GEOCLIM: A global climatology of LAI, FAPAR, and FCOVER from VEGETATION observations for 1999–2010

Aleixandre Verger a,b,c,*, Frédéric Baret c, Marie Weiss c, Iolanda Filella a,b, Josep Peñuelas a,b

a CREA, Cerdanyola del Vallés 08193, Catalonia, Spain
b CSIC, Global Ecology Unit, Cerdanyola del Vallés 08193, Catalonia, Spain
c INRA UMR114 EMMAH, Domaine Saint-Paul, Avignon 84914, France

Abstract

Land-surface modelling would benefit significantly from improved characterisation of the seasonal variability of vegetation at a global scale. GEOCLIM, a global climatology of leaf area index (LAI), fraction of absorbed photosynthetically active radiation (FAPAR)—both essential climate variables—and fraction of vegetation cover (FCOVER), is here derived from observations from the SPOT VEGETATION programme. Interannual average values from the GEOV1 Copernicus Global Land time series of biophysical products at 1-km resolution and 10-day frequency are computed for 1999 to 2010. GEOCLIM provides the baseline characteristics of the seasonal cycle of the annual vegetation phenology for each 1-km pixel on the globe. The associated standard deviation characterises the interannual variability. Temporal consistency and continuity is achieved by the accumulation of multi-year observations and the application of techniques for temporal smoothing and gap filling. Specific corrections are applied over cloudy tropical regions and high latitudes in the Northern Hemisphere where the low number of available observations compromises the reliability of estimates. Artefacts over evergreen broadleaf forests and areas of bare soil are corrected based on the expected limited seasonality. The GEOCLIM data set is demonstrated to be consistent, both spatially and temporally. GEOCLIM shows absolute differences lower than 0.5 compared with MODIS (GIMMS3g) climatology of LAI for more than 80% (90%) of land pixels, with higher discrepancies in tropical and boreal latitudes. ECOCLIMAP systematically produces higher LAI values. The phenological metric for the date of maximum foliar development derived from GEOCLIM is spatially consistent (correlation higher than 0.9) with those of MODIS, GIMMS3g, ECOCLIMAP and MCD12Q2 with average differences within 14 days at the global scale.

1. Introduction

The state and dynamics of vegetation play key roles in the carbon cycle and global climate. A set of essential climatic variables was identified as both accessible from remote sensing observations and involved in key processes (GCOS, 2010). Among those relating to land surfaces, the leaf area index (LAI) and the fraction of absorbed photosynthetically active radiation (FAPAR) can be derived from observations in the reflective solar domain. These biophysical variables of vegetation are crucial in several processes, including photosynthesis, respiration, and transpiration. LAI is defined as one-half the total area of green elements per unit area of horizontal ground (Chen & Black, 1992; GCOS, 2010). It controls the exchanges of energy, water, and greenhouse gases between the land surface and the atmosphere. FAPAR is defined as the fraction of radiation absorbed by the canopy in the 400–700 nm spectral domain under specified conditions of illumination and is a main input in models of light-use efficiency (McCallum et al., 2009). The fraction of vegetation cover (FCOVER), defined as the fraction of the background covered by green vegetation as seen from the nadir, is also as a very pertinent variable that can be used in models of the surface-energy balance to separate the contribution of the soil from that of the canopy (Gutman & Ignatov, 1998; Su, McCabe, Wood, Su, & Prueger, 2005).

LAI, FAPAR, and FCOVER are routinely estimated from sensors with medium resolution such as VEGETATION (Baret et al., 2013), Moderate Resolution Imaging Spectroradiometer (MODIS) (Myneni et al., 2002) and the Advanced Very High Resolution Radiometer (AVHRR) (Zhu et al., 2013). The European Copernicus Global Land Service delivers global LAI, FAPAR, and FCOVER products from SPOT VEGETATION data from 1999 to the present with a spatial sampling close to 1 km. The products, known as GEOV1 products, have benefited from the development and validation of existing products (Baret et al., 2013). Camacho, Cernicharo, Lacaze, Baret, and Weiss (2013) demonstrated that GEOV1 products were more accurate and precise than current products.

