Evaluation of Agreement Between Space Remote Sensing SPOT-VEGETATION fAPAR Time Series

Michele Meroni, Clement Atzberger, Christelle Vancutsem, Nadine Gobron, Member, IEEE, Frédéric Baret, Roselyne Lacaze, Herman Eerens, and Olivier Leo

Abstract—Satellite-derived time series of the fraction of absorbed photosynthetically active radiation (fAPAR) are widely used to monitor vegetation dynamics and to detect vegetation anomalies. Several global data sets are available for this purpose. They are produced using different algorithms and/or satellite sensors. This paper compares and analyzes three multitemporal fAPAR data sets derived from SPOT-VEGETATION instrument by explicitly distinguishing between spatial and temporal agreement. The first two data sets are currently used by the Joint Research Centre—Monitoring Agricultural Resources Unit (JRC-MARS) for operational yield forecasting and food security assessments. The third time series (named GEOV1) is from a new processing algorithm developed within the European FP7 Geoland2 project. The comparative analysis was conducted for the years 2003 and 2004 over three 10° × 10° regions with different eco-climatic characteristics (Niger, Brazil, and France). Our study revealed that GEOV1 fAPAR estimates were systematically higher than those of JRC-MARS. The spatial analysis showed moderate to high agreement between data sets with specific seasonality in the three study regions. The temporal agreement showed spatial (and land cover-related) variability spanning from very low to almost perfect. Large differences were observed in regions and periods with large cloud occurrence where GEOV1 provides more reliable and smooth temporal profiles due to improved cloud screening and longer compositing periods. Other sources of disagreement between data sets were identified in differences in the fAPAR retrieval algorithm definitions.

Index Terms—CYCLight, fraction of absorbed photosynthetically active radiation (fAPAR), GEOV1-Geoland2, JRC-FAPAR, multitemporal remote sensing (RS) data, SPOT-VEGETATION (VGT).

I. INTRODUCTION

SPACE remote sensing (RS) data and derived products are used operationally by various institutions for monitoring natural vegetation and crops at regional to continental scales. For this purpose, multitemporal long-term archives of RS data are collected and analyzed using dedicated processing chains. In many operational monitoring and early warning systems, the well-known normalized difference vegetation index (NDVI) [1] is employed as a proxy of vegetative status and/or to derive anomaly indicators (deviation from a reference period or long-term average, e.g., [2]). Nevertheless, a growing body of knowledge [3] indicates that the exploitation of RS products that allow a direct access to vegetation key properties may be better suited for environmental applications. The fraction of absorbed photosynthetically active radiation (fAPAR) in the wavelength range 400–700 nm is one of these key variables that present several advantages compared to NDVI as it acts as an integrated indicator of the status and health of vegetation and plays a major role in modeling the primary productivity of the phytosphere [4], [5]; it is also quite robust to changes in the spatial resolution of the observations [6]. Compared to a radiative transfer state variable such as leaf area index (LAI), fAPAR is a flux ratio and has the advantage that its retrieval accuracy is less dependent on knowledge regarding the 3-D structure of the canopy [7], [8].

As a consequence of the relevance of fAPAR for environmental applications, direct validation, assessment, and comparison studies are needed to evaluate the reliability of fAPAR products. In fact, in recent years, the availability of biophysical products derived from RS moderate-resolution sensors has increased, and potential users are now faced with a wide choice of data products and sources, e.g., MODIS NASA [9], SPOT VEGETATION (VGT), SeaWIFS or MERIS Joint Research Centre’s (JRC’s) products [10], VGT Cyclopes [11], and Globsarbon [12]. A rational selection of the most appropriate product could be driven by several factors: whereas spatial and temporal resolution may represent the primary criteria, the availability and accessibility of the required spatial coverage and temporal archive also play a role. The quality of the data sets, assessed for instance through validation against ground measurements, should also be of the highest importance [13].

Unfortunately, in situ measurement activities appear to be limited and, in most cases, are focused on LAI (e.g., [14]–[17]) rather than fAPAR. Moreover, direct validation of coarse-resolution satellite data against point-like field measurements is a nontrivial task [15] with a number of associated...
uncertainties [14]. As a result, validation of RS fAPAR products in different eco-climatic regions of the earth is currently limited to a restricted number of sites distributed globally [18], [19] or to regional exercises [10], [20], [21].

