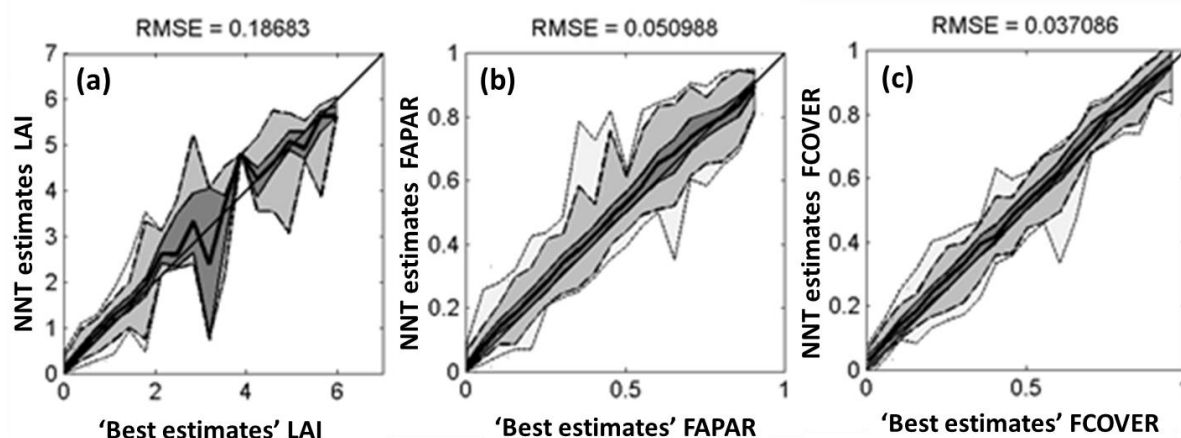


362

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

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

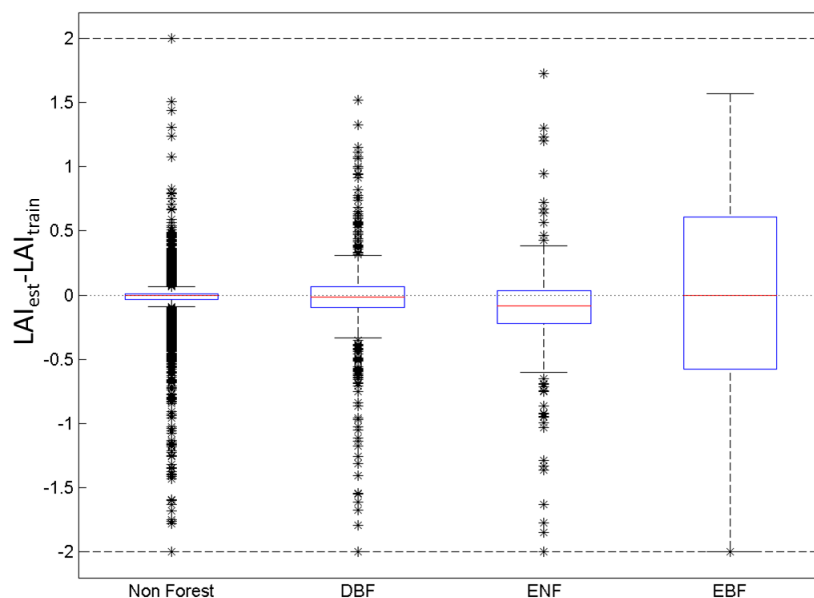
371



372

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

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



385

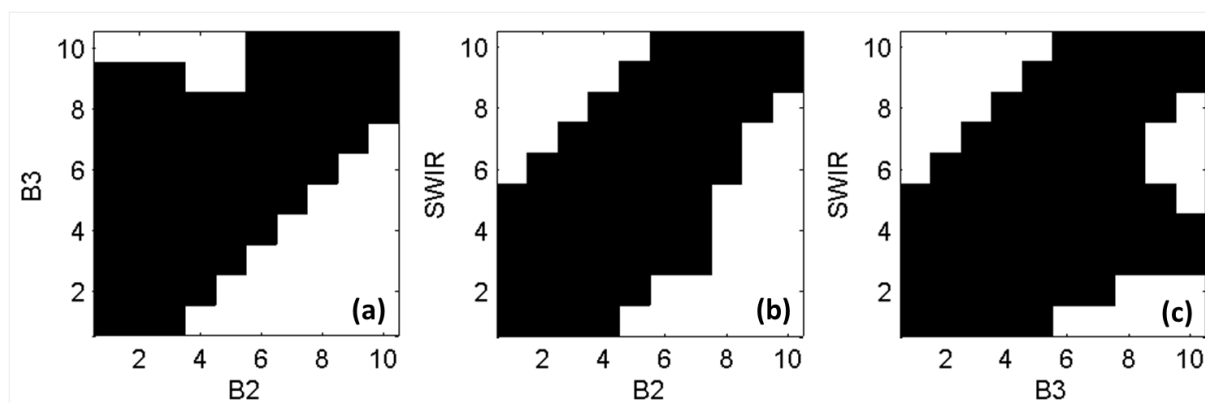
386 **Figure 9. Difference between the LAI estimated from the NNT ( $LAI_{est}$ ) and the ‘best estimates’**  
 387 **LAI value used for the training ( $LAI_{train}$ ) for the four main vegetation types : Evergreen**  
 388 **Broadleaf Forest (EBF), Needle leaf Forest (ENF), Deciduous Broadleaf Forest (DBF) and Non-**  
 389 **Forest. The red line, blue box and black whiskers represent respectively 50%, [25%,75%] and**  
 390 **[5%-95%] percentiles. Outliers are indicated by stars.**

## 391 2.4 Associated uncertainties and quality assessment

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

### 399 2.4.1 Input out of range

400 Since the algorithm is based on a learning machine approach, it is important to verify whether  
 401 the inputs of a given observation keeps within the range of variation of the training dataset  
 402 called here the definition domain. If this condition is not fulfilled, the network will run in  
 403 extrapolation mode, with no warranty about the realism of the outputs. The definition domain  
 404 is limited by the convex hull formed in the BRF feature space by the cases used in the training  
 405 process (Figure 10). For the sake of simplicity and ease of implementation, the 3D feature  
 406 space formed by B2, B3 and SWIR bands was gridded by dividing the range of variation of  
 407 each band (Figure 10 and Table 2) into 10 equally spaced classes. The ensemble of cells  