Some land-surface models (LSMs) for simulating terrestrial water and carbon cycles use the spatiotemporal variation of LAI or FAPAR, described by different lookup tables depending on the type of vegetation (Viterbo and Beljaars 1995). The availability of satellite data in the last
two decades describing the state and evolution of vegetation has allowed a better integration of biophysical variables into LSMs. Previous studies have demonstrated the improved performance of LSMs due to a better characterisation of the seasonal and interannual variability of vegetation functioning provided by the assimilation of satellite data. In particular, data assimilation yields a more realistic parameterisation in phenological models and reduces the models’ prediction errors by 21 and 15% for FAPAR and LAI, respectively (Stöckli, Rutishauser, Baker, Liniger, & Denning, 2011). A number of studies have shown the potential of assimilating LAI observations to correct vegetation model states (Demarty et al., 2007; Gu, Belair, Mahfouf, & Deblonde, 2006) and the implications of introducing the observed seasonal (van den Hurk, Viterbo, & Los, 2003) and interannual (Guillevic et al., 2002) variability of LAI in the annual cycle of hydrological fluxes. Boussetta, Balsamo, Beljaars, Kral, and Jarlan (2013) showed that the assimilation of a MODIS derived LAI monthly climatology, i.e. the interannual average of LAI time series (as opposed to a vegetation-dependent constant LAI), in a model of global numerical weather prediction improved the forecast of near-surface (screen-level) air temperature and relative humidity through its effect on evapotranspiration. Barbou, Calvet, Mahfouf, and Lafont (2014) more recently demonstrated the potential of the assimilation of GEOF1 LAI into an ISBA-A-gs land-surface model to improve the monitoring of droughts. A LAI climatology was also useful for the identification of anomalies and trends in global vegetation (Baret et al., 2012; Brandt, Verger, Diouf, Baret, & Samini 2014; Verger, Baret, Weiss, Kandasamy, & Vermote, 2013). The climatology of the biophysical variables reveals the seasonality inherent to the land-cover type and improves land-cover classification (Verhegghen, Bontemps, & Defourny, 2014). A climatology gap filling can better cope with missing and noise-contaminated data than can standard temporal filters for the most missing data or large gaps in a single annual time series of satellite data, which have a large impact on the accuracy of the phenological metrics extracted from the reconstructed time series (Guyon et al., 2011; Kandasamy, Baret, Verger, Neveux, & Weiss, 2013; Verger et al., 2013). Extraction of phenological information is also sensitive to the temporal (Pouliot, Latifovic, Fernandez, & Otholf, 2011; Zhang, Friedl, & Schaaf, 2009) and spatial (Fisher & Mustard, 2007; Kovalsky, Roy, Zhang, & Ju, 2011) resolution of the satellite data. The climatology derived from time series of moderate spatial resolution sensors preserves the high temporal frequency mandatory for phenological studies (Guyon et al., 2011). Finally, the climatology information can make projections and improve the stability of near real time estimates (Jiang, Liang, Wang, & Xiao, 2010; Verger, Baret, & Weiss, 2014).

Despite the significance of global phenology for earth system monitoring and modelling, there are few data sets that explicitly describe the annual vegetation cycle at global scale. Boussetta et al. (2013) derived a monthly LAI climatology from 2000–2008 MODIS observations to be used in a numerical weather prediction model as indicator of the leaf life development stage. The ECOCLIMAP programme (Faroux et al., 2013) with CYCLOPES products were selected for the training process. The selected MODIS and CYCLOPES products were selected for the training process. The selected products were combined after re-projection onto the VEGETATION Plate-Carrée 1/12° grid, smoothed over time, interpolated at the 1-day frequency, combined, and eventually re-scaled to better fit the expected range of variation. Further details for the training of the neural networks and the generation of the product are provided in Baret et al. (2013). Recent validation studies indicated that GEOF1 outperformed existing products in both accuracy and precision (Camacho et al., 2013). GEOF1 products are freely available at www1.

2. GEOF1 implementation

The generation of GEOF1 was achieved from GEOF1 time series for the period 1999–2010 based on the interannual means of biophysical variables and the application of specific corrections. The input data set and the steps required to produce GEOF1 are described hereafter.

2.1. GEOF1 biophysical products

GEOF1 biophysical products provide global coverage of LAI, FAPAR, and FCOVER from 1998/12/24 to the present at a ground sampling distance of 1/112° (approximately 1 km at the equator) and 10-day steps. A neural-network machine-learning algorithm was used to estimate GEOF1 products (Baret et al., 2013). Directionally normalised VEGETATION reflectances (Roujean, Leroy, & Deschamps, 1992) from the top of the canopy in the red, near-infrared, and short-wave infrared bands derived from the CYCLOPES processing line (Baret et al., 2007) were used as inputs to the neural networks. Based on the validation results for the available biophysical products (Garrigues et al., 2008; Weiss, Baret, Garrigues, Lacaze, & Bicheron, 2007), the MODIS and CYCLOPES products were selected for the training process. The selected products were combined after re-projection onto the VEGETATION Plate-Carrée 1/112° grid, smoothed over time, interpolated at the 1-day frequency, combined, and eventually re-scaled to better fit the expected range of variation. Further details for the training of the neural networks and the generation of the product are provided in Baret et al. (2013).