As an alternative (and complement) to direct validation, comparative analysis has also been performed for selected geographical regions to assess the consistency of different fAPAR products (e.g., [22]–[25]). Comparison between existing products yields important information regarding the feasibility of creating historical fAPAR records combining different sensors [26]–[28]. In addition, when users have already set up an operational processing chain in selected geographical settings based on a given product, there is a natural inertia to adopt a new one, and it becomes more relevant to assess the added value of the new product, i.e., its relative performance with respect to the one being employed and for which an extensive knowledge of the pros and cons has been developed. Therefore, the comparison performed targeting the user’s geographical areas of interest can provide an important assessment of the new product and highlight deficiencies in the employed one, improving our ability to effectively utilize new data sources.

In this manuscript, we focus on such user assessment by comparing the new fAPAR global GEOV1 product developed within the framework of the Geoland2 project (http://www.geoland2.eu) with the fAPAR products currently used by the Monitoring Agricultural ResourceS (MARS) Unit of the JRC for agriculture monitoring (http://www.marsop.info/marsop3). Products are derived from the same sensor, i.e., VGT, so that the differences are restricted to the processing chains and fAPAR retrieval algorithms.

The method used for the comparison capitalizes on previous studies of image comparison [26], [29], [30] and extends these concepts to multitemporal RS data. The novelty is that, in addition to analyzing the “overall” agreement between paired data sets, we explicitly address the seasonality of their spatial coherence and the spatial variability of their temporal coherence. Analyzing the two dimensions allows assessing the consistency of products with respect to the capture of spatial and seasonal patterns. Both information types are of utmost importance in vegetation monitoring programs. The framework of analysis presented in this study may also serve as guideline/protocol for similar studies in the future.

II. METHODS

For the comparison of the variability of two RS image time series (e.g., fAPAR) one can visualize such time series as 3-D data cubes where X and Y represent the geographical dimensions and Z represents time (Fig. 1). To get a measure of “overall agreement,” first some key statistics extracted from the two data sets (e.g., mean, standard deviation, analysis of data distribution, etc.) are computed and compared. In addition, scatterplots and associated statistics (e.g., $R^2$, linear regression coefficients, etc.) are produced. This can be done by matching the observations of the data cubes according to their 3-D coordinates.

The following assessments explicitly focus on the comparison of spatial and temporal variability; in the first case (Fig. 1, left panel), the analysis is performed image by image whereas in the latter case it is performed pixel by pixel (Fig. 1, right panel). The analysis of spatial variability leads to a temporal profile of a given statistic. On the contrary, in most cases, the analysis of temporal variability leads to maps. For both dimensions, land cover maps are used for stratification purposes. The concepts are shown in Fig. 1 using the coefficient of determination ($R^2$) as an example.

If the spatial variability of two data sets is compared (Fig. 1, left panel), individual layers paired by acquisition time (e.g., day, compositing date) are extracted from the time series and stratified by land cover. In the three-class example of Fig. 1, this yields three vectors of fAPAR values for the first data sets and three vectors for the second one. Thus, for each land cover class (a to c), the statistic (e.g., $R^2$) can be calculated. If the process is repeated for all time steps, the statistic can be plotted as a function of time as shown at the bottom of Fig. 1 (left panel).
TABLE I
SUMMARY OF METRICS USED FOR IMAGE COMPARISON

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data distribution</strong></td>
<td></td>
</tr>
<tr>
<td>Empirical Cumulative Distribution Function (ECDF)</td>
<td>ECDF graphs are used to depict and visually compare data distributions. Difference between two ECDFs $F(x)$ and $G(x)$ are quantified by the Kolmogorov-Smirnov statistics: $D_{KS} = \max</td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td></td>
</tr>
<tr>
<td>Scatterplots, Ordinary Least Square (OLS) regression</td>
<td>Traditionally used to describe the correlation and relationship between datasets.</td>
</tr>
<tr>
<td>Geometric Mean Functional Relationship (GMFR) regression analysis</td>
<td>GMFR is preferred to OLS for exploring the relationship between two datasets when both are subjected to error of equal or unknown intensity. GMFR model: $Y = a + bX$, with $b = \pm \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$ and $a = \bar{Y} - b\bar{X}$ where $\bar{X}$ and $\bar{Y}$ are the mean value of $X$ and $Y$, respectively. The sign of $b$ is given by the sign of correlation coefficient.</td>
</tr>
<tr>
<td><strong>Agreement</strong></td>
<td></td>
</tr>
<tr>
<td>Agreement Coefficient (AC, [29])</td>
<td>Used to assess the agreement between two images, non-dimensional, bounded (0-1 for no to perfect agreement), symmetric (no preference to one dataset). $AC = 1 - \left( \frac{SSD}{SPOD} \right)$ where $SSD$ is the sum of squared differences and $SPOD$ the sum of potential differences: $SSD = \sum_{i=1}^{n} (X_i - Y_i)^2$, $SPOD = \sum_{i=1}^{n} \left[ (X_i - \bar{X}) + (Y_i - \bar{Y}) \right] \left[ (X_i - \bar{X}) - (Y_i - \bar{Y}) \right]$.</td>
</tr>
<tr>
<td>Root Mean Square Difference (RMSD) and partitioning into systematic and unsystematic components [26]</td>
<td>The overall RMSD is partitioned into its systematic and unsystematic percentages. The mean square difference (MSD) can be partitioned into the systematic mean product-difference ($MPD_s$) and unsystematic mean product-difference ($MPD_u$): $MPD_s = \frac{1}{n} \sum_{i=1}^{n} \left( (X_i - \hat{X}_i)(Y_i - \hat{Y}_i) \right)$, $MPD_u = MSD - MPD_s$, where $\hat{X}_i$ and $\hat{Y}_i$ are the GMFR model predictions. The systematic component between two datasets can be interpreted as a regularized bias due to known or discoverable factors while the unsystematic element is a random component caused by image noise or unknown factors. The systematic difference can in principle be removed by regression analysis.</td>
</tr>
</tbody>
</table>