 408 containing data used for the training form the definition domain. When the input BRF values  
 409 are outside the definition domain, i.e. outside cells containing data used for the training, an  
 410 ‘input out of range’ flag is raised. Note that the sun zenith angle was not included in the  
 411 description of the definition domain. This would have induced increased complexity for  
 412 marginal gain since the definition domains corresponding to several sun zenith angles are  
 413 largely overlapping.



414

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

420 **Table 2. Range of variation of BRDF values (Minimum and maximum) observed in the training**  
 421 **dataset and used to compute the definition domain.**

	Min	Max
<b>B2</b>	0.000	0.429
<b>B3</b>	0.036	0.547
<b>SWIR</b>	0.000	0.648

422 **2.4.2 Output out of range**

423 The physical limits for the three variables are described in Table 3. However, for *LAI*, the  
 424 upper limit is not a physical limit, but a value just slightly higher than the maximum value  
 425 that can be reached by the MODIS and CYCLOPES original products. Because of the several  
 426 sources of uncertainties associated to the inputs, the algorithm calibration process including  
 427 uncertainties attached to the original MODIS and CYCLOPES products, a tolerance is set for  
 428 the extreme values: when the neural network provides biophysical variable estimates outside  
 429 their definition range the value will be always set to the closest bound of the range, i.e. either  
 430 the minimum or the maximum values (Table 3). The product uncertainty value will be also set  
 431 to its maximum value. However, The output status flag is thus raised only when the output is  
 432 outside the output range enlarged by the tolerance values [ $P_{min}^{tol}$ ,  $P_{max}^{tol}$ ] as defined in Table  
 433 3.