2.2. Climatology generation

For the generation of GEOF1 only the GEOF1 biophysical products with the best quality were selected according to the quality flags on snow, aerosol, reflectance input and biophysical output status (Baret, Weiss, & Kandasamy, 2010). The climatology is defined as the interannual mean of the best quality GEOF1 products accumulated for 1999–2010. It is generated at the pixel scale (1-km spatial resolution) and at a dekadal temporal step (a 10-day period, with 36 dekads per year) within a 30-day compositing window (±15 days). The average values are then computed from three adjacent dekads instead of only 1 dekad along the 12-year period, which allows an increase in the number of points (12 × 3 = 36 compared to only 12), provides more robust and continuous estimates, and induces fewer artefacts because the
dynamics of the products are approximately linear between the 3 dekads (Baret et al., 2010), as shown by Camacho et al. (2013) for the smoothness of the GEOV1 products. A temporal smoothing and gap-filling (TSGF) technique (Verger, Baret, & Weiss, 2011) was applied to correct the residual artefacts, especially when the GEOV1 products were systematically unavailable across the years due to cloud coverage, and to ensure continuity and consistency in GEOCLIM as phenological studies request. Gap filling was achieved by linear interpolation. Temporal smoothing relied on an adaptive Savitzky–Golay second-degree polynomial fitting by processing three valid values on either side of each date (Verger et al., 2011). The composting window may be asymmetric due to possible missing data. TSGF technique was demonstrated to improve other existing temporal filters for reconstruction satellite LAI time series in terms of the accuracy as compared to the original data by ensuring robustness to noise and missing data, while preventing over-smoothing (Kandasamy et al., 2013; Verger et al., 2011, 2013). An example of a climatology computation is illustrated in Fig. 1, and TSGF correction is illustrated in Fig. 2a.

2.3. Correction of specific artefacts

The generated climatology was then corrected for specific problematic behaviours based on available expert knowledge:

- Some artefacts were observed at northern high latitudes during the winter: anomalous seasonality and unexpected increases in LAI (FAPAR, FCOVER) (Fig. 2b) with an artificial maximum peak in winter (Fig. 2b) and high interannual variability resulting in high standard deviations (Fig. 3a). These artefacts were mainly due to snow cover or very poor conditions of illumination that limited the number of valid observations and the reliability of the bidirectional reflectance model applied for the correction of VEGETATION data (Roujean et al., 1992) (Figs. 2b, 3b). The LAI (FAPAR, FCOVER) values are expected to be relatively stable and low due to the low temperatures, short days, and low illumination during winter at these high latitudes. To correct these artefacts at northern latitudes, the GEOCLIM inputs higher than the 20th percentile during winter (defined here as the period for which the sun zenith angle, SZA > 70° at the time of VEGETATION overpass, i.e. around 10:30) were fixed at the minimum pixel values observed over the entire period. We used the minimum values by preferentially selecting the values computed from at least three valid observations because the quality of GEOCLIM inputs is highly correlated with the number of valid observations available for their composition (Figs. 2b, 3). We used the minimum value computed over all dekads if none of the dekads verified this condition. The areas where this specific correction was applied are shown in Fig. 4a. Similar approaches based on representative winter values and thresholds to fill gaps and correct values affected by snow or poor illumination at high latitudes were also considered by Beck, Atzberger, Hogda, Johansen, and Skidmore (2006), Delbart, Kergoat, Le Toan, Lhermitte, and Picard (2005), Zhang, Friedl, Schaaf, and Strahler (2004).

- Significant artefacts were also detected at equatorial and tropical latitudes due to aerosol-cloud contamination that produced high instabilities, artificial seasonalities, and missing data in the GEOV1 products and, consequently, in the derived output (Fig. 2c). The high standard deviations (Fig. 3a) and the low number of available observations (Figs. 2c, 3b) appeared to be good indicators of the high uncertainty associated with the computed output over these tropical areas. Most of these cases corresponded to evergreen broad-leaf forests (EBFs) (cf. Figs. 3, 4b), so a minimum seasonality and high LAI values should be observed. We thus identified a pixel as an EBF if the 90th percentile (P90) of the LAI output was >4.5 and the 20th percentile was >P90-1.5. This method for the detection of EBFs based only on GEOV1 products (Fig. 4a) agreed well with the GLOBCOVER land-cover map (Defourny et al., 2009) (Fig. 4b). For EBFs, the GEOCLIM values were fixed to the 90th percentile computed over the entire period (Fig. 2c).