When the temporal variability of two data sets is compared (Fig. 1, right panel) the processing is carried out pixel by pixel. First, the observations of a given pixel position are extracted from both data sets, i.e., profiles shown as gray ($Z_1$) and red ($Z_2$) vectors. These vectors can then be mutually compared, and statistics can be computed (here $R^2$). When the process is repeated for all pixel positions, it is convenient to show the resulting statistic in the form of a map. The map can be further analyzed, for example by plotting the frequency distributions of the $R^2$ values per land cover class.

Of the plethora of existing statistics used to quantify RS data agreement (for a review see [29]) we focused on a suite of
The analysis is performed on dekadal (here defined as 10-day period) fAPAR products derived from VGT observations. Data sets operationally used by the JRC-MARS Unit have been compared with the Geoland2 BioPar (GEOV1) data set. The analysis is performed on the two first full years of GEOV1 demo data encompassing 72 (2 × 36) time steps (February 2003 to February 2005). Depending on the region investigated, JRC-MARS currently uses two different archives: the AGRI4CAST archive covering an extended European window (hereafter referred to as A4C) and the FOODSEC global archive (FS) which is used outside this window. FS and GEOV1 archives share the same geographic reference system (Geographical Lat/Lon, WGS84, pixel size = 1/112°), while A4C imagery is stored in the INSPIRE-LAE system at 1 km resolution. Since the latter data set required a re-projection (nearest neighbor resampling) to Geographical Lon/Lat reference system, the comparison GEOV1 versus A4C was conducted using an average over a 3 × 3 pixels support area. In this way, we aim to limit possible co-registration errors.

Table II summarizes the main characteristics of the investigated fAPAR products. It is important to observe that, according to product definitions, FS and GEOV1 algorithms have a common trait because they both use CYCLOPES fAPAR as one of the elements used to train the neural networks (details in Sections III-A and C).

With the exception of the empirical cumulative distribution functions (ECDFs) computation which is carried out separately on the two data sets, the analysis was performed using only cloud-free pixels (by masking out the union of the two cloud masks). This ensures that the comparison is unbiased and based on identical sets of dekadal observations.

### III. DATA

**A. FS-fAPAR**

In this data set fAPAR is derived by the CYCLight algorithm [32]. The algorithm is based on a neural network trained with dekadal top of canopy (TOC) reflectances (so-called VGT-S10) in the red and near infrared (NIR) bands, view and solar angles at observation time as input and the CYCLOPES V3 [11] fAPAR derived also from the VGT sensor as output. CYCLOPES products thus correspond to “black sky” (under direct illumination) fAPAR at 10:00 A.M., solar local time. The VGT-S10 inputs are 10-day images composited using the NDVI maximum value composite (MVC) criterion on calibrated, cloud-screened, and atmospherically corrected (SMAC algorithm, [33]) daily top of the atmosphere imagery (so-called VGT-P) as processed by the VGT image processing center (CTIV-VITO).

**B. A4C-fAPAR**

The A4C-fAPAR is produced for Europe only, and it is derived by the VITO implementation of the JRC-IES algorithm [34], [35] for VGT-P data. The algorithm is based on the use of a spectral index optimized to provide an estimate of fAPAR and the so-called rectified reflectance values in the red and NIR spectral bands. These are virtual reflectances largely uncontaminated by atmospheric and angular effects. The relation between the index and fAPAR is specified using the physically based semidiscrete model of Gobron et al. [36] and the 6S atmosphere radiative transfer scheme [37]. The output of this process is the “black sky” and “green leaf” fAPAR at the time of satellite overpass (about 10:30 A.M. local solar time). Dekadal syntheses are generated using the fAPAR MVC rule.