434 **Table 3. Minimum, Maximum, Resolution and Tolerance values used to raise the output**  
 435 **out of range flag.**

	Min	Max	Resol.	Tol <sub>min</sub>	Tol <sub>max</sub>
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

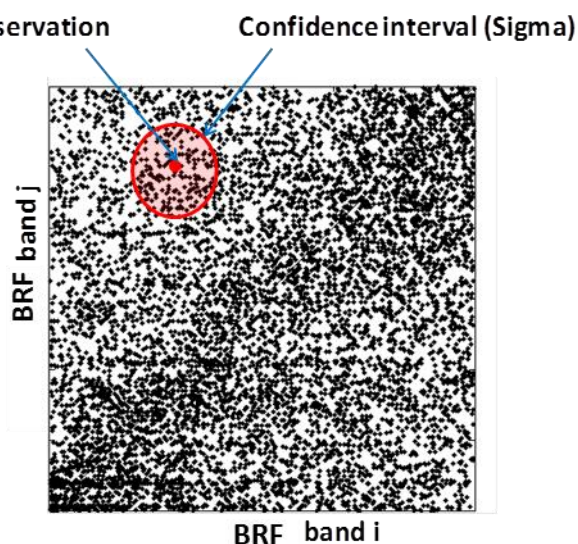
436 **2.4.3 Product uncertainties**

437 The uncertainties associated to each biophysical variable are computed over the training  
 438 dataset: for each case in the training dataset, a RMSE value is computed over the biophysical  
 439 variables that have their corresponding input BRDF values within the uncertainty domain  
 440 (Figure 11). Further, the sun zenith angles selected have also to be within  $\pm 5^\circ$  around the  
 441 direction of the case considered. Uncertainties on reflectance are derived from Baret et al.

442 (2007): the standard deviation,  $\sigma$ , is computed for each waveband according to the uncertainty  
 443 model reported in Table 4:

444  $\sigma = a + b \cdot \text{BRF}$  [5]

445 Where  $a$  and  $b$  are coefficients.



446

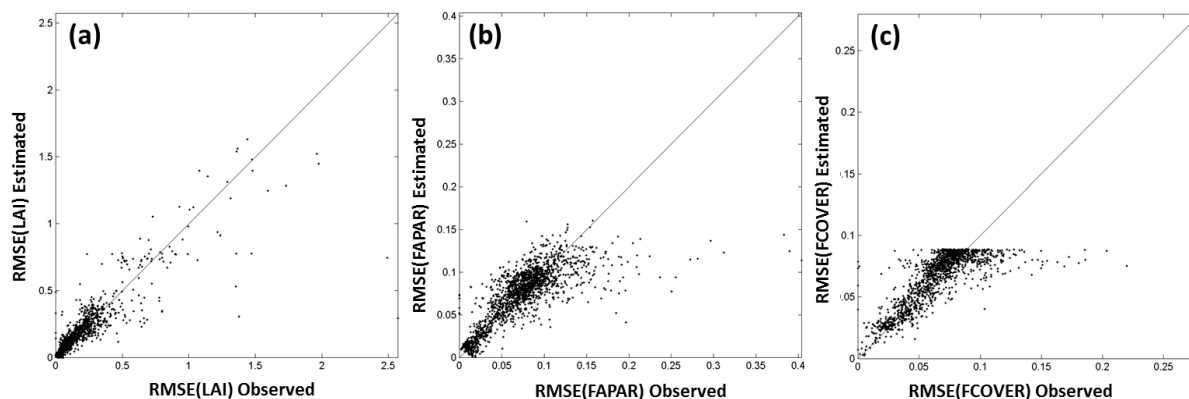
447 **Figure 11. Scheme showing how the uncertainties attached to the products were**  
 448 **computed**

449 **Table 4. Values of the uncertainty model used for the input BRFs. From Baret et al.**  
 450 **(2007).**

	<b>B2</b>	<b>B3</b>	<b>SWIR</b>
<b>a</b>	0.005	0.003	0.005
<b>b</b>	0.05	0.03	0.03

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





459

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

### 463 3 Operational production and dissemination

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

474 Before its integration in the operational processing center, the processing line has been run to  
 475 generate 2 years (2003–2004) of GEOV1 demonstration products used to perform the  
 476 validation exercise presented in Camacho et al. (2012). They have also been supplied to  
 477 Geoland2 users who checked the consistency with their requirements before starting the  
 478 operational production of real time and the whole 12 years time series processing of  
 479 observations. The processing line has been optimized to ensure operational management with  
 480 permanent quality control while benefiting from the variable computation resources available  
 481 for parallel processing.

