Plant phenology evaluation of CRESCENDO land surface models. Part I: start and end of growing season.

Daniele Peano[1], Deborah Hemming[2], Stefano Materia[1], Christine Delire[3], Yuanchao Fan[4,5], Emilie Joetzjer[3], Hanna Lee[4], Julia E.M.S. Nabel[6], Taejin Park[7,8], Philippe Peylin[9], David Wårlind[10], Andy Wiltshire[2,11], and Sönke Zaehle[12]

[1]Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, CSP, Bologna, Italy
[2]Met Office Hadley Centre, Exeter, UK
[3]Centre National de Recherches Météorologiques, UMR3589, Université de Toulouse/Météo-France/CNRS, Toulouse, France
[4]NORCE Norwegian Research Centre AS, Bjerknes Centre for Climate Research, Bergen, Norway
[5]Center for the Environment, Harvard University, Cambridge, USA
[6]Max Planck Institute for Meteorology, Hamburg, Germany
[7]NASA Ames Research Centre, CA, USA
[8]Bay Area Environmental Research Institute, CA, USA
[9]Laboratoire des Sciences du Climat et l'Environnement, Gif-sur-Yvette, France
[10]Department of Physical Geography and Ecosystem Science, Faculty of Science, Lund University, Sweden
[11]Global Systemss Insitute,University of Exeter, Exeter, UK
[12]Max Planck Institute for Biogeochemistry, Jena, Germany

Correspondence: Daniele Peano (daniele.peano@cmcc.it)

Abstract. Plant phenology plays a fundamental role in land-atmosphere interactions, and its variability and variations are an indicator of climate and environmental changes. For this reason, current land surface models include phenology parameterizations and related biophysical and biogeochemical processes. In this work, the climatology of beginning and end of the growing season, simulated by the land component of seven state-of-the-art European Earth System models participating in the CMIP6, is evaluated globally against satellite observations. The assessment is performed using the vegetation metric leaf area index and a recently-developed approach, named four growing season types. On average, the land surface models show a 0.6-month delay in the growing season start, while they are about 0.5 months earlier in the growing season end. Difference with observation tends to be higher in the Southern Hemisphere compared to the Northern Hemisphere. High agreement between land surface models and observations is exhibited in areas dominated by broad-leaf deciduous trees, while high variability is noted in regions dominated by broad-leaf deciduous shrubs. Generally, the timing of the growing season end is accurately simulated in about 25% of global land grid points versus 16% in the timing of growing season start. The refinement of phenology parameterization can lead to better representation of vegetation-related energy, water, and carbon cycles in land surface models, but plant phenology is also affected by plant physiology and soil hydrology processes. Consequently, phenology representation and, in general, vegetation modelling is a complex task, which still needs further improvement, evaluation, and multi-model comparison.

1 Introduction

Plant phenology and its variability have a substantial influence on the terrestrial ecosystem (e.g. *Churkina et al.*, 2005; *Kucharik et al.*, 2006; *Berdanier and Klein*, 2011) and land-atmosphere interactions (e.g. *Cleland et al.*, 2007; *Richardson et al.*, 2013; *Keenan et al.*, 2014). Moreover, recent decades observations show modifications in both spring and autumn phenology under global warming (e.g. *Menzel et al.*, 2006; *Richardson et al.*, 2013; *Park et al.*, 2016; *Zhu et al.*, 2016; *Chen et al.*, 2020; *Zhang et al.*, 2020). For these reasons, phenology variability is one of the indicators of climate change (e.g. *Schwartz et al.*, 2006; *Soudani et al.*, 2008; *Jeong et al.*, 2011).

Given the influence of plant phenology on vegetation productivity, and since green leaves are the primary interface for the exchange of energy, mass (e.g., water, nutrient, and CO_2), and momentum between the terrestrial surface and the planetary boundary layer (*Richardson et al.*, 2012), land surface models (LSMs) need to accurately simulate plant growing season cycles. Limitations may result in biases and uncertainties in representing vegetation productivity and carbon cycle (e.g. *Churkina et al.*, 2005; *Kucharik et al.*, 2006; *Berdanier and Klein*, 2011; *Richardson et al.*, 2012; *Friedlingstein et al.*, 2014; *Savoy and Mackay*, 2015; *Buermann et al.*, 2018). For example, *Kucharik et al.* (2006) show an overestimated April-May net ecosystem production triggered by biases in plant budburst. *Berdanier and Klein* (2011) describe a link between above ground net primary production, growing season length, and soil moisture in high-elevation meadows. They show that the potential impact of changes in active growing season length on biomass production accounts for about 3-4 g m^{-2} d^{-1}. The work by *Richardson et al.* (2012) is an example of a systematic evaluation of LSMs' phenology representation. They evaluate fourteen models participating in the North American Carbon Program Site Synthesis against ten forested sites, within the AmeriFlux and Fluxnet-Canada networks. Their assessment reveals a typical bias of about two weeks in LSMs representation of the beginning and end of the growing season. They also show a low skill in LSMs' reproduction of the observed inter-annual phenology variability. These biases lead to an overestimation of about 235 gC m^{-2} yr^{-1} in the gross ecosystem photosynthesis of deciduous forest sites. However, uncertainties in simulated maximum production partially balance this overestimation. The work by *Buermann et al.* (2018) is another example of a multi-LSMs evaluation. They observe widespread lagged plant productivity responses across northern ecosystems associated with warmer and earlier springs, which is weakly captured by ten evaluated TRENDYv6 current LSMs. Consequently, current LSMs still present biases in simulating timings and the magnitude of the vegetation active season.

The latest generation of LSMs have started including a more detailed description of land biophysical and biogeochemical processes, and they have become able to explicitly represent carbon and nitrogen land-cycles, as well as plant phenology and related water and energy cycling on a global scale (e.g. *Oleson et al.*, 2013; *Lawrence et al.*, 2018). In particular, current LSMs link Leaf Area Index (LAI) and plant phenology to changes in temperature, precipitation, soil moisture and light availability (e.g. *Oleson et al.*, 2013; *Lawrence et al.*, 2018), as displayed in observations (e.g. *Caldararu et al.*, 2012; *Zeng et al.*, 2013; *Tang and Dubayah*, 2017). Besides, some LSMs use satellite-based data assimilation as a tool to constrain the parameters of phenology schemes (e.g. *Knorr et al.*, 2010; *Stöckli et al.*, 2011; *MacBean et al.*, 2015).

In this framework, the European CRESCENDO project (https://www.crescendoproject.eu/) fostered the development of a new generation of LSMs to be used as the land component of the Earth System Models (e.g. *Smith et al.*, 2014; *Olin et al.*,

2015; *Cherchi et al.*, 2019; *Mauritsen et al.*, 2019; *Sellar et al.*, 2020; *Seland et al.*, 2020; *Yool et al.*, 2020; *Boucher et al.*, 2020) employed in the Coupled Model Intercomparison Project Phase 6 (CMIP6, *Eyring et al.*, 2016). In particular, seven novel LSMs, which are part of the CRESCENDO effort, are used in this work, namely Community Land Model (CLM) version 4.5 (*Oleson et al.*, 2013) and version 5.0 (*Lawrence et al.*, 2019), JULES-ES (*Wiltshire et al.*, 2020a), JSBACH (*Mauritsen et al.*, 2019), LPJ-GUESS (*Lindeskog et al.*, 2013; *Smith et al.*, 2014; *Olin et al.*, 2015), ORCHIDEE (*Krinner et al.*, 2005), and ISBA-CTRIP (*Decharme et al.*, 2019).

Given the relevance of plant phenology and its changing variability related to climate, LSMs need routine evaluation against observations (e.g. *Jolly et al.*, 2005; *Richardson et al.*, 2012; *Dalmonech and Zaehle*, 2013; *Kelley et al.*, 2013; *Murray-Tortarolo et al.*, 2013; *Anav et al.*, 2013; *Forkel et al.*, 2015; *Peano et al.*, 2019). This study aims to evaluate the ability and limits of the novel CRESCENDO LSMs to represent the global climatology of start and end of growing season timings. The CRESCENDO LSMs cover a wide range of phenology schemes and vegetation descriptions. This selection may therefore help understand the sources of differences between LSMs' representation of phenology and the regions where plant phenology simulations remain difficult.

Vegetation phenology can be assessed by considering different plant features and indices, such as leaves colour (e.g. normalized difference vegetation index, NDVI, *Churkina et al.*, 2005; *Keenan et al.*, 2014), the fraction of absorbed photosynthetically active radiation (e.g. *Kelley et al.*, 2013; *Forkel et al.*, 2015), or canopy density (e.g. LAI *Murray-Tortarolo et al.*, 2013; *Peano et al.*, 2019). Each methodology presents skills and limitations (e.g *Forkel et al.*, 2015). In this work, the novel Four Growing Season Types (4GST) methodology developed by *Peano et al.* (2019) is used to evaluate phenology. This method proved good skill in capturing the principal global phenology cycles (*Peano et al.*, 2019), and integrates a broader spectrum of growing season modes compared to previous techniques (e.g. *Murray-Tortarolo et al.*, 2013). The set of growing season modes investigated in 4GST are (1) evergreen phenology; (2) single growing season with summer LAI peak; (3) single growing season with summer dormancy; and (4) two growing seasons. 4GST uses LAI data to evaluate phenology. Most ecosystem and climate models introduce 'leaf area' as a fundamental state parameter describing the interactions between the biosphere and the atmosphere. The most common measure of the area of leaves is the LAI, which is generally defined as the one-sided leaf surface area divided by the ground area in m^2/m^2 (*Chen and Black*, 1992). In addition, LAI is the key variable by which LSMs scale-up leaf-level processes to canopy and ecosystem scale exchanges of carbon, energy, and water. This makes the LAI a reasonable choice for the evaluation of the LSMs' phenology (*Murray-Tortarolo et al.*, 2013; *Peano et al.*, 2019).

In this paper, we present a brief description of the methods, LSMs and satellite data used (section 2). Next, we present the main results of the satellite data comparison and evaluation of LSMs against observations (section 3). Finally, we discuss the methodology, data, and results (section 4), followed by concluding remarks (section 5).