**C. GEOV1-fAPAR**

The pre-operational processing chain of the Geoland2 Core Mapping Service BioPar provided the GEOV1 fAPAR product (version 1) [38], [39]. GEOV1 fAPAR estimation capitalizes on existing fAPAR products, namely MODIS Collection 5 [9] and CYCLOPES V3 [11], with the aim of taking advantage of their specific performances while limiting the cases in which they are deficient. These products are combined and linearly scaled to generate a “fused” fAPAR product. Model calibration is done over the BELMANIP2 sites [40] that are assumed to represent all continental biomes.

Two years of fAPAR are generated for these sites and then used to calibrate a neural network able to relate the fused product to pre-processed VGT data, namely: sun zenith angle at observation time, red, NIR, and short-wavelength infrared TOC bidirectional reflectance factor (BRF)-normalized reflectances. The preprocessing includes cloud screening, atmospheric correction and BRF normalization combined with temporal compositing over a window of 30 days. This last step is an iterative process which also allows for the removal of any cloud-contaminated pixels which were not detected during cloud screening. The algorithm is applied at 10-day frequency using a time window of 30 days ranging from 16 days before the day of dekadal update to 13 days after. A Gaussian weighting function is applied to the synthesis period with the maximum weight on the last observation of the period. The outputs of this process are dekadal products of the “black sky” fAPAR at
D. Study Regions

Three study regions have been selected among those currently being monitored by JRC-MARS and corresponding to different eco-climatic regions: France (Europe, temperate to Mediterranean climate, Cw to Cs according to Köppen climate classification scheme [41]), Mato Grosso—Brazil (South America, tropical climate, Am to Aw), and Niger (Africa, steppe to desert, BS to BW) (Fig. 2). The analysis was based on the $10^\circ \times 10^\circ$ tiles ($1120 \times 1120$ VGT pixels) adopted for GEOV1 data distribution. Results of the comparison between A4C and GEOV1 are presented for the France tile whereas those between FS and GEOV1 are presented for Brazil and Niger.

To allow land cover specific analysis, the Global Land Cover 2000 land cover map [42] was used for Niger and Brazil, whereas the Corine Land Cover 2000 [43] was used for the France tile. Original land cover classes were aggregated and re-coded into three broad vegetation classes plus a residual one (“others”) comprising the remaining classes. This recoding allowed us to focus on the main vegetated land cover types. For the Brazil and France tiles, the three classes are represented by forests, cropland, and shrub- and grassland. The remaining pixels were mainly covered by water and other land uses. For the Niger tile, the three main classes are composed of wood- and shrubland, cropland, and grassland. In this last region, the residual class is mainly composed of bare soil and sparse vegetation.

IV. RESULTS AND DISCUSSION

A. Cloud Masks and Valid Observations

For the three study regions, the fraction of valid observations is greater than 80% (Fig. 3). A4C and more noticeably FS have a greater fraction of valid observations than GEOV1 regardless of the land cover type. The Brazil tile, severely affected by cloud coverage, shows the largest difference in valid observation between the data sets, with GEOV1 having 7.5% less valid observations compared to FS. Differences in the number of valid observations vary with time/season and geographic locations in the images. The largest differences were observed between FS and GEOV1 during the wet season in specific geographic areas of Brazil (Fig. 4). This wet season...
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 4. Brazil tile, percentage of valid observations maps for FS (a) and GEOV1 (b) time series, respectively. White areas correspond to water bodies. The temporal profile of the mean fraction of valid observations for the forest land cover is presented in panel (c) FS (black curve) and GEOV1 (red curve). Time is expressed in months since February 2003.

In conclusion, the difference in the fraction of valid observations is due to the different cloud detection performances, the one adopted by GEOV1 being more conservative and very likely, more realistic. This finding was as also confirmed by visual inspection of the imageries.

B. Visual Inspection of Temporal Profiles

Temporal profiles of individual pixels are shown in Fig. 5 to provide a visual and qualitative example of the products. Fig. 5 shows the profiles of three randomly extracted pixels (one for each land cover class considered) for each of the study regions.

Without any generalization intent, we observe that these individual JRC-MARS profiles show lower values and more erratic trends compared to GEOV1. Although the bias between GEOV1 and FS seems about stable with time, it increases in winter over France as compared to A4C. Such qualitative observation is further developed in the following statistical analysis.

