Towards an Operational Delivery of a Time-Series of Essential Climate Variables for Global Land Monitoring

Bruno Smets[1], Roselyne Lacaze[2], Philippe Pacholczyk[3], Hassan Makmahra[3], Frédéric Baret[4], Fernando Camacho[5], Patrick Claes[1]
VITO, Mol, Belgium; {bruno.smets, patrick.claes}@vito.be
HYGEOS, Toulouse, France; rl@hygeos.com
CNES, Toulouse, France; {philipipe.pacholczyk, hassan.makhmara}@cnes.fr
INRA, Avignon, France; baret@avignon.inra.fr
EOLAB, Valencia, Spain; camacho@eolab.es

Abstract. Geoland2 project is part of the Global Monitoring for Environment and Security (GMES) European-lead initiative intends to prepare, validate and demonstrate pre-operational service chains and products of the Land Monitoring Core Service (LMCS). One of 3 Core Mapping Services (CMS), BioPar is providing in near real-time and off-line biophysical variables describing the continental vegetation state, the radiation budget at the surface, and the water cycle. A close cooperation between several teams has resulted in the pre-operation of a CEOS validated global LAI and fAPAR dataset with 1km spatial resolution, 10 days sampling frequency, capable to be applied consistently to several sensors and with known accuracy and consistency. These products can be envisioned as an important step towards the Essential Climate Variables (ECV) as defined by GCOS and are freely disseminated by a Spatial Data Infrastructure (SDI) in a format in accordance to international standards (ISO and INSPIRE directive). This paper introduces the ECV products and emphasizes the development and integration platform to make optimal use of the available computer resources and on the content of the free operational products.

Keywords. Environment, Global Land Monitoring, Operations, Essential Climate Variables.

1. Introduction

The European Global Monitoring for Environment and Security (GMES) initiative provides a political framework for future implementations of Services Centres related to environmental applications. Building upon the results of preceding projects funded by the European Commission, and the European Space Agency, geoland2 is the last brick towards the implementation of fully mature GMES Land Services. Its goal is to build, validate and demonstrate operational processing lines and land products on a user-driven basis.

1.1. The geoland2 project

Geoland2 is made of 3 Core Mapping Services (CMS) providing "basic" geo-information on the land state covering a wide variety of thematic content, spatial and temporal scales; and 7 Core Information Services (CIS) which are thematic elements that use the CMS products to generate more "elaborated" information addressing specific European policies [1]. The 3 CMS deliver global, regional and local maps in the following areas of interest : (i) Euroland provides periodic land cover change and urban atlas, (ii) BioPar provides near real-time and off-line biophysical parameters and (iii) SATChMo provides area frame sampling for seasonal monitoring and land cover change. The 7 CIS thematic services provide products for (i) terrestrial land carbon monitoring, (ii) global crop monitoring, (iii) natural resource monitoring in Africa, (iv) impact of agricultural practices on envi-

ronment, (v) forest detection and monitoring, (vi) water resources monitoring and quality, and (vii) spatial planning to detect and forecast urban land use changes.

All Products and services provided by the geoland2 project are available progressively to users on the Spatial Data Infrastructure (SDI) portal, located at http://www.geoland2.eu.

1.2. Essential climate variables in BioPar Core Mapping Service

The, Bio-geophysical Parameter (BioPar) CMS aims at setting-up operational infrastructures to provide regional, continental, and global variables describing the vegetation state, the radiation budget at the surface, and the water cycle, both in near real time and off-line mode (Figure 1).

Product	NRT / Off-line	Spatial Resolution	Spatial coverage	Temporal Resolution	Sensor
Continental Vegetation					
LAI, FCover, FAPAR, DMP, NDVI	NRT	1 km	Global	10-days	VGT
Times series of vegetation products	Off-line	4 km	Global	10-days	AVHRR + VGT
Burnt areas + seasonality	NRT	1 km	global	Daily	VGT
MERIS FR biophysical variables	NRT	300 m	Europe	10-days	MERIS
HR biophysical products	Off-line	10 m	Pilot Areas	4 times/year	SPOT
Energy Budget					
Downwelling Shortwave Surface Flux Downwelling Longwave Surface Flux	NRT	~5 km	Global	hourly	ΣGEO + AVHRR
Land Surface Temperature	NRT	~5 km	Global	hourly, daily, 10-days	ΣGEO + AVHRR
Surface Albedo	NRT	1 km	Global	10-days	VGT
Surface Albedo	NRT	~5 km	Global	10-days	ΣGEO + AVHRR
Water cycle					
Water bodies + seasonality	NRT	1km, 250m	Africa	10-days	VGT, MODIS
Soil Moisture + Freeze/Thaw	NRT	0.1°	Global	Daily	ASCAT
Time series of soil moisture products	Off-line	25 km	Global	Daily	ERS-1&2 Scatt