2 Method, models and data

2.1 Satellite observation

To perform a comprehensive global phenology evaluation, a satellite-based observational dataset is required. LAI satellite observations present uncertainties and limitations related to the assumptions and algorithms applied in the LAI calculation (Section 4.2 e.g. *Fang et al.*, 2013; *Forkel et al.*, 2015; *Jiang et al.*, 2017). For this reason, three different satellite observational products are considered in this work:

1. the full time series of LAI3g data is generated by an artificial neural network (ANN) algorithm that is trained with the overlapping data of the third generation Global Inventory Modeling and Mapping Studies (GIMMS) NDVI3g and Terra Moderate Resolution Imaging Spectroradiometer (MODIS) LAI products (see *Zhu et al.*, 2013). It covers the 1982-2011 period with a 15-day temporal frequency and a 1/12° spatial resolution;

2. The MODIS (MOD15A2H and MYD15A2H version 6, https://lpdaac.usgs.gov/ ; *Myneni et al.*, 2015a, b) LAI algorithm is based on a three-dimensional radiative transfer equation that links surface spectral bi-directional reflectance factors to vegetation canopy structural parameters (see *Yan et al.*, 2016a). It covers the 2000-2017 period with a 4- or 8-day temporal frequency and a 500 m spatial resolution;

3. The Copernicus Global Land Service (CGLS, LAI 1km version 2, https://land.copernicus.eu/global/products/LAI ; *Maisongrande et al.*, 2004; *Drusch et al.*, 2012) LAI dataset is obtained through a neural network applied on top-of-atmosphere input reflectances in red and near-infrared bands derived from SPOT and PROBA-V. The instantaneous LAI estimates obtained in this way go through a temporal smoothing and small gap filling, which discriminate between evergreen broadleaf forest and no-evergreen broadleaf forest pixels (see *Verger et al.*, 2019; *Verger et al.*, 2011). It covers the 1999-2019 period with a 10-day temporal frequency and a 1km spatial resolution. Note that CGLS has a reduced latitudinal coverage compared to MODIS and LAI3g since it covers up to 75°N versus the 90°N of the other two products.

The 2000-2011 period is common to the three satellite datasets and it is used in the present analysis. The satellite observations are aggregated into monthly values and regridded, by means of a first order conservative remapping scheme (*Jones*, 1999), to a regular 0.5° x 0.5° grid for consistency with the LSMs' output. The regridding operation is directly applied to the gap-filled satellite data. Note that regridding does not employ any specific treatment for differences in land cover.

To perform biome-level analysis, the observed ESA CCI land cover map (https://www.esa-landcover-cci.org/) has been used to define a standard regional vegetation distribution. In particular, *Li et al.* (2018) aggregated the original 37 ESA-CCI land cover classes into half-degree spatial resolution and translate them into 14 Plant Functional Types (PFTs) based on an adjusted cross-walking table. These data have been used to obtain an observed dominant PFT map for the 2000-2011 period. Based on *Li et al.* (2018), all vegetation types are classified into ten categories: Broad-leaf Evergreen Trees (BET); Broad-leaf Deciduous Trees (BDT); Needle-leaf Evergreen Trees (NET); Needle-leaf Deciduous Trees (NDT); Broad-leaf Evergreen

Shrubs (BES); Broad-leaf Deciduous Shrubs (BDS); Needle-leaf Evergreen Shrubs (NES); Needle-leaf Deciduous Shrubs (NDS); grass covered areas (Grass); crop covered areas (Crop).

2.2 Land surface models

Seven European LSMs, which are part of the CRESCENDO project, are evaluated in this study. Further details on each of these LSMs are provided in the following sections, and briefly summarized in Table 1.

2.2.1 Community Land Model version 4.5

The Community Land Model (CLM) is the terrestrial component of the Community Earth System Model version 1.2 (CESM1.2, http://www.cesm.ucar.edu/models/cesm1.2/). In its version 4.5 (CLM4.5, *Oleson et al.*, 2013) and biogeochemical configuration (i.e. BGC compset, *Koven et al.*, 2013), it is the land component of the CMCC coupled model version 2 (CMCC-CM2, *Cherchi et al.*, 2019). CLM4.5-BGC features fifteen PFTs, in which crop is represented as a generic C3 crop. The PFTs time-evolution follows the area changes described in the Land Use Harmonization version 2 (LUH2, *Hurtt et al.*, 2020). CLM4.5-BGC resolves explicitly plant phenology (*Oleson et al.*, 2013; *Koven et al.*, 2013), which is described by means of three specific parameterization: (1) evergreen plant phenology; (2) seasonal-deciduous plant phenology; (3) stress-deciduous plant phenology (*Oleson et al.*, 2013).

The CLM4.5 representation of phenology is based on soil temperature, soil moisture, and day-length. In particular, the leaf onset in the seasonal-deciduous plant phenology starts when the soil temperature of accumulated growing-degree-day (GDD) passes a critical threshold. The leaf litterfall, instead, starts when the day-length exceeds another specific threshold (*Oleson et al.*, 2013). In the stress-deciduous plant phenology, soil moisture and soil temperature drive the start and end of the growing season. For example, the leaf onset is soil moisture-driven in areas characterized by year-round warm conditions. Finally, the evergreen plant phenology is characterized by a background litterfall, which is a continuous leaf fall and fine roots turnover distributed throughout the year. A PFT-specific leaf longevity parameter drives this mechanism. Further details can be found in *Oleson et al.* (2013).

2.2.2 Community Land Model version 5.0

CLM version 5.0 (CLM5.0) is the terrestrial component of the Community Earth System Model version 2 (CESM2, http://www.cesm.ucar.edu/models/cesm2/) and of the Norwegian Earth System Model version 2 (NorESM2, *Seland et al.*, 2020).

CLM5.0 uses the same number of default PFTs, while the crop module uses two C3 crop configurations: C3 rainfed and C3 irrigated. The irrigation area is based on crop type and region, and the irrigation triggers for crop phenology are newly updated from the CLM4.5.

CLM5.0 uses the same three specific plant phenology parameterization applied in CLM4.5 (*Lawrence et al.*, 2018). Differently from CLM4.5, CLM5.0 include also precipitation in the stress-deciduous phenology scheme. In particular, antecedent

rain is required to start leaf onset, this is done to reduce the occurrence of anomalous green-up during the dry season driven by upwards water movement from wet to dry soil layers (*Dahlin et al.*, 2015).

Several major changes have been made in the CLM5.0. One of the physiological changes includes maximum stomatal conductance, which now uses the Medlyn conductance model (*Medlyn et al.*, 2011), rather than the previously used Ball-Berry stomatal conductance model. In CLM5.0, the *Jackson et al.* (1996) rooting profiles are used for both water and carbon, where the rooting depths were increased for broadleaf evergreen and broadleaf deciduous tropical tree PFTs. These features impact on soil moisture and plant hydrology that control stress-deciduous plant phenology. Other modifications that might influence phenology include nutrient dynamics, hydrological and snow parameterizations, plant hydraulic functions, revised nitrogen cycling with flexible leaf stoichiometry, leaf N optimization for photosynthesis, and carbon costs for plant nitrogen uptake (*Lawrence et al.*, 2019).

2.2.3 JULES-ES

JULES-ES is the Earth System configuration of the Joint-UK Land Environment Simulator (JULES). JULES-ES is the terrestrial component of the new UK community ESM, UKESM1 (*Sellar et al.*, 2020). It is based on the core physical land configuration of JULES (JULES-GL7) as described in *Wiltshire et al.* (2020a). The simulations described here used a near-final configuration of JULES-ES prior to the final tuning performed as part of UKESM1 (*Yool et al.*, 2020). JULES-ES is run offline forced by global historic meteorological data as described in section 2.3.

JULES-ES includes a full carbon and nitrogen cycle with dynamic vegetation (*Wiltshire et al.*, 2020b), 13 Plant Functional Types with trait based physiology (*Harper et al.*, 2016), and a representation of crop harvest and land use change. In JULES-ES, the allometrically defined maximum LAI varies with the carbon status (*Clark et al.*, 2011) and extent of the underlying vegetation. In the case of natural grasses, maximum LAI can vary rapidly sub-seasonally whereas tree PFTs have a smaller variation. Phenology operates on top of this variation for Deciduous Broadleaf and Needleleaf PFTs based on an accumulated thermal time model. Consequently, JULES-ES features one phenology scheme, which relies on thermal conditions.

2.2.4 JSBACH

JSBACH3.2 is the land component of the MPI-ESM1.2 (*Mauritsen et al.*, 2019). For the simulations described here, JSBACH3.2 is run offline at T63 ($\sim$1.9°) resolution. Simulations were conducted without natural changes in the land cover, instead, a static map of natural land cover based on *Pongratz et al.* (2008) was used. Anthropogenic land cover changes were applied using land-use transitions (see *Reick et al.*, 2013) derived from the LUH2 forcing, whereby rangelands were treated as natural vegetation (see also *Mauritsen et al.*, 2019). Simulations were conducted according to the common CRESCENDO protocol as described in section 2.3, with the only difference that land-use change was already simulated starting 1700 to avoid a cold start problem when applying land-use transitions. JSBACH3.2 contains a multilayer hydrology model (*Hagemann and Stacke*, 2015) and a representation of the terrestrial nitrogen cycle (*Goll et al.*, 2017).

JSBACH3.2 is run with its default phenology model, called LoGro-P, as evaluated in *Böttcher et al.* (2016) and *Dalmonech et al.* (2015). This phenology is based on a logistic equation for the temporal development of the LAI. Under ideal environ-

mental conditions, the LAI approaches a maximum value representing a prescribed PFT specific physiological limit. Growth and leaf shedding rates of the logistic equation are functions of environmental conditions, chosen differently according to the phenology type (see e.g. *Böttcher et al.*, 2016). JSBACH3.2 distinguishes the following phenology types: (1) evergreen, (2) summergreen, (3) raingreen, (4) grasses, as well as (5) tropical and extra-tropical crops.