- Some artefacts were also detected in the raw output for areas of bare soil (BS) where the observed seasonality was of the same order of magnitude of the precision of the GEOV1 product (Fig. 2d). A pixel was identified as BS if the 90th percentile of the LAI output was <0.05 (compare Fig. 4a to the land-cover map in Fig. 4b). For BS, the GEOCLIM values were fixed to the 50th percentile computed over the entire period (Fig. 2d).

3. Evaluation of GEOCLIM

This section assesses the performance of GEOCLIM. The validation data sets are first described. The spatial and temporal consistencies of GEOCLIM are discussed and evaluated by comparison with climatologies derived from both AVHRR and MODIS data. For the sake of brevity, the results focus on LAI, because of the three variables LAI, FAPAR, and FCover, GEOCLIM is used most by the scientific community. The comparison was performed at 0.5° spatial sampling on a Plate-Carrée grid. The 0.5° spatial resolution corresponds to the typical resolution of global models and reduces computation time. For comparison purposes, the different LAI data sets were averaged at monthly time step because one month corresponds to the lowest temporal sampling among the validation datasets.

The global phenology derived from the temporal seasonality of GEOCLIM is also investigated in this section. For simplicity, we focus on the date of maximum foliar development, i.e. the timing of the peak of the growing season in the LAI annual cycle (Brown, de Beurs,
Fig. 2. Illustration of TSGF correction for LAI for four GLOBCOVER biome classes (Defourny et al., 2009): (a) grassland, (b) needleleaf forest, (c) evergreen broadleaf forest, and (d) bare soil. The thin lines and grey intervals represent the raw model output computed as interannual means and the associated standard deviations. The thick lines represent the final GEOCLIM output corrected for artefacts. The dashed lines represent the number of available observations (Nb). The latitudes and longitudes of the sites are indicated. DOY, day of the year.

Fig. 3. Global maps of (a) the maximum standard deviations (Max. SD) of interannual LAI values observed over the 36 dekads and (b) the minimum number (Min. Nb) of valid observations over the 36 dekads for computing GEOCLIM. The areas in grey correspond to pixels with no data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
In addition to the phenological metrics derived from the LAI climatologies, we use also the MCD12Q2 MODIS phenological product (Zhang et al., 2003).

3.1. Validation data sets

3.1.1. ECOCLIMAP product

ECOCLIMAP is a database at 1/112° resolution on a Plate-Carrée grid resolution that includes a classification of ecosystems and a consistent set of associated land-surface variables, including LAI, at 10-day temporal sampling (Faroux et al., 2013). We used the latest Open-ECOCLIMAP v1 version, available since June 2014 at www2. It combines the global database of the first version, ECOCLIMAP-I, and an upgraded version, ECOCLIMAP-II, for Europe. ECOCLIMAP-I contains 215 ecosystems obtained by combining existing land covers, climatic maps, and NDVI seasonal profiles from AVHRR data acquired between April 1992 and March 1993 (Masson et al., 2003). For each class of vegetation, the maximum and minimum LAI values are fixed based on in situ knowledge, and the annual cycle of LAI is constrained by the NDVI AVHRR temporal profiles using a linear relationship between NDVI and LAI. The second version, ECOCLIMAP-II, contains 573 ecosystems across Europe based on more recent land-cover maps, and the annual LAI profiles are derived from MODIS Collection 5 for the years 2002–2006 (Faroux et al., 2013).

The ECOCLIMAP LAI data set at the original 1/112° resolution was aggregated at 0.5° spatial sampling and averaged at monthly temporal sampling.

3.1.2. MODIS climatology

The MODIS Collection 5 Boston University (BU) LAI product at 0.25° latitude/longitude grid is the extracted best quality of standard MODIS LAI product based on MOD15A2 and MOD13A2 quality flags (Samanta et al., 2011). The standard MODIS LAI products relies on a biome dependent look-up table inversion of a radiative transfer model which ingests red and near infrared bidirectional reflectance factor values, their associated uncertainties, the view-illumination geometry, and biome type (within eight types based on MOD12Q1 land cover map). Further details on the retrieval algorithm are provided in Myneni et al. (2002), Yang, Tan, et al. (2006). Valid 1 km 8-day values are averaged to obtain monthly LAI (Samanta et al., 2011). The monthly LAI 1 km sinusoidal product is aggregated and projected onto a 0.25° Plate-Carrée projection.

We derived a monthly MODIS climatology as the interannual average from 2000 to 2010 MODIS BU LAI product. The 0.25° product was aggregated to 0.5° spatial resolution for comparison purposes.