Figure 1: BioPar CMS Portfolio

The importance to continuously monitor the Earth's surface was recently recognized by the Global Climate Observing System (GCOS)[2]: a set of Essential Climate Variables (ECV) was identified as being both accessible from remote sensing observations and intervening with key processes. The Global Terrestrial Observing System (GTOS) is the programme for observations, modelling and analysis of terrestrial ecosystems to support sustainable development and detect and manage global and regional environmental changes. Among the ECVs related to land surfaces, LAI (Leaf Area Index)[3] and fAPAR (Fraction of Absorbed Photosynthetic Active Radiation)[4] may be derived from observations in the reflective solar domain and play a key role in processes, such as photosynthesis, respiration and transpiration.

To provide LAI and fAPAR products as terrestrial ECV, three important objectives are to be met: (i) Consistent (long) time-series: the algorithm is capable to be applied consistently to several sensors, and the operation provides a sustainable service in Near-Real Time (NRT) and fast off-line (re-)processing taking advantage of existing EO data archives; (ii) Accurate and with known accuracy: propagation of quality information throughout the operational chain and validation results according to CEOS guidelines; (iii) Easy accessible products through a free data policy mode.

The close cooperation between the research team (INRA), the development team (CNES), the operations team (VITO) and the validation team (EOLAB) under supervision of the task manager (HYGEOS) realizes a major step towards the establishment of T10 and T11 ECV datasets by providing the GEOV1 LAI and fAPAR datasets.

2. Algorithm

LAI is defined as half the total developed area of leaf elements per unit horizontal ground area [5]. fAPAR is defined as the fraction of radiation absorbed by the canopy in the 400-700 nm spectral

domain under specified illumination conditions [6]. Both products, generated by the biophysical algorithm GEOV1, are based on already existing products to capitalize on the efforts accomplished and obtain a larger consensus from the user community. Following the published literature on products validation, the best performing products, the CYCLOPES and MODIS products, were selected and fused to take advantage of their specific performances while limiting the situations where these products show deficiencies [7]. Neural networks were calibrated and run in parallel to provide consistent estimates of the biophysical variables for each sensor, along with the quality flags and quantitative uncertainties for both biophysical variables LAI and fAPAR. The top of canopy reflectance acquired by the VEGETATION sensor in the red (B2), near infrared (B3) and short wave infrared (SWIR) are used as inputs after being preprocessed(cloud screening, atmospheric correction) and normalized from bidirectional effects [8]. In addition, the cosine of the sun zenith angle ($\cos(\theta_s)$) at the observation time is also used as input. To benefit from the better performances observed for CYCLOPES $fAPAR$ and LAI products for the low values, and for MODIS products for the large values, the MODIS and CYCLOPES products are fused using a weight factor [7]. The similar processing applied to both LAI and $fAPAR$ is expected to keep a good consistency between the both products, and on top provides a cover fraction (FCover) product defined as the fraction of background covered by green vegetation as seen from nadir that can be used when separating the contribution of the soil from that of the canopy.

Quality assessment indicators are provided along with the products: (i) Input out of range: this represents the consistency of the measured VEGETATION input reflectances with those used in the training data base. A flag is raised when observations are outside the training definition domain. (ii) Output out of range: this represents the consistency of the actual network outputs (the biophysical variables) with those used in the training data base. (iii) Estimated uncertainties: this represents the expected error expressed in Root Mean Square Error (RMSE) between the estimated and the actual biophysical values as derived from the theoretical performances of the algorithm evaluated over an independent data set. The reflectance uncertainties are used to define a confidence interval. The LAI and $fAPAR$ with corresponding reflectance inside the confidence interval are then used to compute the RMSE. A specific network is finally trained to relate the estimated uncertainties to the input reflectance and observation geometry values [7]. (iv) Quality indicators: they are a replication of the quality indicators computed in the preprocessing algorithmic steps, including those related to the atmospheric correction and cloud filtering.