In general, JSBACH3.2 features a higher amount of phenology schemes (i.e. five) compared to the other LSMs, which are driven by soil temperature, air temperature, soil moisture, and net primary productivity (NPP). In particular, the phase changes in evergreen and summergreen phenologies are determined by temperature thresholds calculated by the alternating model of *Murray et al.* (1989) from heat sums, chill days, and critical soil temperatures. The raingreen phenology has a non-zero growth rate whenever the soil moisture exceeds the wilting point and the NPP is positive. The shedding rate depends on the relative soil water content. The grass phenology resembles the raingreen phenology but further requires the air temperature and soil moisture to exceed a critical value for a non-zero growth rate. Because grass roots are less deep than tree roots, the soil moisture is taken from the upper soil layer for the grass phenology. The crop phenology is modeled as a function of NPP and distinguishes tropical and extra-tropical crops in order to reflect different farming practices in dependence of the prevailing climatic conditions. The vegetation in the conducted JSBACH3.2 simulations was represented by 12 PFTs, each of which is linked to one of the phenology types: one forest type with evergreen phenology, one forest and one shrub type with summergreen phenology, two forest and one shrub type with raingreen phenology, C3 and C4 grasses as well as C3 and C4 pastures with grass phenology, and C3 and C4 crops with extra-tropical and tropical crop phenology, respectively.

2.2.5 LPJ-GUESS

The Lund-Potsdam-Jena General Ecosystem Simulator version 4.0 (LPJ-GUESS; *Lindeskog et al.*, 2013; *Smith et al.*, 2014; *Olin et al.*, 2015), a process-based 2nd generation dynamic vegetation and biogeochemistry model, is the terrestrial biosphere component used in the European community Earth-System Model (EC-Earth-Veg, http://www.ec-earth.org/; *Hazeleger and Bintanja*, 2012; *Döscher et al.*, in review; *Miller et al.*, in prep.). It simulates vegetation dynamics, land use and land management following LUH2 (*Hurtt et al.*, 2020). LPJ-GUESS features 25 PFTs, ten woody and two herbaceous PFTs compete in the natural stand fractions, whereas two herbaceous species, C3 and C4 photosynthesis-pathways, compete in pasture, urban and peatland fractions. Crop stands have each five crop functional types representing the properties of global crop types and corresponds to the classes found in LUH2, namely both annual and perennial C3 and C4 crops, and C3 N fixers, and two herbaceous cover crops (C3 and C4) that are grown in-between the main agricultural growing seasons.

Similar to CLM4.5 and CLM5.0, LPJ-GUESS plant phenology is described by means of three specific parameterization: (1) evergreen plant phenology; (2) seasonal-deciduous plant phenology; (3) stress-deciduous plant phenology (*Smith et al.*, 2014). An explicit phenological cycle is simulated only for leaves and fine roots in seasonal-deciduous and stress-deciduous PFTs, whereas evergreen PFTs have a prescribed background litterfall for leaves, fine roots and sapwood. Seasonal-deciduous plant phenology is based on a PFT-dependent accumulated GDD sum threshold for leaf onset, with leaf area rising from 0 to the pre-determined annual maximum leaf area linearly with an additional 200 (100 for herbaceous and needleleaved tree PFTs) degree days above a threshold of 5 °C. For seasonal-deciduous PFTs, growing season length is fixed, all leaves being

shed after the equivalent of 210 days with full leaf cover. Stress-deciduous plant phenology PFTs shed their leaves whenever the water stress scalar ω drops below a threshold, $\omega_{\min}$, signifying the onset of a drought period or dry season. New leaves are produced, after a prescribed minimum dormancy period, when ω again rises above $\omega_{\min}$ (*Smith et al.*, 2014). Crop PFT sowing and harvest decisions are modelled based on climate variability (*Waha et al.*, 2011; *Lindeskog et al.*, 2013) and climatic thresholds (*Bondeau et al.*, 2007).

2.2.6 ORCHIDEE

The ORCHIDEE model used for the CRESCENDO simulations is the land component of the IPSL (Institut Pierre Simon Laplace) ESM used for the CMIP6 simulations (*Boucher et al.*, 2020). The surface heterogeneity is described with 15 different PFTs that can be mixed in each grid cell. The annual evolution of the PFT distribution is derived from the LUH2 database as described in more detail in *Lurton et al.* (2019). In each grid cell, the PFTs are grouped into three soil tiles according to their physiological behavior: high vegetation (forests) with eight PFTs, low vegetation (grasses and crops) with 6 PFTs, and bare soil with one PFT. An independent hydrological budget is calculated for each soil tile, to prevent forests from exhausting all soil moisture. In contrast, only one energy budget (and snow budget) is calculated for the whole grid cell. Note that since its first description in *Krinner et al.* (2005), the model has substantially evolved; we describe below only the main features relevant for this study.

A Phenology module describes leaf onset and leaf senescence for deciduous PFTs based on temperature and soil moisture. In temperate and boreal regions, leaf onset is driven by an accumulation of warm temperature in spring, following the concept of GDD. In addition, a minimum period of cold temperature, based on a Number of Chilling Days (NCD), is used to avoid buds dying with late frosts. Both criteria are combined, with PFT-specific GDD and NCD thresholds to be met, before leaves can start growing (see *Botta et al.* (2000) for more details). For the dry tropics and semi-arid ecosystems, a moisture availability criteria is used based on water accumulated in the soil. A minimum of 5 consecutive days with soil moisture increase (root zone) should occur after the 1 January for northern hemisphere and the 1 July for southern hemisphere, with the addition of a filter for small rises in soil moisture (see model 4b in *Botta et al.* (2000)). Both temperature and moisture criteria are combined for grasses and crops, and the different parameters of the leaf onset parameterisation have been calibrated with satellite data (*Botta et al.*, 2000). Leaves are then further separated into four age classes with different photosynthetic efficiency. Leaf fall is controlled by different turnover processes. The first one is a simple aging process and a second senescence process based on climatic conditions (either based on air temperature or soil moisture conditions) is applied.

2.2.7 ISBA-CTRIP

ISBA-CTRIP is the land surface model of CNRM-ESM2-1 (http://www.umr-cnrm.fr/spip.php?article1092). It is used within the SURFEX version 8 modeling platform representing SURFace EXchanges between ocean, lakes, and land. It solves the energy, carbon and water budgets at the land surface and was recently thoroughly upgraded (*Decharme et al.*, 2019). The model distinguishes 16 vegetation types (9 tree, 1 shrub, 3 grass, and 3 crop types) alongside desert, rocks, and permanent snow. *Decharme et al.* (2019) give a detailed description of the physical aspects of the model.

Differently from the other LSMs, leaf phenology results directly from the daily carbon balance of the leaves. Leaf turnover time is dependent on potential leaf longevity reduced when 10-day assimilation rates start to decrease. Leaf area index is diagnosed from leaf biomass and specific leaf area index, which varies as a function of leaf nitrogen concentration and plant functional type. To allow for leaf growth after dormancy there is an imposed minimum leaf biomass. Crops have the same phenology as grasses. A detailed description of the terrestrial carbon cycle can be found in *Delire et al.* (2020).

2.3 Experimental setup

In this study, the S3 CRESCENDO simulations were used, characterized by transient CO_2, climate, and land-use forcing. Each model spin-up is obtained by recycling climate mean and variability from the period 1901-1920, with the pre-industrial (1860) atmospheric CO_2 concentration until carbon pools and fluxes reach a steady state. The 1861-1900 period is simulated using the same climate forcing as the spin-up, but with time-varying atmospheric CO_2 and land-use forcing. Finally, the LSMs are forced over the 1901-2014 period with changing CO2, climate, and land-use forcing. All LSMs are commonly driven by the atmospheric forcing reanalysis CRUNCEP version 7 (*Viovy*, 2018), and the land-use data is taken from the Land Use Harmonization version 2 (*Hurtt et al.*, 2020). Note that the use of LUH2 land cover transitions differs across the models (see model description). CRUNCEPv7 provides for 2m air temperature, precipitation, wind, surface pressure, shortwave radiation, long-wave radiation, and air humidity.

Each LSM is run on different spatial resolutions (Table 1), but the outputs of these simulations are provided on a regular 0.5° x 0.5° grid, over which simulations and observations are compared. CLM4.5, JULES-ES, JSBACH, and ISBS-CTRIP perform their simulations at a coarser resolution. Their output are regridded by applying a first-order conservative remapping method (*Jones*, 1999). The LAI monthly mean output from these simulations are used in the present analysis.

2.4 Growing season analysis

The times of start and end of growing season (GSS and GSE, respectively) are evaluated using the Four Growing Season Types (4GST) method introduced by *Peano et al.* (2019). 4GST has been shown to adequately capture the main global phenology cycles for evaluation of LSMs.

The 4GST method allows to evaluate start and end of the growing season and the global spatial distribution of four main growing season types: (1) evergreen (EVG); (2) single growing season peaking in summer (SGS-S); (3) single growing season with summer dormancy (SGS-D); (4) two growing seasons (TGS). The EVG type is identified when relative changes in LAI annual cycle are smaller than 25% of local LAI mean value. Note that GSS and GSE timings are not detected in EVG areas. The other three types are identified based on LAI slopes and transition timings as illustrated and summarized in Figure 1. When one single growing season is identified, SGS-S and SGS-D are discerned based on the peak-month (i.e. in the Northern Hemisphere (NH) SGS-S is detected when LAI peak occurs between April and September, otherwise, SGS-D is detected, vice versa in the Southern Hemisphere (SH)). The timings of start and end of the growing season are identified through a critical threshold set to 20% of the annual LAI cycle (Figure 1). TGS, instead, is identified when two growing seasons at least three-month-long are detected and GSS and GSE timings are identified for each cycle. Further details can be found in *Peano et al.* (2019). Note that

in this analysis, the timings of the TGS GSS correspond to the GSS timings of the first growing season cycle, while the GSE are the second GSE timings, as described in *Peano et al.* (2019). The 4GST method is applied on monthly LAI data in this work, instead of 15-day time-scale used in *Peano et al.* (2019).