3.1.3. GIMMS3g climatology

The GIMMS3g LAI product derived from AVHRR data is available at 15-day temporal steps and 1/12° spatial resolution for the period July 1981 to December 2011. The principles used for the generation of this LAI data set are based on the use of neural networks which were trained first with GIMMS NDVI3g and MODIS LAI products for the overlapping period 2000–2009. The trained neural network algorithm is then applied using the land-cover class, the latitude and longitude coordinates, and the NDVI3g as the input data to generate the full temporal coverage of the GIMMS3g LAI data set. Further details on the algorithm for GIMMS3g retrieval can be found in Zhu et al. (2013).

We derived the GIMMS3g climatology as the interannual mean of GIMMS3g LAI time series at 15-day temporal step for the period 1999–2010. We aggregated the 1/12° products at 0.5° spatial resolution and averaged at monthly temporal sampling for comparison purposes.

3.1.4. MCD12Q2 product

MCD12Q2 (Collection 5) (Ganguly et al., 2010; Zhang et al., 2003) provides global yearly vegetation phenologies at 500 m from 2001 to 2010 MODIS time series. The MCD12Q2 algorithm uses a series of piecewise logistic functions fitted over the annual cycle of EVI data (Zhang et al., 2003). Among the transition dates provided by the MCD12Q2 product, specific corrections were applied in GEOCLIM.
product, the parameter “onset of greenness maximum” is used here for comparisons with the parameter “peak of growing season” derived from the LAI climatologies, i.e. the date for which the climatology reaches its maximum value in the LAI annual cycle. The onset of maximal greenness conceptually corresponds to the transition date at which the annual cycle of the vegetation reaches maturity. This date is

Fig. 5. GEOCLIM global maps of (a) the maximum LAI at the peak of the growing season (Max. LAI), (b) the mean annual LAI (Mean LAI), (c) the standard deviation of interannual LAI values for the date of the peak (interannual variability) (SD Max. LAI), and (d) the standard deviation of the mean LAI annual cycle (seasonal variability) (SD Mean LAI). The areas in dark grey correspond to pixels with no data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
thus expected to be earlier than the date of maximum vegetation. Zhang, Friedl, and Schaaf (2006) compared the MCD12Q2 parameter to in situ measurements and found that it corresponded to the time at which 85–90% of the individual leaves reached their final size.

The MCD12Q2 500 m sinusoidal product was projected onto a 0.5° Plate-Carrée projection using the MODIS re-projection tool (www3). Yearly MCD12Q2 values from 2001 to 2010 were then averaged to provide a typical phenology for comparison with the phenological metrics derived from the LAI climatology data sets.

3.2. Spatiotemporal consistency

The GEOCLIM biophysical variables had highly consistent spatial and temporal patterns (Fig. 5a), in agreement with the global distributions of biomes (Fig. 4b). The seasonal patterns of GEOCLIM LAI also reflected the expected regimes of vegetation at the global scale. Evergreen broadleaf forests exhibited null seasonality (Fig. 5d) in the tropical belt where LAI was near 5 throughout the year (Fig. 5b). Deserts also had no seasonality (Fig. 5d) where LAI was near zero (Fig. 5b). These results were expected given the forcing applied for evergreen broadleaf forests and bare soils (cf. Section 2.3). As expected, deciduous broadleaf forests and crops had the highest seasonalities (Fig. 5d). The observed seasonality in needleleaf forests (Fig. 5d) with a LAI \( \leq 4 \) (Fig. 5a) and means near 2 (Fig. 5b) agreed with the observed seasonality of the understory layer, which can reach a LAI of \(-2\) or more in summer but which is often near zero in winter (Chen, Rich, Gower, Norman, & Plummer, 1997; Jiao, Liu, Liu, Pisek, & Chen, 2014; Masson et al., 2003).

The areas with the highest interannual variabilities in GEOCLIM (Fig. 5c) corresponded to cropland in the USA and Eurasia, with intrinsic variabilities due to crop rotation or management, but also regions of severe drought and fire in South America, Africa, and Asia, regions of land-cover change such as the deforestation in Amazonian and Indonesian

![Fig. 6. Maps of the mean LAI differences between (a) GEOCLIM and ECOCLIMAP, (b) GEOCLIM and MODIS, and (c) GEOCLIM and GIMMS3g. The percentage of land pixels for each interval of mean LAI differences is indicated on the right of the colour bar. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
tropical forests, and regions of extreme events such as drought and heat waves in Europe, eastern China, and Australia. High interannual variability, however, may also indicate a problem with the computed GEOCLIM value due to instabilities in the GEOV1 data or to insufficient available GEOV1 data, as observed in the Gulf of Guinea (compare Figs. 3b and 5c). In most regions, the interannual variability (Fig. 5c) was significantly lower than the seasonal variability (Fig. 5d), demonstrating that GEOCLIM provided a baseline vegetation annual cycle that was representative of the current phenology and that smoothed most of the anomalies.