C. Overall Agreement

Data distribution is examined comparing the ECDFs over the three study regions (Fig. 6). GEOV1 values are generally higher than the corresponding FS and A4C. Differences in ECDF shapes are noticeable for France (GEOV1 versus A4C) [Fig. 6(c)]. For example, the fraction of observations with fAPAR lower than 0.5 units is 29% and 73% for GEOV1 and A4C data sets, respectively. A better agreement is found for Niger [Fig. 6(a)]. Differences of up to 0.1 fAPAR units regarding the fAPAR values at which ECDF reaches 100% (i.e., saturation value) are present in all study regions.

Quantitatively, differences in ECDFs are described by the Kolmogorov–Smirnov statistic ($D_{KS}$) reported in Fig. 6, showing the largest value for France and the lowest for Niger. The two-sample Kolmogorov–Smirnov test at the 5% significance level rejected the null hypothesis that ECDFs are from the same continuous distribution in all regions and for all classes analyzed.

The presence of horizontally flat sections of the ECDF curves (missing classes of values) is more pronounced for GEOV1 because of its restricted dynamic range used to digitally store the data (95 levels in the range 0–0.94) with respect to that of FS and A4C (201 levels in the range 0–1). Such an effect is not present in France because original data were averaged over a $3 \times 3$ pixels spatial support. The restriction of the dynamic range of GEOV1 eventually resulted in an unnecessary compression of information.
Fig. 6. Empirical cumulative distribution functions (ECDFs) across all land cover classes. All available dekads are pooled together. GEOV1 and FS ECDFs are depicted in (a) and (b) for Niger and Brazil, respectively. GEOV1 and A4C ECDFs are depicted in (c) for France.

Fig. 7. Density scatterplots GEOV1 versus FS [(a) and (b), Niger and Brazil], A4C [(c), France], all available dekads are pooled together. Darker areas in the plots indicate higher density of sample points. The linear OLS regression line is shown in red, the 1:1 line in black.

Fig. 7 refers to the overall correlation analysis and shows the density scatterplots of GEOV1 versus FS and A4C fAPAR for the three regions of interest together with the linear ordinary least square (OLS) regression line. High level of correlation is found between GEOV1 and FS data sets ($R^2$ equal to 0.94 and 0.86 in Niger and Brazil, respectively). However, the relationship between the two data sets is different in the two regions with Niger being characterized by a small positive offset and a slope $> 1$ ($b_0 = 0.015, b_1 = 1.109, p < 0.001$) and Brazil showing a consistent positive offset with a slope close to one ($b_0 = 0.073, b_1 = 1.032, p < 0.001$). For the France tile, the scatterplot of GEOV1 versus A4C shows a reduced $R^2$ (0.60) with the OLS regression line being far away from the 1:1 line ($b_0 = 0.310, b_1 = 0.740, p < 0.001$). This indicates low correspondence between data sets with highest differences for low to intermediate fAPAR values. Results over France are comparable with those presented by Camacho and Cernicharo [19] where the JRC-FAPAR algorithm was applied to the SeaWiFS sensor data instead of VGT. Even if the input data are different, both analyses show that GEOV1 products present larger values than those achieved using the JRC-FAPAR approach. Similarly, in a comparison study by McCallum et al. [22] in Northern Eurasia, the SeaWiFS JRC-fAPAR estimates (equivalent to A4C) were also found lower than those of MODIS and CYCLOPES, the two algorithms used to produce GEOV1.

A number of differences in definition and computation of fAPAR contribute to explain the observed bias between GEOV1 and A4C products. First, A4C algorithm is based on a “green” leaf scattering assumption and provides lower fAPAR values than using less constraint on the leaf single scattering albedo, as with GEOV1. These different assumptions on leaf optical properties may have a significant impact on fAPAR and explain differences in the order of 0.05–0.20 fAPAR units depending on vegetation density [23]. Second, GEOV1 algorithm explicitly models the length of light path in the vegetation as a function of the sun position, resulting in higher fAPAR values during wintertime when the sun zenith angles are larger and the light path in the vegetation increases. Third, GEOV1 algorithm uses a linear scaling factor of 1.05 applied to fAPAR estimation to match the expected maximum value of 0.94.