3 Results

3.1 Satellite data comparison

We inspect the main differences between LAI3g, MODIS and CGLS by plotting the spatial distribution of the four growing season types, GSS, and GSE (Figure 2).

The three products show a high consistency in the distribution of growing season types (agreement of about 80%, Table 2), with the main differences occurring in tropical regions, such as in Amazon and Congo basins, and in semi-arid areas, such as central Australia (Figures 2a,d,g). Compared to MODIS, LAI3g differs mainly in EVG regions (Table 2) due to an underestimation of EVG areas in the Tropics (Supplementary Figure 2). These regions are characterized by high canopy density, which saturates to high LAI in the satellite data (e.g. *Myneni et al.*, 2002), resulting in limited seasonal variability. In addition, the AVHRR sensor used to derive LAI3g is less responsive to changes in vegetation compared to MODIS and SPOT/PROBA data (*Piao et al.*, 2020). Both LAI3g and CGLS differ from MODIS in areas featured by the TGS type (Table 2). The Horn of Africa is the only region where all three satellite products place a TGS type (Figure 2).

Larger differences among satellite products are found in the assessment of GSS and GSE (Figure 2), especially in the NH where LAI3g and CGLS clearly anticipate GSS (Figures 2e,h) with respect to MODIS. The three satellite products present a consistency similar to the one reached by the growing season type distribution (about 75%) when a one-month tolerance is considered (Table 3), since time-resolution of the products has been homogeneizied to one month (see Section 2.4).

Keeping these differences in mind, the MODIS data are used as a graphical reference in Figures 3, 4, and 5. These figures keep track of the agreement among satellite data despite the choice of MODIS as reference. Figures using CGLS and LAI3g as a graphical reference are presented in the supplementary material.

3.2 Growing season types distribution

The 4GST allows estimating the ability of each LSM in capturing the observed spatial distribution of the four growing season types (Figure 3). In general, all the LSMs capture the single growing season that peaks in summer (SGS-S type) reasonably well, especially in the NH mid- and high-latitude regions. The majority of LSMs are also able to correctly represent the two growing seasons (TGS) in the Horn of Africa region (Figure 3). Most LSMs are unable to reproduce the observed growing-season-type distribution in the SH, except for the evergreen (EVG) tropical areas. A partial exception is LPJ-GUESS, which shows large SGS-S type areas in South America, Southern Africa and Northern Australia, in agreement with the satellite products (Figure 3f). The high number of PFTs used by LPJ-GUESS can be the source of this skill (Table 1).

LSMs used in this study are primarily able to capture the observed EVG and SGS-S regions with agreement between 36.0% and 95.4%, and between 44.3% and 79.5%, respectively (Table 2). In contrast, the TGS regions are seldom reproduced by LSMs, and the agreement rate with MODIS ranges between 0.4% and 19.1%, (Table 2). Overall, the CRESCENDO Multi-Model Ensemble (MME) mean reproduce the same MODIS growing season type distribution over about 69.5% of global land-surface area, with a 45.4% to 74.0% range among models (Figure 3 and Table 2). It is noteworthy that the evergreen type is correctly detected in the broad-leaf evergreen tropical areas in both satellite observations and LSMs (Figure 3). On the contrary, the high-latitude needle-leaf evergreen regions are partially represented in LSMs, while satellite data do not catch these areas due to satellite limitations resulting from the impact of cloud and snow cover on light availability during the winter season (see Section 4.2). Besides, the variability of understory and secondary PFTs may influence LAI seasonality representation.

This initial evaluation highlights that LSMs have difficulties in accurately representing SH phenology. The correct location of the less common types, i.e. single growing season with summer dormancy (SGS-D) and TGS, is as well hardly captured by the LSMs. Similar results are obtained when CGLS and LAI3g satellite observations are used as references (Supplementary Figures 2,3 and Supplementary Tables 1,2).

3.3 Variability of growing season start and end

4GST is then applied to evaluate the ability of LSMs to represent the GSS and GSE timing in vegetated areas not classified as EVG-type (white regions in Figures 4 and 5 correspond to not-vegetated and EVG-type domains).

On average at the global scale, LSMs approximately exhibit a disagreement of 0.6 months and 0.5 months in GSS and GSE, respectively, with LSMs simulating a later GSS and an earlier GSE, practically shortening the growing season by one month (Table 4). This bias is not evenly distributed around the globe. LSMs reproduce the correct growing season length in about 17% of the global land grid-cell, but sometimes growing season is affected by a shift in seasonality, as in the case of JULES-ES (Table 3). Differently from the other LSMs, the LAI cycle in JULES-ES starts from a climatological condition (*Wiltshire et al.*, 2020a), which can lead to the detected shift.

Generally, the GSE simulated by the LSMs show a better agreement with MODIS (about 25% agreement in vegetated grid-cell, ranging from 4.9% to 26.4%, Table 3) compared to GSS timings (15.8% agreement in vegetated land grid-cell, ranging from 2.7% to 19.1%, Table 3). Considering a one-month-tolerance to account for the downgraded time-resolution, the agreement between LSMs and MODIS increases to $\sim$45% and $\sim$31%, respectively (Table 3).

LSMs exhibit larger agreement with MODIS GSS and GSE timings in the NH compare to the SH (Figures 4, 5 and Tables 3, 4). Only CLM 5.0 and LPJ-GUESS show similar results in both hemispheres (Table 3). In particular, LPJ-GUESS shows good skill (agreement with observation larger than 15%) in capturing both GSS and GSE timings in both hemispheres (Figures 3f, 4f, and 5f).

LPJ-GUESS is the model that shows the highest agreement with MODIS (Table 3) and the lowest bias in average GSS and GSE timings (0.4 and 0.1 months, respectively, Table 4). JULES-ES shows the lowest agreement with MODIS (Table 3) and the highest bias in the average GSS and GSE timings (1.2 and -2.3, respectively, Table 4). This result may be associated with

the representation of PFTs in the two models used to describe global vegetation. LPJ-GUESS, indeed, is the model featuring the largest number of PFTs, while JULES-ES uses the least (Table 1). Moreover, JULES-ES and LPJ-GUESS differ also on the details of the phenology parameterization. LPJ-GUESS features three phenology schemes driven by temperature and soil moisture versus one parameterization only based on the temperature in JULES-ES (Sections 2.2.3, 2.2.5 and Table 1). Similar to JULES-ES, JSBACH presents a small number of PFTs, but it reaches better results thanks to the five implemented phenology schemes (Section 2.2.4 and Table 1).

The two Community Land Model versions (i.e. CLM4.5, and CLM 5.0, Table 3) show very different outcomes, with CLM5.0 exhibiting larger biases in GSS and GSE timings compared to CLM4.5 (Figures 4b,c, and 5b,c, and Table 4). The two model versions differ in the crop representation, plant physiology, and phenology parameterization (Section 2.2 and Table 1). The implementation of an antecedent rain requirement trigger for stress deciduous PFTs (*Dahlin et al.*, 2015) help improved phenology in semi-arid regions (e.g. the Sub-Sahara, Figures 4b,c, and 5b,c). Nonetheless, *Zhang et al.* (2019) show that the same upgrade influences the leaf senescence in temperate grasslands. On the other hand, the irrigation scheme in the CLM5.0 crop model allows for the improvement in crop-dominated regions, such as the Indian peninsula (Figures 4b,c, and 5b,c). Further differences occur between CLM 4.5 and CLM 5.0 (Figure 4b,c, and 5b,c), which could be ascribed to the changes in plant physiology, soil hydrology, and rooting profile. For example, CLM5.0 applies a different rooting profile scheme and soil moisture threshold (Table 1) affecting the representation of the soil moisture impact on phenology.

CGLS and LAI3g support the results obtained with MODIS in the NH mid-latitude, Africa, and Brazil (shaded cross pattern in Figures 4, 5). Only LAI3g supports MODIS outcomes in the NH high-latitude (shaded slash pattern in Figures 4, 5). In general, the direct comparison of LSMs with LAI3g and CGLS satellite observations exhibits results following MODIS ones (Supplementary Figures 5-8, and Supplementary Tables 3-6).

3.4 Latitudinal variability

The MME zonal average shown in Figure 6 highlights the LSMs' abilities and limitations in simulating the observed GSS and GSE timings at different latitudes. The GSS bias ranges between -1.8 months (earlier GSS) just south of the Equator and +2.0 months (delayed GSS) south of 50°S (Figure 6a). The GSE bias ranges between -3.0 months in the 0-10°N latitudinal band and +1.3 months in the southern sub-tropics. The CRESCENDO LSMs correctly simulate the GSE timings north of 60 °N. The Spearman correlation of the GSS and GSE latitudinal distributions is 0.67 ± 0.07 and 0.51 ± 0.11, respectively. These values are significant at the 95% level based on a Monte Carlo approach.

In the NH mid- and high-latitude, the LSMs' GSS exhibit an average delay of up to 1.6 months, especially in North America (Supplementary Figure 9a). This bias and the spread among LSMs might be driven by differences in temperature schemes and thresholds used by LSMs (Table 1). Note that also differences among satellite data occur in the NH mid- and high-latitude, highlighting potential differences among these three products (see Section 4.2). Large LSM biases in NH tropical region GSE timings and southern sub-tropical GSS timings are driven by premature values in Africa (Supplementary Figures 9c,d). These discrepancies may derive from difficulties in the LSM's ability to simulate the observed phenology type and the response to soil moisture in Africa (Figure 3, and Supplementary Figure 10). Large variability is spotted in the region below 40°S. The

reduced amount of vegetated land area may cause this behaviour. A different growing season type detection in this area, such as a different size of the evergreen region (Figure 2), may, indeed, extensively influence the GSS and GSE detection, which is the case for the satellite products (Figure 6), especially LAI3g.

Observed latitudinal distributions highlight an increasing northward trend in the NH mid-latitude GSE timings (GSE around May-June at ∼20°N and around September-October at ∼40°N, Figure 6b), and an increasing southward trend in the 30-55°S latitudinal band (GSS around July at ∼30°S and around September at ∼55°S, Figure 6a). Similar trends are reproduced by the LSMs, but with a higher magnitude (Figure 6). In the NH, the difference between simulated and observed trends may be driven by an overestimated influence of radiation and temperature on leaf senescence in LSMs. In the SH, the discrepancies between observed and modeled trends may be related to relatively large phenology variability in the SH associated with the small vegetated land area in this hemisphere.