3.3. Comparison with ECOCLIMAP, MODIS and GIMMS3g climatologies

The map of annual mean differences between GEOCLIM and the LAI climatologies derived from AVHRR and MODIS data (Fig. 6) shows LAI differences of ±0.5 for 54%, 83% and 91% of the land pixels as compared with ECOCLIMAP, MODIS and GIMMS3g, respectively. GEOCLIM produced systematically lower values than ECOCLIMAP for the remaining 46% of pixels, with larger differences for dense forests (northern boreal and tropical forests) but with significant differences also for crops (e.g. USA and eastern Asia) (Fig. 6a). These systematic negative bias of GEOCLIM as compared to ECOCLIMAP was evident across latitudes and along the year (Fig. 7). GEOCLIM produced also systematically lower values than MODIS (Fig. 6b) over tropical forests with differences ~0.5 along the year (Fig. 7b) and over northern deciduous broadleaf forest during the maximum growing leaf development (lower frequencies for the maximum values in Fig. 8). On the contrary GEOCLIM produced slightly higher LAI values than GIMMS3g (Fig. 6c) over Amazon and Indonesian evergreen broadleaf forests and over boreal needle leaf forests in Russia and USA during the winter time (Fig. 7d).

Despite the large discrepancies in the magnitude of LAI between the different datasets, due in part to the differences in sensors and retrieval algorithms, seasonality and its phasing generally agreed well (Fig. 7). Seasonality was inverted in the Southern Hemisphere relative to the Northern Hemisphere (compare Fig. 7a and c). In the Northern Hemisphere, LAI seasonality decreased in the length of season (active growth period) with latitude (compare Fig. 7c and d). In the tropical latitudes (~20°–10°) characterised by very limited seasonality GEOCLIM and GIMMS3g systematically showed lower values than MODIS and ECOCLIMAP (Fig. 7b). The largest differences in terms of seasonality were in the Northern Hemisphere at high latitudes (40°–70°) where ECOCLIMAP produced longer growing seasons as compared to other LAI datasets and higher values in the period of active growth (Fig. 7d). Nevertheless, all the data sets agreed well for the base level of LAI during the dormancy period for the 40°–70° latitudes validating a posteriori the reliability of the specific correction applied in winter (SZA > 70°) to the GEOCLIM values for high northern latitudes (Section 2.3).

Histograms of the LAIs (Fig. 8) indicated very similar distributions between the different LAI datasets for shrubs/savannah/bare soil. Some similarities in the position of the maximum frequency were also observed for crops and grassland. Some discrepancies, however, were observed: ECOCLIMAP produced low frequencies for LAIs of zero at the expense of higher intermediate values, while GEOCLIM, MODIS and GIMMS3g produced a smoother transition. Forests had the largest discrepancies between the different data sets. GEOCLIM and GIMMS3g produced a bimodal distribution for deciduous broadleaf forests, with a peak for low values (LAI near 1) corresponding to the dormant period of the vegetation in winter and a second mode for the period of active growth with values higher than 6 in few occasions. MODIS produced also a peak for LAI near 1 and a smooth transition up to maximum values around 6.5. ECOCLIMAP produced an even distribution for deciduous forests, but with unrealistic peaks. Evergreen broadleaf forests had relatively consistent narrow distributions between GEOCLIM, ECOCLIMAP and MODIS but with significant differences in the magnitudes (i.e. similar shapes but shifted distributions). The LAI modes were 5 for GEOCLIM, and 6 for ECOCLIMAP and MODIS. GIMMS3g produced broader distributions with the LAI mode ~4. Needleaf forests had similar distributions for GEOCLIM, MODIS and GIMMS3g but with higher frequencies for low values compared to those in ECOCLIMAP. The LAI mode around 1 for GEOCLIM (MODIS and GIMMS3g) for deciduous broadleaf forests (Fig. 8c) and needleleaf forests (Fig. 8e) corresponded to the winter LAI value and reproduced the expected seasonality in northern high latitudes (Fig. 7d) while ECOCLIMAP produced unrealistic LAI distributions and peaks’ locations (Fig. 8c and e).

3.4. Assessment of global phenology

The spatial pattern of the phenology derived by GEOCLIM (Fig. 9) reflected the distributions of climate and biome type (Fig. 4b). Seasonality was strongly dependent on temperature at northern latitudes >30°, and the timing of maximum greenness had a clear latitudinal gradient indicating a delay in the date of peak development with latitude (Fig. 10). In other regions, seasonality had more complex spatial patterns that were driven mostly by biome type, land use, and the seasonal variation in rainfall (Fig. 9).