It is also noticed that the sample points almost fill the entire sector above the 1:1 line in the three scatterplots. This behavior can be explained by the combined effect of two GEOV1 features: a larger compositing temporal window and greater cloud screening effectiveness of the processing chain. In fact, for any GEOV1 fAPAR level ($y$-axis), we can observe similar FS and A4C levels (i.e., data points lying close to the regression line) and a set of FS and A4C observations characterized by lower levels, eventually as low as almost zero. Such negatively biased observations found in the FS and A4C data sets can be attributed to the presence of undetected clouds resulting in erroneous (underestimated) fAPAR values. This effect is clearly visible in the two regions characterized by higher levels of cloudiness (i.e., Brazil and France) and less evident for the semi-arid region of Niger. Only in a few cases do the FS and A4C data sets show higher fAPAR values compared to GEOV1. Further investigation of temporal profiles extracted retaining all observations for each data set confirmed that when FS or A4C fAPAR show unrealistic and negatively biased spikes, GEOV1 estimation is not/less affected or missing [see for example of Fig. 5(b) in February]. This potentially problematic issue with current FS and A4C products can be minimized by smoothing the time series using an algorithm that makes the data approach their upper envelop (e.g., [2], [44], and [45]).
D. Spatial Agreement

The analysis of the temporal evolution of selected spatial metrics aims to highlight specific periods of spatial disagreement. As an example, Fig. 8 shows the mean fAPAR profiles and the root mean square difference (RMSD) between paired data sets [Fig. 8(a)–(c)] for the forest land cover. In addition, using the geometric mean functional relationship regression (GMFR), the RMSD is partitioned into its systematic and unsystematic components [Fig. 8(d)–(f)]. Finally, [Fig. 8(g)–(i)] reports the temporal profile of the agreement coefficient (AC).

The AC and the RMSD show a pronounced seasonality in all regions. Periods of minimum agreement between GEOV1 and FS are found during seasons characterized by highest precipitations in the southern part of Niger (June to September) and Brazil (October to March). The average agreement between GEOV1 and A4C (France tile) is reduced and shows minima during wintertime [Fig. 8(i)].

The systematic component of the RMSD is greater than the unsystematic one for the three study regions [Fig. 8(d)–(f)] with some exceptions in Niger during the start of the growing season. Periods of lowest agreement yield an increase of unsystematic differences in Niger and Brazil [Fig. 8(d) and (e)] and of systematic differences in France [Fig. 8(f)]. This behavior suggests that temporally limited events of spatial disagreement are due to random “noise” in Niger and Brazil, while a systematic difference in fAPAR retrieval exists in France.

A closer investigation prompted by this observation revealed that for Niger and Brazil (GEOV1 versus FS), events of spatial disagreements are due to two causes. The first one is related to the different compositing window length and cloud screen-

E. Temporal Agreement

The maps of the temporal AC (Fig. 9) and RMSD (Fig. 10) show that the strength of the temporal agreement between data sets varies spatially.

The Niger tile shows very low temporal correlation between the two data sets in desert zones (i.e., upper part of the image) where the fAPAR values are consistently low throughout the year because vegetation is very sparse or absent. These low AC values are expected in such situations due to the lack of temporal variability. A high agreement is instead found in
the southern part of the tile, where vegetation with a distinct seasonality is present. In this tile, the spatial heterogeneity of agreement is mainly driven by the existence (or not) of temporal fAPAR variability and should not be interpreted as a result of different fAPAR algorithms.

For Brazil, we observe that AC minima are located in the forested areas in the northern part of the tile as well as in the Pantanal wetlands (southwest part). The low AC is again due to the fact that forests and flooded wetlands exhibit reduced fAPAR dynamics during the year. In addition, as mentioned in Sections IV-A and B, the frequent presence of clouds in the area (as described by the maps of fraction of valid observations; Fig. 4) contributes to reduce the agreement between the two data sets because of the weaknesses of the FS processing chain on cloud detection and decontamination. For the remaining areas, high AC values are observed.

Even if the AC is spatially heterogeneous, Fig. 10 shows a small RMSD for Niger and Brazil (lower than 0.1 and 0.2 fAPAR units, respectively). Neglecting non-vegetated areas, the highest proportion of systematic difference is found for the classes exhibiting the lowest fAPAR seasonality, i.e., grassland and forest classes in Niger and Brazil, respectively.

In the France tile, most areas show a reduced agreement between the two data sets (Fig. 9) and a large RMSD, often exceeding 0.3 fAPAR unit (Fig. 10). Areas characterized by such high RMSD show the largest differences during late autumn and winter, when GEOV1 values exceed 0.5 fAPAR units over large areas. The systematic component dominates the RMSD with some exceptions for agricultural areas. As already observed in Section IV-D, the existence of such large discrepancies between the two products during wintertime deserves further investigation.