3.5 Regional variability

To assess sources of biases in the LSMs, different biomes derived from the ESA CCI land cover map (*Li et al.*, 2018, Figure 7a) are investigated. The GSS timings are generally delayed compared to observations, except for the Broadleaf Evergreen Trees (BET) and Broadleaf Deciduous Shrub (BDS) biomes (Figures 7f,k). In BDS-dominated regions, the Multi-Model Ensemble mean (MME) falls within the observational range (Figures 7f,k), but a large spread among LSMs exists. The BDS-dominated regions are semi-arid and transition areas, where LSMs' parameterization could be more sensitive to climate conditions and parameter selection, especially, soil moisture. The large spread among LSMs, then, might be mostly linked to the differences in the implementation of soil moisture in the phenology schemes (Table 1). It is noteworthy that this biome covers a small fraction of the global vegetated regions. The largest biome (i.e. Grass in the North Hemisphere, Figure 7g), instead, exhibits a mean delay of one month, which is common among the LSMs except for LPJ-GUESS which falls within observed range. Besides, large biome variability is visible in the SH crop biome (Figure 7m). In general, LSMs show a larger variability in the South Hemisphere (SH) compared to the North Hemisphere (NH).

GSE timings display heterogeneous outcomes (Figure 8). In general, a larger variability is observed compared to GSS timings. The NH Grass biome, which covers about 33% of the global vegetated area, exhibits a mean delay of one month which is mainly driven by JULES-ES (Figure 8g). The SH BDS area displays a large variability among models (Figure 8k) ranging from May (LPJ-GUESS) to November (JULES-ES). Large biome variability appears in Broadleaf Evergreen Tree (BET), Grass and Crop SH biomes (Figures 8i,l,m) and NH Crop (Figure 8h). This result highlights the need for further investigation on the representation of crop phenology in the LSMs since only a few LSMs (i.e. JSBACH and ORCHIDEE) treat crops with a specific parameterization (Section 2).

In general, LSMs show a higher agreement in representing GSS timings compared to GSE timings. Consequently, the different approaches used to describe the start of the growing season are relatively consistent among LSMs. In comparison, the representation of the end of the vegetative season requires further investigation and development. Note that this regional evaluation is performed based on the observed biome distribution (i.e. ESA-CCI map). However, each LSM treats differently

the land cover and biome distribution (Section 2). For this reason, part of the obtained spread among LSMs derives from differences in PFTs representation and distribution (Section 2.2, and Table 1), which affect phenology representation in LSMs.

4 Discussion

4.1 Land surface models

The plant phenology growing season start and end are mainly triggered by changes in solar radiation, temperature, and soil moisture conditions (e.g. *Caldararu et al.*, 2012; *Zeng et al.*, 2013; *Tang and Dubayah*, 2017). State-of-the-art LSMs represent the phenological transitions using different parameterizations based on the climate conditions (Section 2.2). Many of these parameterizations (see Section 2.2) are based on values derived from localized observations (e.g. *White et al.*, 1997; *Thornton et al.*, 2002; *Jolly et al.*, 2005; *Savoy and Mackay*, 2015). Consequently, the phenology parameters are calibrated on specific regions of the globe, which may be one reason for the large spread of values seen in the present analysis.

Generally, phenology calibration areas are located in the NH, where LSMs exhibit better results and larger coherence compared to the SH. Among the LSMs evaluated here, LPJ-GUESS, CLM4.5, and ORCHIDEE show good skill (agreement with observation larger than 15%) in the SH (Table 3). On the other hand, CLM5.0 and JULES-ES do not reach such agreement in the NH (Table 3). High skill (agreement with observation larger than 20% for at least one timing) in the NH are obtained by CLM4.5, ORCHIDEE, and ISBA-CTRIP (Table 3). The different performance between models can occur from differences in phenology parameterization as well as different vegetation cover types (Plant and Crop Functional types), soil characterization, and initial spatial resolution (Table 1).

Among the LSMs evaluated here, JULES-ES shows relatively lower skill in simulating GSS and GSE timings compared to the other LSMs (Table 3). This result may be ascribed to the smaller number of PFTs (see Table 1) and details of the phenology parameterization that characterize this LSM (Section 2.2.3 and Table 1). JSBACH accounts for a similar number of PFTs (Table 1) but features a more complex phenology scheme (Section 2.2.4 and Table 1). For this reason, JSBACH exhibits a higher skill than JULES-ES in reproducing GSS and GSE timings (Table 3).

Similar to JSBACH, ORCHIDEE feature a PFT-oriented phenology scheme (Section 2.2.6 and Table 1), which contributes to the high skill noted for ORCHIDEE.

CLM 4.5, CLM 5.0, and LPJ-GUESS use three phenology schemes: (1) evergreen; (2) seasonal-deciduous; (3) stress-deciduous (Sections 2.2.1, 2.2.2, 2.2.5). Among these schemes, the seasonal-deciduous one employs calendar thresholds (summer and winter solstices and day-length threshold in CLM, and fixed 210-day phenology in LPJ-GUESS) that may improve the results of LPJ-GUESS and CLM 4.5. On the other hand, this may mean that the seasonal-deciduous type may be less responsive to future climate change.

Contrary to the other LSMs, ISBA-CTRIP uses the daily leaf carbon balance to simulate plant phenology, and it reaches good skill (Tables 3, 4). Consequently, ISBA-CTRIP highlights the opportunity to attain results aligned with the other LSMs using leaf carbon availability instead of climatic conditions.

The improvement of the phenology parameterization can lead to better representation of vegetation in the LSMs. However, other vegetation features affect the plant phenology representation, as in the case of the two CLM versions. CLM4.5 and CLM5.0 share similar phenology parameterization (Sections 2.2.1 and 2.2.2) but differ in the crop irrigation scheme, soil and plant hydrology, and carbon and nitrogen cycling (*Lawrence et al.*, 2018). Since soil moisture has a significant control on plant phenology (e.g. *Caldararu et al.*, 2012), CLM5.0 revision of stomatal response to rising CO_2 concentrations through a new Medlyn stomatal conductance scheme (*Fisher et al.*, 2019; *Medlyn et al.*, 2011) and the use of a revised mechanistically based soil evaporation parameterization that accounts for the rate of diffusion of water vapor through a dry surface layer (*Swenson and Lawrence*, 2014) are likely to be principal sources of differences between CLM5.0 and CLM4.5.

In general, this comparison highlights the complexity of vegetation phenology modelling and the strong inter-linkages between climate, hydrology, soil, and plants.

4.2 Satellite data

Satellite-based LAI datasets have been used in this work as a benchmark for the evaluation of the LSMs' phenology performance globally. However, satellite observations present some caveats and uncertainties (e.g. *Myneni et al.*, 2002; *Fang et al.*, 2013; *Jiang et al.*, 2017). For this reason, three separate satellite LAI products obtained from different acquisition sensors (namely AVHRR for LAI3g, MODIS for MODIS LAI, and SPOT/PROBA VEGETATION for CGLS) have been used in this study. The comparison between these datasets shows major issues associated with LAI products derived from satellite reflectance observations. For example, large differences between LAI3g, MODIS, and CGLS occur at high latitudes and in tropical regions (Figure 6), where thick clouds and snow cover can affect the data reconstruction (e.g. *Delbart et al.*, 2006; *Kandasamy et al.*, 2013; *Yan et al.*, 2016b). LAI satellite data are also affected by the applied regridding and gap-filling algorithms, which could create spurious seasonal cycles as well as smooth the observed phenology season (e.g. *Kandasamy et al.*, 2013; *Chen et al.*, 2017). In addition, the observed reflectance saturates in regions characterized by dense canopies reaching prescribed LAI upper limits (e.g. 7.0 m^2/m^2 in MODIS and LAI3g *Myneni et al.*, 2002; *Maignan et al.*, 2011). This issue can affect the identification of growing season cycles in thickly forested areas leading to possible overestimation of evergreen type detection.

4.3 4GST limitations

Contrary to previous phenology analysis that focused on specific biomes (e.g. *Dahlin et al.*, 2015; *Zhang et al.*, 2019) or NH mid- and high-latitude regions (above 30° N *Anav et al.*, 2013; *Murray-Tortarolo et al.*, 2013), 4GST accounts explicitly for different phenology types on the global scale. In particular, it takes into account SGS-D and TGS types that were neglected in previous analyses (*Murray-Tortarolo et al.*, 2013) due to their reduced coverage (Table 2).

Regions with multiple growing seasons per year (TGS) are difficult to capture on a global scale, despite their important influence on climate (e.g. *Zhang et al.*, 2003; *Dalmonech and Zaehle*, 2013; *Peano et al.*, 2019). The state-of-the-art LSMs, indeed, exhibit a low skill in reproducing this specific growing season type (Figure 3 and Table 2). Two growing seasons usually occur in regions characterized by two separate rain seasons, in semi-arid areas or in cropland regions (e.g. *Zhang et al.*,

2003, 2005; *Martiny et al.*, 2006). In this analysis, observations and LSMs, except for JULES-ES and ORCHIDEE, agree on a TGS type only in the Horn of Africa. This region features two distinct precipitation seasons (e.g. *Liebmann et al.*, 2012; *Peano et al.*, 2019), which trigger the TGS phenology type.