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

### 488 4 CONCLUSION

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

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

500 The approach undertaken here was not calibrated using ground validation measurements,  
501 preserving the required independency between calibration and validation processes. However,  
502 this led to use a more pragmatic approach based on heuristic arguments that explains the  
503 sometimes subjective selection of criteria used in the development of the GEOV1 algorithm  
504 and several alternative solutions could have been proposed. Nevertheless the GEOV1  
505 products resulting from the largely inductive approach undertaken yield robust, consistent and  
506 accurate estimates of these key biophysical variables as demonstrated in Camacho et al.  
507 (2012). The validation shown in the companion paper demonstrates that GEOV1 significantly  
508 improved the performances of currently existing products, both regarding accuracy and spatial  
509 as well as temporal consistency. It provides quantified estimates of uncertainties, although  
510 these are simply derived from the training dataset that reflects mostly the sensitivity of the  
511 product to input reflectance values. Further investigations should be directed towards a better  
512 quantification of the uncertainties, including ‘model’ assumptions since both MODIS and  
513 CYCLOPES are based on radiative transfer model inversion.

514 A clear understanding of the actual definition of the retrieved variables and their consistency  
515 with application requirements is one of the key aspects to consider when deriving biophysical  
516 variables from remote sensing observations. For *FCOVER*, no particular question is raised  
517 since the *FCOVER* definition is simple and clear if restricted to the green vegetation elements  
518 including overstory and understory. For *FAPAR*, because of the good consistency between  
519 MODIS and CYCLOPES definitions, the output product is also well defined: black sky  
520 *FAPAR* (green element including over and understory) at 10:15 (actually between 10:00 and  
521 10:30). For *LAI*, the question is more complex since the original MODIS and CYCLOPES  
522 *LAI* products are defined differently, at least with regards to the assumptions embedded in the  
523 radiative transfer models used. For MODIS, clumping at the tree scale is accounted for, but  
524 not at the shoot, nor at the landscape scales. For CYCLOPES, clumping is accounted for at  
525 the landscape scale only. However, its effect will be significant mainly for the larger *LAI*  
526 values as discussed by Garrigues et al. (2006). Therefore, GEOV1 *LAI* product will  
527 marginally account for landscape clumping as demonstrated by the good consistency between  
528 MODIS and CYCLOPES *LAI* products for low to medium *LAI* values (Camacho et al. 2012).  
529 However the special case of savannas with relatively low *LAI* values and significant clumping  
530 (Ryu et al. 2010). In this case, the radiation interaction between plants that are separated by  
531 significant distances is minimal, and clumping described at the plant level in MODIS may be  
532 approximated at the landscape scale as in CYCLOPES. Clumping at the shoot scale will not  
533 be accounted for in GEOV1 *LAI* product since neither MODIS, nor CYCLOPES account for  
534 it. Conversely, clumping at the plant scale was accounted mainly for forests and savannas in  
535 the original MODIS product but not in the CYCLOPES ones. However, clumping is mainly  
536 observed for the higher *LAI* values (overlap between leaves within the tree volume when the  
537 number of leaves increases) (Rochdi 2003). Therefore, clumping at the plant scale as  
538 described in MODIS products should be partly preserved in the GEOV1 *LAI* product. The  
539 validation achieved in the companion paper (Camacho et al. 2012) over a limited number of

540 sites demonstrates that the GEOV1 *LAI* products are close to the values estimated from  
541 indirect techniques and that include the clumping effect.

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

550 Although the development of this series of biophysical products should represent an important  
551 step towards a more effective use of remote sensing observations, improvements are expected  
552 mainly through several aspects. Clouds constitute obviously the major limitation of optical  
553 systems that suffer from large areas and periods without data. Recent studies demonstrate that  
554 the fusion between several sensors improves data continuity (Hagolle et al. 2005; Verger et al.  
555 2010b; Yang et al. 2006b). Alternative approaches based on enhanced time series processing  
556 may help removing outliers, filling gaps due to missing observations and smoothing the  
557 temporal profiles (Kandasamy et al. 2013; Verger et al. 2012). However, the main limitation  
558 in such global products comes mainly from the little a priori information available and  
559 required to regularize the inversion process (Combal et al. 2002). The use of global  
560 classification is likely to be insufficient because of its limited accuracy (Defourny et al. 2009;  
561 Herold et al. 2008; Yang et al. 2006a) and because the variability in canopy architecture and  
562 optical properties within a given class is probably as large as between classes when the  
563 seasonal variability is considered. Further, the often mixed nature of kilometeric pixels poses  
564 both a scaling issue and the difficulty to identify the several co-existing surface patches.  
565 These problems call for an improved spatial resolution that will allow resolving most of the  
566 vegetation patches and will authorize identifying the corresponding vegetation type from the  
567 past observations and use it as prior information. Such systems are currently being available  
568 with hectometric resolutions, such as the PROBA-V (300m daily), Sentinel-3 (300 m every 2  
569 days), VIIRS (370m daily). However, decametric systems such as Sentinel 2 or LDCM in  
570 combination with the previous hectometric ones would probably provide the most efficient  
571 observation system.