3. Validation

As for the retrieval methodology, the GEOV1 processing line is a heritage of the CYCLOPES V3 processing line which has been consolidated, and adapted to fits the BioPar requirements. Before its integration in the operation center, the processing line has been run off-line to generate 2 years (2003-2004) of GEOV1 demonstration products (Figure 2) used to perform the validation exercise. These products have also been supplied to geoland2 users who checked they matched their requirements before moving into operation.

The validation exercise was defined to be consistent with the best practices proposed by the CEOS WGCV LPV subgroup [9]. First an inter-comparison with the existing global products (MODIS C5, CYCLOPES V3.1, GLOBCARBON V2) was performed to examine the spatial and temporal consistency of GEOV1. Second, a direct validation approach used ground reference maps to quantify the overall performance of the products. The products were inter-compared over the BELMANIP-2 global network of homogeneous sites that was designed to represent the variability of continental landscapes [10]. The products were analyzed for 7 generic classes, namely: Ever-

green Broadleaf Forest, Deciduous Broadleaf Forest, Needle-leaf Forest, Shrublands, Herbaceous, Cultivated, Sparse and bare areas. Several criteria of performance were evaluated, including spatio-temporal continuity, spatial distribution of retrievals, dynamic range of retrievals, statistical consistency per biome type, temporal variations and smoothness of temporal profiles, precision and accuracy. The products were compared over a similar spatial support area and temporal support period. They were thus re-sampled over Plate Carrée projection at 1/112° (about 1 km at equator) spatial resolution. The comparison was deducted using an average value over 3x3 pixels to reduce co-registration errors between products and differences in their sensor Point Spread Function (PSF) which determines the actual footprint of the data. The accuracy of the products was compared with a number of high resolution maps set-up from ground measurement collected and processed according to CEOS/WGCV LPV guidelines.

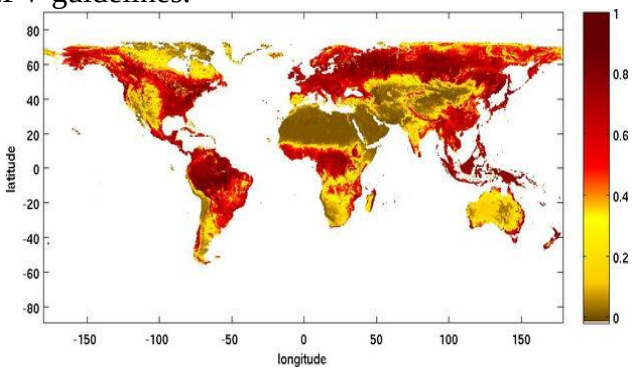

Figure 2: Example of global FAPAR. July 2003.

The results show that GEOV1 products display reliable spatial distribution, smooth temporal profiles, stable from year to year, good dynamic range with realistic magnitude from bare areas to densest forests, and optimal performances with ground-based maps. They clearly outperform the quality of reference global products [11]. Their main limitation is the lack of spatio-temporal continuity, mainly at high latitudes and equatorial regions, as in the other global products. A limitation that could be partly solved with a gap filled version of the products, which is currently under development.

4. Operations

4.1. Principles

The entire GEOV1 processing chain is automated using different independent subsystems to retrieve all required input data, process this data according the algorithm principles, archive the computed data into the Long Term Data storage Facility (LTDF) and package this data into LAI and fAPAR ECV products before making them freely accessible.

The core computing uses a Job Management controller cluster, while performing permanent quality control by a Centric Quality Control Manager (Figure 3). The actual computing is done on processing (CPU) nodes by making use of either Local Storage or one or more Network Storages. The GEOV1 processing line, a heritage of the CYCLOPES-V3 processing line, handles both LAI and fAPAR computing and is executed by geographical division of the working area. The processing is data-triggered and starts execution as soon as all required input data is available and qualified. The processing line is designed according the Master Worker design pattern [12] to make optimal use of computing nodes, data storage, etc. in a data-centric architecture system.