Cropland areas can present multi growing season behavior because of irrigation and crop rotation, such as in South Asia, and China (e.g. *Wu et al.*, 2010; *Gumma et al.*, 2016). Unlike the Horn of Africa, these regions are not captured as TGS in the present analysis due to assumptions and limitations within LSMs, for example CLM4.5 represents all annual crops by a generic C3 PFT (Section 2.2.1, and Table 1), and 4GST assumptions. 4GST TGS type detection adopts a minimum length of three months to detect a growing season. This assumption derives from the need to avoid the detection of small oscillations within the same growing seasons (*Peano et al.*, 2019). Consequently, this assumption affects the model recognition of multiple growing seasons, especially in cropland areas. South Asia, for instance, is characterized by different timing and phenology intensity for each crop growing season (*Gumma et al.*, 2016). Therefore, only some specific crops can be detected based on the 4GST assumptions and growing season signature (*Gumma et al.*, 2016). *Wu et al.* (2010) distinguish multiple growing seasons in China using local maximum and detection threshold. This method improves the multi growing season identification, especially for the crops identified by a strong phenology cycle but may exclude crop characterized by a weak phenological cycle. For these reasons, more specific analyses of semi-arid and crop regions based on higher spatial and temporal data are needed and will be the focus of a future study. Besides, LSMs' crop phenology parameterizations require further development to improve the description of each specific crop.

Another limitation of the present evaluation is the monthly temporal frequency. Data at a higher frequency, indeed, might lead to a more detailed bias assessment. The use of a different temporal frequency may also influence phenology type detection. For example, *Peano et al.* (2019), that uses 15-day LAI data, detect a slightly different distribution of CLM4.5 SGS-D and TGS types in Australia, Horn of Africa, and Brazil. Similarly, *Zhang et al.* (2019), which analyses CLM4.5 in Northeast China with 8-day LAI data, obtain TGS type in areas recognized as SGS-S in the present analysis.

LSMs evaluate phenology at PFT level, but the final LAI values are returned at the grid-cell level. For this reason, a more detailed evaluation of the parameterization would require using PFT-level values. However, global coverage of observed PFT level phenology values are missing, making this analysis limited to specific biomes, such as through PhenoCAM data (*Richardson et al.*, 2018). This analysis will be the focus of a future work.

5 Conclusions

This study evaluates the ability of the land component (LSMs) of seven state-of-the-art European Earth System models participating in the CMPI6 to reproduce the timings of start and end of the plant growing season at the global level. The assessment is performed based on the novel four growing season types methodology, and uses a set of three satellite observation products as a benchmark to account for some of the uncertainty in observations.

In general, LSMs exhibit better agreement with observations in the NH compared to the SH, where large variability associated with the small vegetated land area is present. LSMs also show higher ability to simulate the timing of growing season

end compared to the timing of the growing season start. On average, LSMs show 0.6 months delay in estimating start of the growing season and about 0.5 months premature end of the growing season, leading to about one month shorter phenology active season. High discrepancies between LSMs and satellite products are noted for growing season start (GSS) timings in the region poleward of 50°S, where simulated GSS is delayed by about two months. The growing season end (GSE) shows high differences between LSMs and observations in the 0-10°N latitudinal band, where LSMs simulate a three-month earlier GSE. On the contrary, the LSMs accurately simulate the GSE timings poleward of 60 °N and the GSS in the 30-40°S and 10-30°N latitudinal bands. At the biome scale, LSMs correctly simulate the GSS and GSE timings in Broad-leaf Evergreen Trees dominated areas. High intra-model variability remains in the Broad-leaf Deciduous Shrubs and Crop dominated areas.

Despite a lower ability of LSMs to represent SH phenology, LPJ-GUESS, CLM4.5, and ORCHIDEE show reasonably good outcomes in these regions. In the NH, high skill is achieved by CLM4.5, ORCHIDEE, and ISBA-CTRIP. Uncertainties and spread among LSMs remain, which might affect our understanding of present-day and future impact of land and vegetation interactions with the climate and carbon cycle. Therefore, further improvements in LSMs will be necessary.

Improvements in the phenology parameterization can lead to better representation of vegetation in the LSMs. However, phenology in LSMs is influenced by vegetation and hydrological parameterizations and land surface boundary conditions (e.g. PFT distribution), as shown by the CLM4.5 and CLM5.0 phenology differences.

This study highlights the complexity of vegetation phenology modeling and the strong inter-linkages between climate, hydrology, soil, and plants, which need further details and generalization inside the LSMs code.

Code and data availability. The LAI3g satellite observation data are available from R. Myneni (http://sites.bu.edu/ cliveg/datacodes/); the MODIS satellite observation data are available from T. Park; the CGLS satellite observation data are available from COPERNICUS (https://land.copernicus the atmospheric forcing, CRUNCEP v7, are available from N. Viovy (https://rda.ucar.edu/ datasets/ds314.3/); the land surface models simulations are part of CRESCENDO project and they are stored at the CEDA JASMIN service (http://www. ceda.ac.uk/); the 4GST python script is available online (https://github.com/daniele-peano/4GST)

Author contributions. DP wrote the paper, performed the analysis, and provided CLM 4.5 data; DH and SM were strongly involved in the discussion of result and in drafting the manuscript; TP provided MODIS data; DW provided LPJ-GUESS data; YF and HL provided CLM 5.0 data; AW provided JULES-ES data; EJ and CD provided ISBA-CTRIP data; PP provided ORCHIDEE data; JEMSN provided JSBACH data; all co-authors discussed the results and contributed to writing the manuscript. Authors after SM are listed in alphabetic order.

Competing interests. The authors declare no competing interests.

Acknowledgements. We acknowledge the editor and reviewers for their comments that helped in improving this work. This work has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement 641816 (CRESCENDO). DH and AW also received support from the Met Office Hadley Centre Climate Programme (HCCP) funded by BEIS and Defra. DW acknowledges financial support from the Strategic Research Area MERGE (Modeling the Regional and Global Earth System - www.merge.lu.se), and from the Swedish national strategic e-science research program eSSENCE (http://essenceofescience.se/) LPJ-GUESS simulations were performed on the Tetralith supercomputer of the Swedish National Infrastructure for Computing (SNIC) at Linköping University, under project no. SNIC 2018/2-11 (S-CMIP). TP acknowledges support from NASA's Carbon Monitoring System program (80NSSC18K0173-CMS).

References

Anav, A., P. Friedlingstein, M. Kidston, L. Bopp, P. Ciais, P. Cox, C. Jones, M. Jung, R. Myneni, and Z. Zhu (2013), Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models, *J. Clim.*, *26*, 6801–6843, https://doi.org/10.1175/JCLI-D-12-00417.1.

Berdanier, A. B., and J. A. Klein (2011), Growing Season Length and Soil Moisture Interactively Constrain High Elevation Aboveground Net Primary Production *Ecosystems*, *14*, 963–974, https://doi.org/10.1007/s10021-011-9459-1.

Bondeau, A., P. C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, and D. Gerten, D. (2007), Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Glob. Change Biol.*, *13*, 679–706, https://doi.org/10.1111/j.1365-2486.2006.01305.x.

Böttcher, K., T. Markkanen, T. Thum, T. Aalto, M. Aurela, C. Reick, P. Kolari, A. Arslan, and J. Pulliainen (2016), Evaluating biosphere model estimates of the start of the vegetation active season in boreal forests by satellite observations, *Remote Sensing*, *8*, 580, https://doi.org/10.3390/rs8070580.

Botta, A and Viovy, N and Ciais, P and Friedlingstein, P and Monfray, P, A global prognostic scheme of leaf onset using satellite data, *Global Change Biology, 6, 7*, 709–725

Boucher, Olivier and Servonnat, Jérôme and Albright, Anna Lea and Aumont, Olivier and Balkanski, Yves and Bastrikov, Vladislav and Bekki, Slimane and Bonnet, Rémy and Bony, Sandrine and Bopp, Laurent and others, Presentation and evaluation of the IPSL-CM6A-LR climate model, *Journal of Advances in Modeling Earth Systems*, e2019MS002010

Buermann, W., M. Forkel, M. O'Sullivan, S. Sitch, P. Friedlingstein, V. Haverd, A. K. Jain, E. Kato, M. Kautz, S. Lienert, D. Lombardozzi, J. E. M. S. Nabel, H. Tian, A. J. Wiltshire, D. Zhu, W. K. Smith, and A. D. Richardson (2018), Widespread seasonal compensation effects of spring warming on northern plant productivity, *Nature*, *562*, 110–114, https://doi.org/10.1038/s41586-018-0555-7.

Caldararu, S., P.I. Palmer, and D.W. Purves (2012), Inferring Amazon leaf demography from satellite observations of leaf area index *Biogeosciences*, *9*, 1389–1404, https://doi.org/10.5194/bg-9-1389-2012.

Canadell, J., R. B. Jackson, J. B. Ehleringer, H. A. Mooney, and E.-D. Schulze (1996), Maximum rooting depth of vegetation types at the global scale *Oecologia*, *18*, 583–595, https://doi.org/10.1007/BF00329030.

Chen, J.M. and T.A. Black (1992), Defining leaf area index for non-flat leaves *Plant, Cell and Environment*, *15(4)*, 421–429.

Chen, C., Y. Knyazikhin, T. Park, K. Yan, A. Lyapustin, Y. Wang, B. Yang, and R. B. Myneni (2017), Prototyping of LAI and FPAR Retrievals from MODIS multi-angle implementation of atmospheric correction (MAIAC) data *Remote Sens.*, *9*, 370, https://doi.org/10.3390/rs9040370.

Chen, L., Hänninen, H., Rossi, S., Smith, N. G., Pau, S., Liu, Z., Feng, G., Gao, J., and Liu, J. (2020), Leaf senescence exhibits stronger climatic responses during warm than during cold autumns, *Nature Climate Change*, *10*, 777–780, https://doi.org/10.1038/s41558-020-0820-2.

Cherchi, A., P. G. Fogli, T. Lovato, D. Peano, D. Iovino, S. Gualdi, S. Masina, E. Scoccimarro, S. Materia, A. Bellucci, and A. Navarra (2019), Global Mean Climate and Main Patterns of Variability in the CMCC–CM2 Coupled Model, *Journal of Advances in Modeling Earth Systems*, *11*, 185–209. https://doi.org/10.1029/2018MS001369.

Churkina, G., D. Schimel, B. H. Braswell, and X. Xiao (2005), Spatial analysis of growing season length control over net ecosystem exchange, *Global Change Biology*, *11*, 1777–1787, https://doi.org/10.1111/j.1365-2486.2005.01012.x.