572

### 573 *Acknowledgments*

574 The research leading to these results has received funding from the European Community's  
575 Seventh Framework Program (FP7/2007-2013) under grant agreement n°218795.

### 576 *WWW references*

577 **WWW1:** <http://calvalportal.ceos.org/cvp/web/guest/olive>.

578 **WWW2:** [https://www4.paca.inra.fr/emmah\\_eng](https://www4.paca.inra.fr/emmah_eng)

579 **WWW3:** <http://postel.mediasfrance.org>

580 **WWW4:** <http://www.geoland2.eu>

### 581 *References*

- 582 Avissar, R., & Pielke, R.A. (1989). A parameterization of heterogeneous land surfaces for  
583 atmospheric numerical models and its impact on regional meteorology. *Monthly*  
584 *Weather Review*, 117, 2113-2136
- 585 Bacour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation  
586 of LAI, fAPAR, fCover and LAIxCab, from top of canopy MERIS reflectance data:  
587 principles and validation. *Remote Sensing of Environment*, 105, 313-325
- 588 Baret, F., Bacour, C., Béal, D., Weiss, M., Bruniquel, V., & Regner, P. (2005). Algorithm  
589 Theoretical Basis Document for MERIS Top of Atmosphere Land Products  
590 (TOA\_VEG). In (p. 32). Avignon: INRA-CSE
- 591 Baret, F., Clevers, J.G.P.W., & Steven, M.D. (1995). The robustness of canopy gap fraction  
592 estimates from red and near infrared reflectances: a comparison of approaches. *Remote*  
593 *sensing of environment*, 54, 141-151
- 594 Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR  
595 assessment. *Remote Sensing of Environment*, 35, 161-173
- 596 Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Weiss, M.,  
597 Samain, O., Roujean, J.L., & Leroy, M. (2007). LAI, fAPAR and fCover CYCLOPES  
598 global products derived from VEGETATION. Part 1: Principles of the algorithm.  
599 *Remote Sensing of Environment*, 110, 275-286
- 600 Baret, F., Morissette, J., Fernandes, R., Champeaux, J.L., Myneni, R., Chen, J., Plummer, S.,  
601 Weiss, M., Bacour, C., Garrigue, S., & Nickeson, J. (2006). Evaluation of the  
602 representativeness of networks of sites for the global validation and inter-comparison of  
603 land biophysical products. Proposition of the CEOS-BELMANIP. *IEEE Transactions on*  
604 *Geoscience and Remote Sensing*, 44, 1794-1803
- 605 Baret, F., Nightingale, J., Garrigues, S., & Nickeson, J. (2009). Report on the CEOS Land  
606 Product Validation sub-group meeting Missoula, Montana, June 15th 2009. *The Earth*  
607 *Observer*, 21, 26-30
- 608 Camacho-de Coca, F., Jiménez-Muñoz, J.C., Martínez, B., Bicheron, P., Lacaze, R., & Leroy,  
609 M. (2006). Prototyping of the fCover product over Africa based on existing  
610 CYCLOPES and JRC products for VGT4Africa. In J. Sobrino (Ed.), *Proceedings of the*  
611 *RAQRS'II 2nd International Symposium on Recent Advances in Remote Sensing*, (pp.  
612 722-726.). Valencia (Spain): Univ. Valencia
- 613 Camacho, F., Cernicharo, J., Lacaze, R., Baret, F., & Weiss, M. (2012). GEOV1: LAI, FAPAR  
614 Essential Climate Variables and FCover global time series capitalizing over existing  
615 products. Part 2: Validation and inter-comparison with reference products. *Remote*  
616 *Sensing of Environment*, submitted
- 617 Carlson, T.N., & Ripley, D.A. (1997). On the relationship between NDVI, fractional cover  
618 and leaf area index. *Remote Sensing of Environment*, 62, 241-252
- 619 Chen, J.M., & Black, T.A. (1992). Defining leaf area index for non-flat leaves. *Plant, Cell and*  
620 *Environment*, 15, 421-429