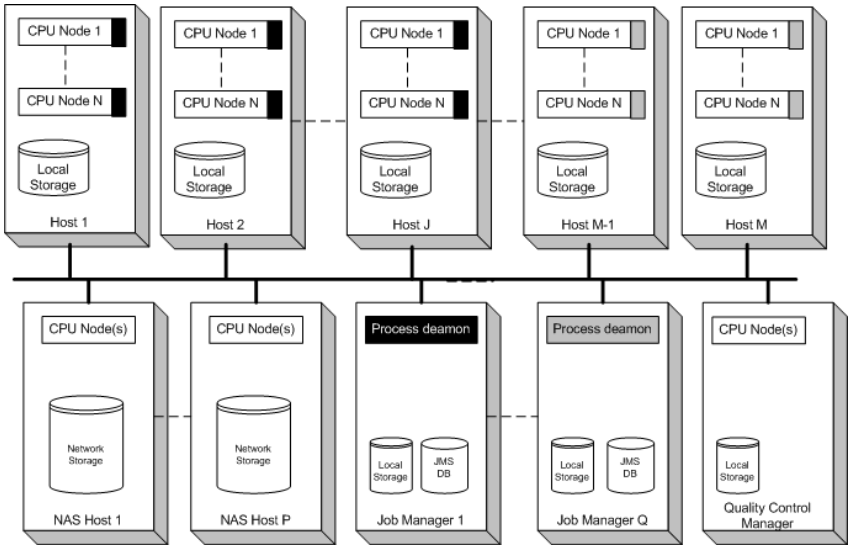

Figure 3: Deployment pattern for standalone distributed nodes.

The trade-off between throughput time and cost is configurable by selecting how many threads are executed in parallel for every processing step and which data storage segment to use. The Job Manager handles this trade-off while the production Quality Control Manager (QCM) performs three major tasks: (i) image control by performing automatic error detection and visual inspection of images during several processing steps, (ii) process control to monitor the executing of jobs and to create reports to gather statistics; and (iii) infrastructure control to monitor the systems to guarantee a maximum uptime.

The Quality Control System (QCS) keeps track of all products and their relationships (parent/child) and all processes with their hierarchy (ParentChain/Chain/Step) using a data model (Figure 4). The data model enables the QCM to verify the execution logic, to keep track of the status of all products and their history, and to analyze the system performances. Following each processing step, an optional quicklook image can be established and registered for a manual visual check to identify and reject bad products. Without this measure, bad input segments could spoil the quality of the final products.

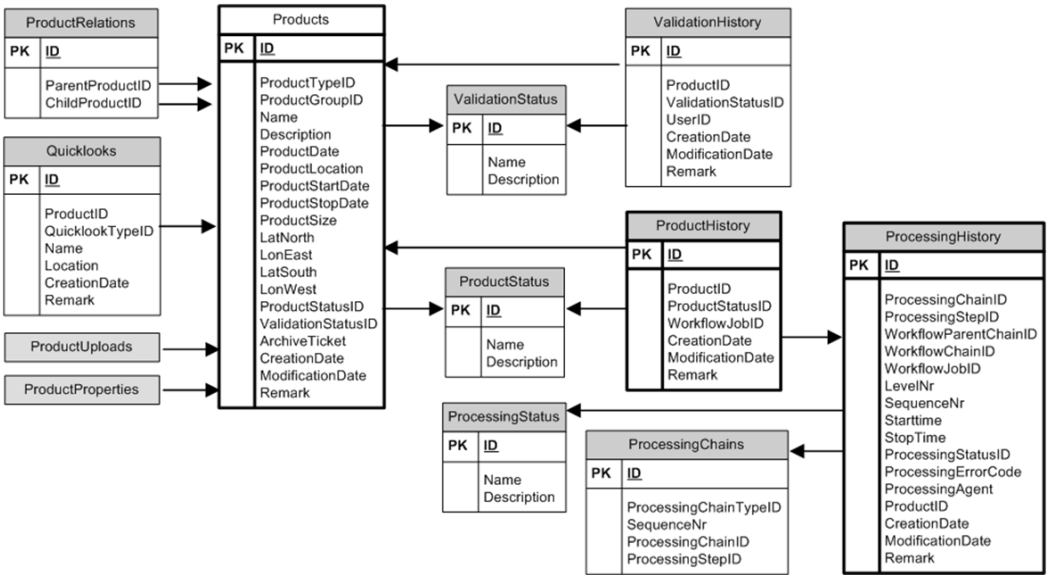

Figure 4: Quality Control System data model

Since both input data and disseminated data (ProductUploads) is registered into the QCS, the QCM is able to provide an overview of the all processing steps from start till end which enables a strong controlling and reporting mechanism.