Clark, D. B., L. M. Mercado, S. Sitch, C. D. Jones, N. Gedney, M. J. Best, M. Pryor, G. G. Rooney, R. L. H. Essery, E. Blyth, O. Boucher, R. J. Harding, C. Huntingford, and P. M. Cox (2011), The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, *4*, 701–722, https://doi.org/10.5194/gmd-4-701-2011.

Cleland, E. E., I. Chuine, A. Menzel, H. A. Mooney, and M. D. Schwartz (2007), Shifting plant phenology in response to global change, *Trends in Ecology and Evolution*, *22*, 357–365. https://doi.org/10.1016/j.tree.2007.04.003.

Collatz, G. J. and Ribas-Carbo, M. and Berry, J. A., Coupled photosynthesis-stomatal conductance model for leaves of C4 plants, *Aust. J. Plant Physiol., 19*, 519–539, https://doi.org/10.1071/PP9920519

Dahlin, K. M., R. A. Fischer, and P. J. Lawrence (2015), Environmental drivers of drought deciduous phenology in the Community Land Model, *Biogeosciences*, *12*, 5061–5074, https://doi.org/10.5194/bg-12-5061-2015.

Dalmonech, D., and S. Zaehle (2013), Towards a more objective evaluation of modelled land-carbon trends using atmospheric CO2 and satellite-based vegetation activity observations, *Biogeosciences*, *10*, 4189–4210, https://doi.org/10.5194/bg-10-4189-2013.

Dalmonech, D., S. Zaehle, G. Schürmann, V. Brovkin, C.H. Reick, and R. Schnur (2015), Separation of the effects of land and climate model errors on simulated contemporary land carbon cycle trends in the MPI Earth System Model version 1, *Journal of Climate*, *28*, 272–291, https://doi.org/10.1175/JCLI-D-13-00593.1.

Decharme B., C. Delire, M. Minvielle, J. Colin, J.-P. Vergnes, A. Alias, D. Saint-Martin, R. Séférian, S. Sénési, and A. Voldoire (2019), Recent changes in the ISBA-CTRIP land surface system for use in the CNRM-CM6 climate model and in global off-line hydrological applications, *Journal of Advances in Modeling Earth Systems*, *11*, https://doi.org/10.1029/2018MS001545.

Delbart N., T. Le Toan, L. Kergoat and V. Fedotova (2006), Remote sensing of spring phenology in boreal regions: A free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–2004), *Remote Sensing of Environment*, *101*, 52–62, https://doi.org/10.1016/j.rse.2005.11.012.

Delire C., R. Séférian, B. Decharme, R. Alkama, J-C Calvet, D. Carrer, A-L. Gibelin, E. Joetzjer, X Morel, M. Rocher, D. Tzanos. (2020), The global land carbon cycle simulated with ISBA: improvements over the last decade, *Journal of Advances in Modeling Earth Systems*, *12*, https://doi.org/10.1029/2019MS001886.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q. (in review, 2021), The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, *Geosci. Model Dev. Discuss., [preprint]*, https://doi.org/10.5194/gmd-2020-446.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini (2012), Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, *Remote Sensing of Environment*, *120*, 25–36, https://doi.org/10.1016/j.rse.2011.11.026.

Eyring, V., S. Bony, G.A. Meehl, C.A. Senior, B. Stevens, R.J. Stouffer, and K.E. Taylor (2016), Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geosci. Model Dev.*, *9*, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Fang, H., C. Jiang, W. Li, S. Wei, F. Baret, J. M. Chen, J. Garcia-Haro, S. Liang, R. Liu, R. B. Myneni, B. Pinty, Z. Xiao, and Z. Zhu (2013), Characterization and intercomparison of global moderate resolution leaf area index (LAI) products: Analysis of climatologies and theoretical uncertainties, *Journal of Geophysical Research: Biogeosciences*, *118*, 529–548, https://doi.org/10.1002/jgrg.20051.

Farquhar, G. V. and von Caemmerer, S. V. and Berry, J. A. (1980), A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, *Planta*, 149(1), 78-90

Fisher, R.A., Wieder, W.R., Sanderson, B.M., Koven, C.D., Oleson, K.W., Xu, C., Fisher, J.B., Shi, M., Walker, A.P., and D.M. Lawrence (2019), Parametric Controls on Vegetation Responses to Biogeochemical Forcing in the CLM5, *Journal of Advances in Modeling Earth Systems*, *11*, 2879–2895. https://doi.org/10.1029/2019MS001609.

Forkel, M., Migliavacca, M., Thonicke, K., Reichstein, M., Schaphoff, S., Weber, U. and Carvalhais, N. (2015), Codominant water control on global interannual variability and trends in land surface phenology and greenness, *Glob Change Biol*, *21*, 3414–3435. https://doi.org/10.1111/gcb.12950.

Friedlingstein, P., M. Meinshausen, V.K. Arora, C.D. Jones, A. Anav, S.K. Liddicoat, and R. Knutti (2014), Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, *J. Climate*, *27*, 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1.

Friedlingstein, Pierre and Joel, G and Field, Chris B and Fung, Inez Y, Toward an allocation scheme for global terrestrial carbon models, *Global Change Biology, 5, 7*, 755–770,

Goll, D. S., A. J. Winkler, T. Raddatz, N. Dong, I. C. Prentice, P. Ciais, and V. Brovkin (2017), Carbon–nitrogen interactions in idealized simulations with JSBACH (version 3.10), *Geoscientific Model Development*, *10(5)*, 2009–2030. https://doi.org/10.5194/gmd-10-2009-2017.

Goudriaan J., van Laar H.H., van Keulen H., Louwerse W. (1985) Photosynthesis, CO2 and Plant Production. In: Day W., Atkin R.K. (eds) Wheat Growth and Modelling. NATO ASI Science (Series A: Life Sciences), vol 86. Springer, Boston, MA

Gumma, M. K., P. S. Thenkabail, P. Teluguntla, M. N. Rao, I. A. Mohammed, and A. M. Whitbread (2016) Mapping rice-fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data, *International Journal of Digital Earth*, *9:10*, 981–1003, https://doi.org/10.1080/17538947.2016.1168489.

Harper, A. B., P. M. Cox, P. Friedlingstein, A. J. Wiltshire, C. D. Jones, S. Sitch, L. M. Mercado, M. Groenendijk, E. Robertson, J. Kattge, G. Bönisch, O. K. Atkin, M. Bahn, J. Cornelissen, Ü. Niinemets, V. Onipchenko, J. Peñuelas, L. Poorter, P. B. Reich, N. A. Soudzilovskaia, and P. V. Bodegom (2016) Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information, *Geosci. Model Dev.*, *9*, 2415–2440, https://doi.org/10.5194/gmd-9-2415-2016.

Hagemann, S., and T. Stacke (2015) Impact of the soil hydrology scheme on simulated soil moisture memory, *Clim Dyn*, *44*, 1731–1750, https://doi.org/10.1007/s00382-014-2221-6.

Hazeleger, W., and R. Bintanja (2012) Studies with the EC-Earth seamless earth system prediction model, *Clim Dyn*, *39*, 2609–2610, https://doi.org/10.1007/s00382-012-1577-8.

Hemming, D., D. Peano, T. Park (in prep.) Plant phenology evaluation of CRESCENDO land surface models.Part II: peak and trough variations, *Biogeosciences*.

Hurtt, G. C., L. Chini, R. Sahajpal, S. Frolking, B. L. Bodirsky, K. Calvin, J. C. Doelman, J. Fisk, S. Fujimori, K. K. Goldewijk, T. Hasegawa, P. Havlik, A. Heinimann, F. Humpenöder, J. Jungclaus, J. Kaplan, J. Kennedy, T. Kristzin, D. Lawrence, P. Lawrence, L. Ma, O. Mertz, J. Pongratz, A. Popp, B. Poulter, K. Riahi, E. Shevliakova, E. Stehfest, P. Thornton, F. N. Tubiello, D. P. van Vuuren, and X. Zhang (2020) Harmonization of Global Land-Use Change and Management for the Period 850–2100 (LUH2) for CMIP6, *Geosci. Model Dev. Discuss.*, https://doi.org/10.5194/gmd-2019-360.

Jackson, R.B., J. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze (1996), A global analysis of root distributions for terrestrial biomes, *Oecologia*, *108*, 389–411, https://doi.org/10.1007/BF00333714.

Jeong, S.-J., C.-H. Ho, H.-J. Gim, and M. E. Brown (2011), Phenology shifts at start vs end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008, *Global Change Biology*, *17*, 2385–2399, https://doi.org/10.1111/j.1365-2486.2011.02397.x.

Jiang, C., Y. Ryu, H. Fang, R. Myneni, M. Claverie, and Z. Zhu, (2017) Inconsistencies of interannual variability and trends in long-term satellite leaf area index products, *Global Change Biology*, *23*, 4133–4146, https://doi.org/10.1111/gcb.13787.

Jolly, W.M., Nemani, R. and Running, S.W. (2005), A generalized, bioclimatic index to predict foliar phenology in response to climate, *Global Change Biology*, *11*, 619–632, https://doi.org/10.1111/j.1365-2486.2005.00930.x.

Jones, P. W. (1999), First and second order conservative remapping schemes for grids in spherical coordinates, *Monthly Weather Review*, *127*, 2204–2210.

Kandasamy, S., F. Baret, A. Verger, P. Neveux, and M. Weiss (2013), A comparison of methods for smoothing and gap filling time series of remote sensing observations - application to MODIS LAI products, *Biogeosciences*, *10*, 4055–4071, https://doi.org/10.5194/bg-10-4055-2013.

Keenan, T. F., J. Gray, M. A. Friedl, M. Toomey, G. Bohrer, D. Y. Hollinger, J. W. Munger, J. O'Keefe, H. P. Schmid, I. S. Wing, B. Yang, and A. D. Richardson (2014), Net carbon uptake has increased through warming-induced changes in temperate forest phenology, *Nature Climate Change*, *4*, 598–604, https://doi.org/10.1038/NCLIMATE2253.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., and Willis, K. O. (2013), A comprehensive benchmarking system for evaluating global vegetation models, *Biogeosciences*, *10*, 3313—3340, https://doi.org/10.5194/bg-10-3313-2013, 2013.