- 621 Chen, J.M., Deng, F., & Chen, M. (2006). Locally adjusted cubic spline capping for  
 622 reconstructing seasonal trajectories of a satellite-derived surface parameter. *IEEE*  
 623 *Transactions on Geoscience and Remote Sensing*, 44, 2230-2238
- 624 Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R.,  
 625 Knyazikhin, Y., & Wang, L. (2002). Retrieval of canopy biophysical variables from bi-  
 626 directional reflectance data. Using prior information to solve the ill-posed inverse  
 627 problem. *Remote Sensing of Environment*, 84, 1-15
- 628 De Kauwe, M.G., Disney, M.I., Quaife, T., Lewis, P., & Williams, M. (2011). An assessment  
 629 of the MODIS collection 5 leaf area index product for a region of mixed coniferous  
 630 forest. *Remote Sensing of Environment*, 115, 767
- 631 Defourny, P., Schouten, L., Bartalev, S., Bontemps, S., Cacetta, P., de Wit, A.J.W., di Bella,  
 632 C., Gerard, B., Giri, C., Gond, V., Hazeu, G.W., Heinimann, A., Herold, M., Knoop, J.,  
 633 Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mucher, S.A., Nonguierma, A., Stibig,  
 634 H.J., Van Bogaert, E., Vancutsem, C., Bicheron, P., Leroy, M., & Arino, O., . (2009).  
 635 Accuracy Assessment of a 300 m Global Land Cover Map: the GlobCover Experience.  
 636 In, *Proceedings of the 33rd International Symposium on Remote Sensing of*  
 637 *Environment*. Stresa (Italy)
- 638 Deng, F., Chen, J.M., Chen, M., & Pisek, J. (2006). Algorithm for global leaf area index  
 639 retrieval using satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*,  
 640 44, 2219-2229
- 641 Ganguly, S., Samanta, A., Schull, M.A., Shabanov, N.V., Milesi, C., Nemani, R.R.,  
 642 Knyazikhin, Y., & Myneni, R.B. (2008a). Generating vegetation leaf area index earth  
 643 system data record from multiple sensors. Part 1: Theory. *Remote Sensing of*  
 644 *Environment*, 112, 4318-4332
- 645 Ganguly, S., Schull, M.A., Samanta, A., Shabanov, N.V., Milesi, C., Nemani, R.R.,  
 646 Knyazikhin, Y., & Myneni, R.B. (2008b). Generating vegetation leaf area index earth  
 647 system data record from multiple sensors. Part 2: Evaluation. *Remote Sensing of*  
 648 *Environment*, 112, 4333
- 649 Garrigues, S., Allard, D., Baret, F., & Weiss, M. (2006). Influence landscape spatial  
 650 heterogeneity on the non-linear estimation of leaf area index from moderate spatial  
 651 resolution remote sensing data. *Remote Sensing of Environment*, 105, 286-298
- 652 Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., Fernandes, R.,  
 653 Plummer, S., Shabanov, N.V., Myneni, R., & Yang, W. (2008). Validation and  
 654 Intercomparison of Global Leaf Area Index Products Derived From Remote Sensing  
 655 Data. *Journal of Geophysical Research*, 113
- 656 GCOS (2006). GCOS-107. Supplemental details to the satellite based component of the  
 657 "implementation plan for the global observing system for climate in support of the  
 658 UNFCCC". In (p. 103). Geneve (Switzerland): GCOS/WMO
- 659 Gobron, N., Pinty, B., Aussedat, O., Chen, J.M., Cohen, W.B., Fensholt, R., Gond, V.,  
 660 Huemmrich, K.F., Lavergne, T., Melin, F., Privette, J.L., Sandholt, I., Taberner, M.,  
 661 Turner, D.P., Verstraete, M.M., & Widlowski, J.L. (2006). Evaluation of fraction of