F. Temporal Smoothness

Starting from the assumption that fAPAR varies smoothly over vegetated land when mapped at 1 km² resolution (with possible exceptions for natural disturbances such as fires or strong human interventions such as grassland cutting or crop harvesting), we investigated the temporal smoothness of the two data sets. This feature can be represented by the mean absolute value of the first derivative of fAPAR over time as shown in Fig. 11(a) and (b) for the Brazil tile as an example.

The FS data set [Fig. 11(a)] shows generally much larger short-term variability compared to GEOV1 [Fig. 11(b)]. For forest and cropland land cover classes, this is also demonstrated by the first derivative ECDFs [Fig. 11(c) and (d)]. For FS, about 40% of all absolute dekadal variation shows values greater than 0.05 fAPAR units. Such high occurrence of large high frequency fluctuations is indeed implausible in the given geographical setting. On the contrary, the smoothness of GEOV1 data set looks more realistic and also reveals the expected influence of land cover [Fig. 11(b)]. In fact, one expects (on average) a much stronger dynamic in agricultural land as compared to permanent forests. The main agricultural production areas of Mato Grosso clearly appear (in greenish colors) in the GEOV1 product [Fig. 11(b)].

V. Conclusion

The study compared different fAPAR time series derived from SPOT-VGT: the new GEOV1 product with two products currently used for operational crop monitoring and yield forecasting by the JRC-MARS Unit. To describe the agreement (or disagreement) between paired data sets, we employed a selection of statistical metrics and a procedure that enabled us to compare their temporal and spatial variability. This provided additional information to better evaluate the agreement between two RS image time series. Ultimately, this should lead to improved processing algorithms. In addition, the methodology presented may serve as guideline in future studies.

The overall comparison between data sets, i.e., space and time considered together, showed only low to moderate agreement, thus questioning the reliability of the different processing chains. GEOV1 products were systematically higher than
those of JRC-MARS and showed greater agreement with FS rather than A4C data. The higher agreement found between the GEOV1 and FS time series may reflect the fact that both algorithms have been calibrated using a common input (CYCLOPES derived fAPAR product).

The analysis of the temporal evolution of the spatial agreement showed specific seasonality in the three study regions. When comparing GEOV1 with FS (Niger and Brazil tiles), peaks in spatial disagreement were found to be related to the greater effectiveness of the GEOV1 cloud screening method and overall greater stability of GEOV1 dekadal syntheses during periods presenting adverse atmospheric conditions. The lack of BRF normalization in the FS product was identified as an additional cause of spatial disagreement. Both GEOV1 features, cloud screening and BRF normalization, represent in our opinion a significant improvement with respect to FS archive.

The largest differences between GEOV1 and A4C are observed in wintertime when GEOV1 presents higher fAPAR values as compared to A4C data. The higher sun zenith angles experienced for the higher latitudes that increase the optical path in the remaining green vegetation have been discussed as a possible cause contributing to this non-negligible wintertime differences. From a user point of view, it is recommended that further research is undertaken to verify the reliability of the products, including the evaluation of wintertime GEOV1 and A4C fAPAR evolutions against in situ measurements.

FS and A4C time series presents sudden falls due to data points affected by cloud contamination. This is also observed from the analysis of the first derivative of fAPAR over time, showing unrealistic short-term fluctuations for the JRC-MARS data with respect to more plausible and smooth temporal profiles of GEOV1. As a consequence, it is recommended that FS and A4C data are treated with an algorithm that smooths out the negatively biased atmospheric noise and makes the data approach their upper envelop. It must be noted that the observed smoothness of GEOV1 is also due to a longer synthesis period (30 days) with respect to JRC-MARS (FS and A4C) data sets (10 days). Even if the most recent observations in the GEOV1 have higher weights in the computation of the fAPAR value, the longer synthesis period may represent a limit for near real-time applications in which the interest is focused on the most recent information only. In this case, the use of JRC-MARS data would be preferable, provided that some quality checks are performed to exclude possible cloud contamination.

The temporal agreement between data sets showed spatial (and land cover related) variability in the study areas spanning from very low to almost perfect agreement. In spite of this variability, Niger and Brazil (GEOV1 versus FS) showed homogeneous and quite low differences in terms of RMSD. On the contrary, in France (GEOV1 versus A4C), spatial variability of the AC was also associated with large fAPAR differences. The difference between data sets was mostly dominated by its systematic component in both the spatial and temporal domains. The elimination of the systematic component through the application of a regression function is in principle feasible but problematic because the relationship between data sets has been found to be variable a cross space and time.

The present study was based on the analysis of the first two years of GEOV1 demonstration products. Further analysis of the full SPOT-VGT archive (1999 to date) is planned to complement the present comparison with an assessment of fAPAR products in operational JRC-MARS applications, that is, testing their relative performances as indicators for the near real-time detection of vegetation anomalies or as statistical predictors of crop yield.