The phenological metrics (Fig. 10) were spatially consistent in the timing of the maximum of the growing season as derived from the different data sets and particularly between GEOCLIM and MODIS (GIMMS3g) with uncertainties of around 14 days in terms of RMSE, bias of less than 1 day, a correlation higher than 0.95 and a slope of the linear regression close to the unity (Table 1). The phenology derived from ECOCLIMAP was also highly spatially consistent with GEOCLIM (correlation about 0.9, slope close to the unity and bias about 6 days, Table 1) but it diverged to some degree (uncertainties of about one
month in terms of RMSE), mostly in the Southern Hemisphere (Fig. 10) in regions with a limited seasonality (Fig. 5d). As expected (Section 3.1.4), the phenological phase for the “onset of greenness maximum” retrieved in MCD12Q2 occurred earlier than the peak date in GEOCLIM (bias of 14 days, Table 1) due to differences in the definitions of the phenological metrics. The phenological metrics derived from GEOCLIM constitutes an intermediate solution across latitudes between MODIS, GIMMS3g, ECOCLIMAP and MCD12Q2 for the date of maximum foliar development (Fig. 10).

4. Discussion

Twelve years (1999–2010) of data from GEOV1 LAI, FAPAR, and FCOVER products were used to compute GEOCLIM outputs for the interannual average seasonal cycle at a pixel scale. The main assumptions were that (i) no land-cover change or abrupt disturbance leading to a change in the phenological annual cycle occurred for the period considered and (ii) the time series were sufficiently long to reduce the sensitivities of the averages to anomalies. Specific correction were applied at northern high latitudes, bare soils and evergreen broadleaf forests to overcome problems associated, respectively, with strong bidirectional effects and snow cover, precision and signal to noise ratio, and aerosol-cloud contamination (Section 2.3). The identification of bare soil and evergreen broadleaf forests was completely driven by the data avoiding possible mis-classification errors introduced by external land cover map information though a good spatial consistency with GLOBCOVER map was observed. In these problematic cases, GEOCLIM was forced to fixed values derived from the input data at the pixel level under the following hypothesis: (i) minimum vegetation activity in winter time at northern latitudes, and no seasonality in (ii) desert areas and (iii) evergreen broadleaf forests where the vegetation is respectively low (LAI ~ 0) and high (LAI ~ 5) throughout the year. The last hypothesis constitutes an oversimplification of the reality because of the possible seasonality of evergreen broadleaf forests. The high uncertainty associated with the data due to poor atmospheric correction and very high cloud occurrence in equatorial and tropical latitudes prevented the extraction of meaningful phenology at the resolution of the individual pixels of 1 km. The high spatial and temporal resolution of forthcoming Sentinel2 sensors should improve the monitoring of vegetation in these problematic areas.

GEOCLIM was indirectly validated based on the comparison with AVHRR and MODIS derived climatologies of LAI. Multitemporal ground data would be preferable for validating GEOCLIM but unfortunately were rarely available. GEOCLIM showed a high agreement with
MODIS (GIMMS3g) climatology of LAI and absolute differences were higher than the Global Climate Observing System (GCOS, 2010) requirements for accuracy, i.e. 0.5 LAI, only in northern boreal and tropical forests representing less than 20% (10%) of land pixels. GEOCLIM systematically produced lower values than MODIS over evergreen broadleaf forest as also observed in the comparison between GEOV1 and MODIS (Camacho et al., 2013; Fang et al., 2013). The difficult observational conditions in tropical latitudes with persistent clouds can cause irregularities in the solution and thus variable but systematic underestimations of LAI (Verger et al., 2011). The specific correction applied to GEOCLIM removed the instabilities in the solution but cannot correct possible biases in the magnitude of original GEOV1 products used as input data for GEOCLIM. Previous studies have also shown that GEOV1 products produce slightly higher values than MODIS for needleleaf forest in winter (Fang et al., 2013). The specific correction applied in GEOCLIM at northern high latitudes reduced these differences but may result in some underestimation of the seasonal amplitude in winter time. Accurate estimation of LAI in needleleaf forests in winter is challenging because contamination by clouds and snow limits the reliability of the reflectances used as inputs in the algorithms (Camacho et al., 2013). Further, the strong bidirectional effects of surface-reflectance at very high latitudes are not well simulated by the radiative transfer models currently used for product generation (Yang, Shabanov et al. 2006). In addition, the understory and foliage clumping are not well accounted for (Jiao et al., 2014; Pisek, Chen, Alikas, & Deng, 2010).