- 662 absorbed photosynthetically active radiation products for different canopy radiation  
 663 transfer regimes: Methodology and results using Joint Research Center products derived  
 664 from SeaWiFS against ground-based estimations. *Journal of Geophysical Research-*  
 665 *Atmospheres*, 111
- 666 Gutman, G.G. (1991). Vegetation indices from AVHRR data: an update and future prospects. ,  
 667 *Remote Sensing of Environment*, 35, 121-136
- 668 Hagolle, O., Lobo, A., Maisongrande, P., Cabot, F., Duchemin, B., & De Pereyra, A. (2005).  
 669 Quality assessment and improvement of temporally composited products of remotely  
 670 sensed imagery by combination of VEGETATION 1 and 2 images. *Remote Sensing of*  
 671 *Environment*, 94, 172-186
- 672 Hagolle, O., Nicolas, J.-M., Fougnie, B., Cabot, F., & Henry, P. (2004). Absolute calibration  
 673 of VEGETATION derived from an interband method based on the Sun glint over ocean.  
 674 *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1472 -1481
- 675 Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., & Schmullius, C. (2008). Some  
 676 challenges in global land cover mapping: An assessment of agreement and accuracy in  
 677 existing 1 km datasets. *Remote Sensing of Environment*, 112, 2538
- 678 Kandasamy, S., Baret, F., Verger, A., Neveux, P., & Weiss, M. (2013). A comparison of  
 679 smoothing and gap filling methods for time series of remote sensing observations.  
 680 Application to MODIS LAI products. *Biogeosciences*, 9, 17053-17097
- 681 Kraus, T. (2008). Ground-based Validation of the MODIS Leaf Area Index Product for East  
 682 African Rain Forest Ecosystems. In. Erlangen: Universitätsbibliothek der Universität  
 683 Erlangen-Nürnberg
- 684 Lacaze, R., Atzberger, C., Bartholomé, E., Combal, B., Calvet, J.C., Lefèvre, V., & Olsson, B.  
 685 (2009). BioPar User Requirements. In. Toulouse
- 686 Liu, G.R., Lin, T.H., & Kuo, T.H. (2002). Estimation of aerosol optical depth by applying the  
 687 optimal distance number to NOAA AVHRR data. *Remote Sensing of Environment*, 81,  
 688 247-252
- 689 McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., &  
 690 Nilsson, S. (2009). Satellite-based terrestrial production efficiency modeling. *Carbon*  
 691 *Balance and Management*, 4:8
- 692 McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., &  
 693 Nilsson, S. (2010). Comparison of four global FAPAR datasets over Northern Eurasia  
 694 for the year 2000. *Remote Sensing of Environment*, 114, 941
- 695 Morissette, J., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J., Garrigues, S., Shabanov, N.,  
 696 Weiss, M., Fernandes, R., Leblanc, S., Kalacska, M., Sanchez-Azofeifa, G.A., Chubey,  
 697 M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, D., Lewis,  
 698 T., Iames, J., Colombo, R., Meroni, M., Busetto, L., Cohen, W., Turner, D., Warner, D.,  
 699 Petersen, G.W., Seufert, G., & Cook, R. (2006). Validation of global moderate resolution  
 700 LAI Products: a framework proposed within the CEOS Land Product Validation  
 701 subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1804-1817