REFERENCES


Christelle Vancutsem received the M.Sc. degree in agronomy and the Ph.D. degree in agronomic science and biological engineering from the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 2000 and 2008, respectively.

She has 12 years of research experience in remote sensing and geographic information systems, in three different institutions (University of Louvain-la-Neuve, Columbia University, Joint Research Centre). She has contributed to various land cover mapping projects. She was responsible for the development of an automatic classification system combining multispectral and temporal information for the production of a global land cover map at 300 m (Globcover, ESA). She has a relevant experience on the spatio-temporal analysis of remotely sensed and climatic time series for health and agriculture applications: evaluation of remotely sensed products to monitor air temperature in epidemiological studies, validation of a vectorial capacity model for malaria transmission, dynamic mapping of cropland areas using MODIS time series, production of bulletins on the assessment of the crop season, and production of a global crop mask combining existing land cover/use data sets.

Nadine Gobron (M’01) received the Maîtrise de Physique degree, the DEA degree, and the doctorat d’Université en Physique de l’Atmosphère from the Université Blaise Pascal, Clermont-Ferrand, France, in 1993, 1994, and 1997, respectively.

She was a European Space Agency (ESA) Research Fellow in 1997–1998 and a Grant-Holder at the Space Applications Institute of the EC JRC in 1998–2000. Currently, she is with the Institute for Environment and Sustainability, Ispra, Italy. She is the author or coauthor of over 60 refereed publications. She was awarded the “Group Achievement Award” by NASA in 2001 and the best young JRC Scientific in 2004. Her main interests currently include research on the theory of radiation transfer in plant canopies and, more generally, the development of tools to quantitatively interpret satellite remote sensing data in the optical spectral domain via inverse methods. She is a contributor to the MERIS Scientific Advisory Group of ESA and an associated member of the MISR Science Team of NASA/JPL. She is the Chair of Committee on Space Research (COSPAR) Commission A on Space Studies of the Earth’s Surface, Meteorology, and Climate of COSPAR.

Frédéric Baret received the Ph.D. degree in the use of remote sensing for crop monitoring from the University of Paris-Sud (Orsay), in 1986. Currently, he is a Research Director at INRA, leading a remote sensing team. He has coordinated several national and European projects. He is involved in the development of radiative transfer models at several scales (soil, leaf, canopy) and their use for the retrieval of biophysical variables. He developed retrieval algorithms (CYCLOPES, GEOV1) both from satellite and airborne sensors as well as close range remote sensing. He is deeply involved in the validation of remote sensing products. He authored more than 150 research papers.

Roselyne Lacaze received the Ph.D. degree in remote sensing applied to biosphere processes and Meteorology from the University of Toulouse, France, in 1999.

She specializes in the retrieval of vegetation properties from optical and multi-angular EO data. She participated in validation of vegetation parameters for many projects. She coordinates the activities dedicated to the definition, production, and validation of the global biophysical variables in the framework of the EC-funded FP7 geoland2 project. She is involved in the continuity of the GMES land service, namely being the Coordinator of an EC-funded FP7 ImagineS project.

Herman Eerens received the M.Sc. degree in agricultural and applied biological sciences from the University of Leuven, Leuven, Belgium, in 1980.

He worked for about 15 years at the same university, first in the forestry department, from 1985 onward in the field of remote sensing. His main activities included radiometry of crops and soils, modeling of crop reflectance, atmospheric correction, and the classification of high-resolution images. In 2007, he joined the Flemish Institute for Technological Research, in Mol, Belgium. Since then, he has been mostly focusing on the processing of low-resolution data of SPOT-VEGETATION, NOAA/METOP-AVHRR, MODIS, and MSG and the extraction of information for agro-statistical purposes.

Olivier Leo received the M.Sc. degree in agricultural engineering from the Institut National Agronomique, Paris, in 1974.

He is an Agronomist Engineer, with a specialization in soil sciences and natural resources. He has participated in numerous studies for rural development, agro-industrial projects, natural resources inventories, or agricultural statistics, in developing countries and Europe. Since 1994, he has been with the European Commission Joint Research Centre, Ispra, Italy, where he has been in charge of the MARS activities for the design of the Common Agriculture Policy and its implementation by new member-states in the frame of EU enlargement. He leads the FOODSEC action ensuring crop monitoring outside Europe and addressing food security issues (food availability, access, and vulnerability) with a focus on Sub-Saharan Africa and in support to EC policies in this field.