LAI values were systematically higher for ECOClimAP than for GEOCLIM, MODIS and GIMMS3g. Boussetta et al. (2013) reported similar higher LAI values for ECOClimAP than for MODIS. Garrigues et al. (2008) also reported large positive biases for ECOClimAP compared with CYCLOPES, MODIS and GLOBCARBON and with ground measurements. The differences in the temporal period, input data and sensors (VEGETATION for GEOCLIM and AVHRR and MODIS for ECOClimAP, Section 3.1) can partially account for the significant discrepancies between GEOCLIM and ECOClimAP although the relatively good agreement of GEOCLIM with MODIS and GIMMS3g AVHRR derived LAI climatologies indicates that the major source of discrepancies are related to the retrieval algorithms. The linear relationship between NDVI and LAI used to retrieve the ECOClimAP product for pixels out of Europe (Masson et al., 2003) may have introduced some overestimation because the LAI-NDVI relationship is exponential and saturates at medium to high values (e.g. Myneni et al. (2002)). Since ECOClimAP assumes low spatial variability within each class of land cover, it is limited to capture the LAI spatial variability as compared to other LAI datasets (Garrigues et al., 2008). Nevertheless, identifying the source of the differences between ECOClimAP and the other LAI datasets being analysed would require further attention and it is out of the scope of this paper.

Further research should focus on the development of improved LAI datasets with due attention to areas (boreal and tropical latitudes) and periods (winter time) where higher uncertainties exist (Fang et al., 2013). In these cases characterised by high level of noise and missing data, the use of the climatology and temporal smoothing and gap filling techniques applied at daily estimates of biophysical variables may increase the robustness of the solution as compared to the classical composition techniques (Verger, Baret, & Weiss, 2014).

The phenological metrics derived from GEOCLIM was highly spatially consistent (correlation higher than 0.9) with MODIS and AVHRR derived phenologies, including ECOClimAP ones, for the date of maximum foliar development with differences lower than six days in all cases except when compared with MCD12Q2 product (systematic biases of 14 days) due to the differences in the definition. A standardization in the definitions of the phenological metrics appears necessary (White et al., 2009). Disentangling the mechanisms that govern the seasonal and interannual variability in phenology and vegetation-climate dynamics at the global scale would require further analysis.

### Table 1

<table>
<thead>
<tr>
<th>RMSE</th>
<th>Bias</th>
<th>ρ</th>
<th>R</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEOCLIM-MCD12Q2</td>
<td>24.18</td>
<td>14.36</td>
<td>19.46</td>
<td>0.94</td>
</tr>
<tr>
<td>GEOCLIM-ECOClimAP</td>
<td>30.37</td>
<td>5.98</td>
<td>29.78</td>
<td>0.89</td>
</tr>
<tr>
<td>GEOCLIM-MODIS</td>
<td>14.38</td>
<td>0.97</td>
<td>14.35</td>
<td>0.98</td>
</tr>
<tr>
<td>GEOCLIM-GIMMS3g</td>
<td>17.65</td>
<td>-0.41</td>
<td>17.64</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The root mean square error (RMSE), bias, standard deviation (σ), correlation coefficient (R), and slope of the regression line through the origin for comparisons between the dates of maximum foliar development derived from GEOCLIM, ECOClimAP, and MCD12Q2 at a global scale at 0.5°. The areas of bare soil and evergreen broadleaf forests (Fig. 4a) with insufficient seasonality for computing the phenological metrics and 10% outliers were not included.

**Fig. 10.** Latitudinal transects at resolution of 0.5° of the average day of the year (DOY) for maximum foliar development derived from GEOCLIM and mean differences as compared to the phenological metrics derived from ECOClimAP, MODIS, GIMMS3g and MCD12Q2.
Acknowledgements

This research was partially supported by the European observation programme Copernicus (388533), Pôle Thématique Surfaces Continentales THEIA (CNES 140570) and the FP7 geoland2 (217975), GIOBIO (32-566) and LONGLOVE (32-594) projects. This research was also supported by the Spanish Government grant CGL2013-48074-P, the Catalan Government grant SGRC2014-274, and the European Research Council Synergy grant ERC-2013-Syg-610028 IMBALANCE-P. Alexandre Verger is the recipient of a Juan de la Cierva postdoctoral scholarship from the Spanish Ministry of Science and Innovation. The authors are indebted to Ranga Myneni, Taejin Park, the Boston University, MODIS, CNES, VITO, and CNRM-GAME (Météo France, CNRS) teams for providing access to the data.

References


Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., et al. (2013). Global data sets of vegetation Leaf Area Index (LAI3g) and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011. Remote Sensing, 5, 927–948.

WWW sites

www1: GEOVI Biophysical Products.

www2: ECOClimap code and data.

www3: MODIS Reprojection Tool.