4.2. Processing

The data processing entails 4 steps: (1) pre-processing, (2) normalization, (3) bio and (4) data packaging, with an additional optional step (5) to generate maps for quality control. The pre-processing step performs a radiometric calibration on SPOT-VEGETATION segments to obtain consistent values across both instrument versions. After this calibration step, cloud screening and atmospheric correction is performed based on daily water vapor, ozone information from NASA SeaDAS ancillary dataset and a monthly climatology of aerosols. The results of this pre-processing step are split in 10°by10° tiles (L2 Tiles) for further processing. From a processing point of view, the number of available P-segments (typical 25 to 30) for a given day determines the amount of parallelism that can be achieved in the operations.

The normalization step is applied at a dekad frequency on surface reflectances acquired during a moving temporal window of 30 days for each spectral band. This processing step is initiated whenever the necessary L2 tiles are available and qualified. The number of tiles required depends on the latitude and the temporal period, as during the European winter there are less Nordic segments available from SPOT-VEGETATION (see http://www.vgt.vito.be). The degree of parallelism that can be achieved in the operations is dependent on the number of available L2 tiles at the given dekad (typical around 300). Each thread results in a 10°by10° tile (L3A tile).

The third step in the processing chain is the execution of the biophysical algorithm which is from an operational point of view straightforward. The input, a single L3A tile, and the result, a 10° by 10° tile (BIO tile) that contains both LAI and fAPAR value. The degree of parallelism that can be achieved is again identical to the amount of available tiles. The operator performs a visual qualification which acts as a mandatory condition to start the last step of the processing chain.

The fourth and last step in the processing chain is formatting the processed data into the actual LAI and fAPAR products. The products are formatted as Hierarchical Data (HDF) providing several image bands: the variable, its uncertainty, the quality flag, the number of input observations and the land-sea mask. The products are disseminated in tiles of 10° by 10° covering the land surfaces of the whole globe through the geoland2 web platform (http://www.geoland2.eu) where users can browse the catalogue, order the products after registration, and subscribe to receive the future products as soon as they are generated. The products are accompanied with by XML metadata file according INSPIRE standard [13] to support the Spatial Data Infrastructure services. The near real time GEOV1 products are disseminated in continents and countries tiles using the EUMETCast system to African and South American users.

An optional fifth step is used to generate a global composite quicklook as a quality control. When registering the quicklook to QCM, the validation status is set to check. An operator visually checks the composite and in case of errors flags the positions to reject the individual tiles, a less time consuming and more reliable process than checking all individual tiles.

Especially the re-processing of the archive benefits from the parallelism in the execution of the processing line, taking maximum advantage of the available resources (**Table 1**).

The processing line is currently in its final test stage and will soon start running in near real time. In parallel, the re-processing of the SPOT/VEGETATION archive from 1999 will be handled to build a 12-years time series of GEOV1 products.

Table 1 : GEOV1 Processing configurations for typical average sequential (AVG SEQ) and typical parallel (AVG PAR) modes

	Data size	AVG SEQ	Maximum Threads	AVG PAR
L2 processing	~70 GB	6 h /day	30	0.5 h / day
L3A processing	~300 GB	15 h /dekad	300	2 h /dekad
BIO processing	~40 GB	2 h /dekad	300	0.5 h /dekad
Total re-processing archive	n.a.	~ 3 Years	n.a.	~4 Months

5. Conclusion

Geoland2 BioPar provides GEOV1 LAI and fAPAR products based on SPOT-VEGETATION top of canopy directionally normalized reflectance values. They have been developed taking advantage of specific performances of existing global products. Their quality has been evaluated following criteria defined in the CEOS/LPV guidelines: their spatio-temporal profiles are smooth and consistent from year to year; the magnitude of values are realistic according to the land cover types; they compare well with ground measurement, reaching the targeted accuracy and showing no bias.