- 702 Ngia, L.S.H., & Sjoberg, J. (2000). Efficient training of neural nets for nonlinear adaptive  
703 filtering using a recursive Levenberg-Marquardt algorithm. *IEEE Transactions on*  
704 *Signal Processing*, 48, 1915-1927
- 705 Rochdi, N. (2003). Un modèle générique d'agrégation des feuilles dans un couvert végétal:  
706 application à la simulation du transfert radiatif. In (p. 169). Paris (France): Institut  
707 National Agronomique
- 708 Roujean, J.L., Leroy, M., & Deschamps, P.Y. (1992). A bidirectional reflectance model of the  
709 Earth's surface for the correction of remote sensing data. *Journal of Geophysical*  
710 *Research*, 97, 20455-20468
- 711 Ryu, Y., Sonnentag, O., Nilson, T., Vargas, R., Kobayashi, H., Wenk, R., & Baldocchi, D.D.  
712 (2010). How to quantify tree leaf area index in an open savanna ecosystem: A multi-  
713 instrument and multi-model approach. *Agricultural and Forest Meteorology*, 150, 63-76
- 714 Scurlock, J.M.O., Asner, G.P., & Gower, S.T. (2001). Worldwide Historical Estimates and  
715 Bibliography of Leaf Area Index, 1932-2000. In. Oak Ridge, Tennessee, U.S.A.: Oak  
716 Ridge National Laboratory
- 717 Sprintsin, M., Karnieli, A., Berliner, P., Rotenberg, E., Yakir, D., & Cohen, S. (2009).  
718 Evaluating the performance of the MODIS Leaf Area Index (LAI) product over a  
719 Mediterranean dryland planted forest. *International Journal of Remote Sensing*, 30,  
720 5061 - 5069
- 721 Verger, A. (2008). Anàlisi comparativa d'algorismes operacionals d'estimació de paràmetres  
722 biofísics de la coberta vegetal amb teledetecció. In, *Departament de física de la terra i*  
723 *termodinàmica* (p. 277). Valencia (Spain): Universitat de Valencia
- 724 Verger, A., Baret, F., García-Haro, F.J., Camacho, F., & Meliá, J. (2010a). Consistency of  
725 vegetation estimates from SEVIRI/Meteosat observations and operational algorithms. In  
726 J. Sobrino (Ed.), *Third International Symposium on Recent Advances in Quantitative*  
727 *Remote Sensing*. Torrent (Spain)
- 728 Verger, A., Baret, F., & Weiss, M. (2008). Performances of neural networks for deriving LAI  
729 estimates from existing CYCLOPES and MODIS products. *Remote Sensing of*  
730 *Environment*, 112, 2789-2803
- 731 Verger, A., Baret, F., & Weiss, M. (2010b). Fusion of MODIS and VEGETATION  
732 observations for improved consistency and continuity of LAI product time series. In J.  
733 Sobrino (Ed.), *Third International Symposium on Recent Advances in Quantitative*  
734 *Remote Sensing*. Torrent (Spain)
- 735 Verger, A., Baret, F., & Weiss, M. (2011). A multisensor fusion approach to improve LAI  
736 time series *Remote Sensing of Environment*, 115, 2460-2470
- 737 Verger, A., Baret, F., Weiss, M., Kandasamy, S., & Vermote, E., F. (2012). The CACAO  
738 method for smoothing, gap filling and characterizing anomalies in satellite time series.  
739 *IEEE Transactions on Geoscience and Remote Sensing*

- 740 Verger, A., Martínez, B., Camacho-de Coca, F., & García-Haro, F.J. (2009). Accuracy  
741 assessment of fraction of vegetation cover and leaf area index estimates from pragmatic  
742 methods in a cropland area. *International Journal of Remote Sensing*, 30, 2685 - 2704
- 743 Vermote, E., Justice, C., Csiszar, I., Eidenshink, J., Mynemi, R., Baret, F., Masuoka, E., &  
744 Wolfe, R. (2010). A terrestrial surface climate data record for global change studies. In  
745 J. Sobrino (Ed.), *Third International Symposium on Recent Advances in Quantitative*  
746 *Remote Sensing*. Torrent (Spain)
- 747 Weiss, M., Baret, F., Garrigues, S., Lacaze, R., & Bicheron, P. (2007). LAI, fAPAR and  
748 fCover CYCLOPES global products derived from VEGETATION. part 2: Validation  
749 and comparison with MODIS Collection 4 products. *Remote Sensing of Environment*,  
750 110, 317-331
- 751 Yang, W., Huang, D., Tan, B., Stroeve, J.C., Shabanov, N., Knyazikhin, Y., Nemani, R., &  
752 Mynemi, R. (2006a). Analysis of leaf area index and fraction of PAR absorbed by  
753 vegetation products from the Terra MODIS sensor: 2000-2005. *IEEE Transactions on*  
754 *Geoscience and Remote Sensing*, 44, 1829-1842
- 755 Yang, W., Shabanov, N.V., Huang, D., Wang, W., Dickinson, R.E., Nemani, R.R., Knyazikhin,  
756 Y., & Myneni, R.B. (2006b). Analysis of leaf area index products from combination of  
757 MODIS Terra and Aqua data. *Remote Sensing of Environment*, 104, 297-312
- 758 Yuan, H., Dai, Y., Xiao, Z., Ji, D., & Shanguan, W. (2011). Reprocessing the MODIS Leaf  
759 Area Index products for land surface and climate modelling. *Remote Sensing of*  
760 *Environment*, 115, 1171
- 761
- 762