The introduction of a Job Manager and a Quality Control Manager provides a good reliability and enhances the timeliness to provide the GEOV1 products and the ability to monitor and report on system performances. The pre-operational chain is almost running in Near Real Time. The available EO archive, from 1998 onwards, will be re-processed off-line in the coming months and with the coming extension back to 1981 using the LTDR/AVHRR series of observation a long term time series comes available for applications to understand climate change and others, and hence can be seen as a contribution to ECV T10 and T11 datasets according GCOS. These global products are sustainable by the use of AVHRR/METOP sensor as backup and compatible using the PROBA-V (2012) and Sentinel-3 missions (2013) within the GMES Operations context. The spatial resolution will be further increased by using these latter missions to 300m in the near future. The products are provided with a free data access policy.

Acknowledgements

The research leading to the results presented in this paper has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n°218795. All these products are under copyright geoland2. The dissemination via EUMETCast was done in co-operation with the DevCoCast project (see http://www.devcocast.eu), funded under the EU 7th Framework Programme for Research.

References

[1] Lacaze R. et al., *Geoland2 - Towards an operational GMES Land Monitoring Core Service*: the Biogeophysical Parameter Core Mapping Service, Proceedings of the 34th ISRSE, 11th – 14th April 2011, Sydney.

[2] GCOS, 2006. GCOS-107. *Supplemental details to the satellite based component of the "implementation plan for the global observing system for climate in support of the UNFCCC"*. WMO/TN N° 1338, GCOS/WMO, Geneve (Switzerland).

[3] GCOS ECV T11 LAI, 2009. L*eaf Area Index - Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables (ECV)*, Version 10, May 25, 2009 (GTOS-66)

[4] GCOS ECV T10 FAPAR, 2009. *Fraction of Absorbed Photosynthetically Active Radiation - Assessment of the Status of the Development of the Standards for the Terrestrial Essential Climate Variables (ECV)*, Version 8, 11 May 2009 (GTOS-65)

[5] Chen, J.M. and Black, T.A., 1992. *Defining leaf area index for non-flat leaves*. Plant, Cell and Environment, 15: 421-429.

[6] McCallum, I. et al., 2009. *Satellite-based terrestrial production efficiency modeling*. Carbon Balance and Management, 4:8.

[7] Baret, F. , M. Weiss, R. Lacaze, F. Camacho, P. Pacholczyk, H. Makhmara, and B. Smets, *Consistent and accurate LAI, FAPAR, and FCover global products: principles and evaluation of GEOV1 products*, Proceedings of 3rd RAQRS, 27th Sept.- 1st Oct., Torrent, pp 208-213, 2010.

[8] Baret, F., Hagolle, O. et al., 2007. *LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION*. Part 1: Principles of the algorithm. Rem. Sens. Environ., 110: 275-286.Salzano R, R Salvatori & F Dominé, 2007.

[9] Morisette, J., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J., Garrigues, S., Shabanov, N., Weiss, M., Fernandes, R., Leblanc, S. et al., 2006. *Validation of global moderate resolution LAI Products: a framework proposed within the CEOS Land Product Validation subgroup*. IEEE Trans. Geosc. Rem. Sens., 44(7): 1804-1817.

[10] Baret, F. et al., 2006. *Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products.* Proposition of the CEOS-BELMANIP. IEEE Transactions on Geoscience and Remote Sensing, 44(7: special issue on global land product validation): 1794-1803.

[11] Camacho, F., J. Cernicharo, F. Baret, R. Lacaze and M. Weiss, *Quality assessment of the first version of geoland2 biophysical variables produced at global scale*, Proceedings of 3rd RAQRS, 27th Sept.- 1st Oct., Torrent, pp 660-665, 2010.

[12] Freeman, E., Sierra, K., Bates, B. 2004. *Head First Design Patterns*. O'Reilly Media, Inc.

[13] Drafting Team Metadata and European Commission Joint Research Centre. 2010. *INSPIRE Metadata Implementation Rules: Technical Guidelines based on EN ISO19115 and EN ISO19119* (Version 1.2)