Kleidon, A. (2004), Global datasets and rooting zone depth inferred from inverse methods, *J. Clim.*, *17(13)*, 2714–2722, https://doi.org/10.1175/1520-0442(2004)017<2714:GDORZD>2.0.CO;2 .

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P. (2010), Carbon cycle data assimilation with a generic phenology model, *J. Geophys. Res*, *115*, G04017, https://doi.org/10.1029/2009JG001119.

Koven, C. D., W. J. Riley, Z. M. Subin, J. Y. Tang, M. S. Torn, W. D. Collins, G. B. Bonan, D. M. Lawrence, and S. C. Swenson (2013), The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4, *Biogeosciences*, *10*, 7109–7131, https://doi.org/10.5194/bg-10-7109-2013.

Krinner, G. and Viovy, Nicolas and de Noblet-Ducoudré, Nathalie and Ogée, Jérôme and Polcher, Jan and Friedlingstein, Pierre and Ciais, Philippe and Sitch, Stephen and Prentice, I. Colin, A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles, 19, 1*, https://doi.org/10.1029/2003GB002199

Kucharik, C. J., C. C. Barford, M. El Maayar, S. C. Wofsy, R. K. Monson, and D. D. Baldocchi (2006), A multiyear evaluation of a Dynamic Global Vegetation Model at three AmeriFlux forest sites: Vegetation structure, phenology, soil temperature, and CO2 and H2O vapor exchange, *Ecological Modelling*, *196*, 1–31, https://doi.org/10.1016/j.ecolmodel.2005.11.031.

Lawrence, D.M., R.A. Fisher, C.D. Koven, K.W. Oleson, S.C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P.J. Lawrence, F. Li, H. Li, D. Lombardozzi, W.J. Riley, W.J. Sacks, M. Shi, M. Vertenstein, W.R. Wieder,, C. Xu, A.A. Ali, A.M. Badger, G. Bisht, M.A. Brunke, S.P. Burns,, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J.B. Fisher, M. Flanner, A.M. Fox, P. Gentine, F.Hoffman, G. Keppel-Aleks, R., Knox, S. Kumar, J. Lenaerts, L.R. Leung, W.H. Lipscomb, Y. Lu, A., Pandey, J.D. Pelletier, J. Perket,, J.T. Randerson, D.M. Ricciuto, B.M., Sanderson, A. Slater, Z.M. Subin, J. Tang, R.Q. Thomas, M. Val Martin, and X.

Zeng (2018), Technical description of version 5.0 of the community land model (CLM), *Tech. Rep. Ncar Technical Note NCAR*, National Center for Atmospheric Research, Boulder, CO.

Lawrence, D.M., R.A. Fisher, C.D. Koven, K.W. Oleson, S.C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P.J. Lawrence, F. Li, H. Li, D. Lombardozzi, W.J. Riley, W.J. Sacks, M. Shi, M. Vertenstein, W.R. Wieder,, C. Xu, A.A. Ali, A.M. Badger, G. Bisht, M. van den Broeke, M.A. Brunke, S.P. Burns,, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J.B. Fisher, M. Flanner, A.M. Fox, P. Gentine, F.Hoffman, G. Keppel-Aleks, R., Knox, S. Kumar, J. Lenaerts, L.R. Leung, W.H. Lipscomb, Y. Lu, A., Pandey, J.D. Pelletier, J. Perket,, J.T. Randerson, D.M. Ricciuto, B.M., Sanderson, A. Slater, Z.M. Subin, J. Tang, R.Q. Thomas, M. Val Martin, and X. Zeng (2019), The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty., *Journal of Advances in Modeling Earth Systems*, *11*, 4245–4287, https://doi.org/10.1029/2018MS001583

Levis, S., A. Badger, B. Drewniak, C. Nevison, and X. Ren (2018), CLMcrop yields and water requirements: avoided impacts by choosing RCP 4.5 over 8.5, *Climatic Change*, *146*, 501–515, https://doi.org/10.1007/s10584-016-1654-9

Li, W., N. MacBean, P. Ciais, P. Defourny, C. Lamarche, S. Bontemps, R. A. Houghton, and S. Peng (2018), Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015), *Earth System Science Data*, *10*, 219–234, https://doi.org/10.5194/essd-10-219-2018

Liebmann, B., I. Bladé, G. N. Kiladis, L. M. V. Carvalho, G. B. Senay, D. Allured, S. Leroux, and C. Funk (2012), Seasonality of African Precipitation from 1996 to 2009, *Journal of Climate*, *25*, 4304–4322, https://doi.org/10.1175/JCLI-D-11-00157.1

Lindeskog, M., A. Arneth, A. Bondeau, K. Waha, J. Seaquist, S. Olin, and B. Smith (2013), Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, *Earth Syst. Dynam.*, *4*, 385–407, https://doi.org/10.5194/esd-4-385-2013

Thibaut Lurton and Yves Balkanski and Vladislav Bastrikov and Slimane Bekki and Laurent Bopp and Patrick Brockmann and Patricia Cadule and Anne Cozic and David Cugnet and Jean-Louis Dufresne and Christian Éthé and Marie-Alice Foujols and Josefine Ghattas and Didier Hauglustaine and Rong-Ming Hu and Masa Kageyama and Myriam Khodri and Nicolas Lebas and Guillaume Levavasseur and Marion Marchand and Catherine Ottlé and Philippe Peylin and Adriana Sima and Sophie Szopa and Rémi Thiéblemont and Nicolas Vuichard and Olivier Boucher, Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model, *JAMES*, https://doi.org/2019MS001940

MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F.-M., and Ciais, P. (2015), Using satellite data to improve the leaf phenology of a global terrestrial biosphere model, *Biogeosciences*, *12*, 7185–7208, https://doi.org/10.5194/bg-12-7185-2015.

Maignan, F., F.-M. Bréon, F. Chevallier, N. Viovy, P. Ciais, C. Garrec, J. Trules, and M. Mancip (2011), Evaluation of a global vegetation model using time series of satellite vegetation indices, *Geosci. Model Dev.*, *4*, 1103–1114, https://doi.org/10.5194/gmd-4-1103-2011.

Maisongrande, P., B. Duchemin, and G. Dedieu (2004), VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products, *International Journal of Remote Sensing*, *25:1*, 9–14, https://doi.org/10.1080/0143116031000115265.

Martiny, N., P. Camberlin, Y. Richard, and N. Philippon (2006), Compared regimes of NDVI and rainfall in semi-arid regions of Africa, *Int. J. Remote Sens.*, *27*, 5201–5223, https://doi.org/10.1080/01431160600567787.

Mauritsen, T., J. Bader, T. Becker, J. Behrens, M. Bittner, R. Brokopf, V. Brovkin, M. Claussen, T. Crueger, M. Esch, I. Fast, S. Fiedler, D. Fläschner, V. Gayler, M. Giorgetta, D. S. Goll, H. Haak, S. Hagemann, C. Hedemann, C. Hohenegger, T. Ilyina, T. Jahns, D. Jimenéz-de-la-Cuesta, J. Jungclaus, T. Kleinen, S. Kloster, D. Kracher, S. Kinne, D. Kleberg, G. Lasslop, L. Kornblueh, J. Marotzke, D. Matei, K. Meraner, U. Mikolajewicz, K. Modali, B. Möbis, W. A. Müller, J. E. M. S. Nabel, C. C. W. Nam, D. Notz, S.-S. Nyawira, H. Paulsen, K. Peters, R. Pincus, H. Pohlmann, J. Pongratz, M. Popp, T. J. Raddatz, S. Rast, R. Redler, C. H. Reick, T. Rohrschneider, V. Schemann, H. Schmidt, R. Schnur, U. Schulzweida, K. D. Six, L. Stein, I. Stemmler, B. Stevens, J.-S. von Storch, F. Tian, A. Voigt, P. Vrese, K.-H. Wieners, S.

Wilkenskjeld, A. Winkler, and E. Roeckner (2019), Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO2., *Journal of Advances in Modeling Earth Systems*, *11*, 998–1038, https://doi.org/10.1029/2018MS001400.

Medlyn, B.E., Duursma, R.A., Eamus, D., Ellsworth, D.S., Prentice, I.C., Barton, C.V.M., Crous, K.Y., De Angelis, P., Freeman, M., and L. Wingate (2011), Reconciling the optimal and empirical approaches to modelling stomatal conductance: RECONCILING OPTIMAL AND EMPIRICAL STOMATAL MODELS, *Global Change Biology*, *17*, 2134–2144, https://doi.org/10.1111/j.1365-2486.2010.02375.x.

Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavská, O., Briede, A., Chmielewski, F.M., Crepinsek, Z., Curnel, Y., Dahl, Å., Defila, C., Donnelly, A., Filella, Y., Jatczak, K., Måge, F., Mestre, A., Nordli, Ø., Peñuelas, J., Pirinen, P., Remisŏvá, V., Scheifinger, H., Striz, M., Susnik, A., Van Vliet, A.J.H., Wielgolaski, F.-E., Zach, S. and Zust, A. (2006), European phenological response to climate change matches the warming pattern, *Global Change Biology*, *12*, 1969–1976, https://doi.org/10.1111/j.1365-2486.2006.01193.x.

Miller et al. (in prep.) Coupled dynamics of climate, vegetation, land use and terrestrial biogeochemistry in the EC-Earth-Veg ESM - implementation and evaluation.

Murray, M. B., M.G.R. Cannell and R.I. Smith (1989), Date of budburst of fifteen tree species in Britain following climatic warming, *J. Appl. Ecology*, *26*, 693–700.

Murray-Tortarolo, G., A. Anav, P. Friedlingstein, S. Sitch, S. Piao, Z. Zhu, B. Poulter, S. Zaehle, A. Ahlström, M. Lomas, S. Levis, N. Viovy, and N. Zeng (2013), Evaluation of land surface models in reproducing satellite-derived LAI over the high-latitude northern hemisphere. part I: uncoupled DGVMs, *Remote Sens.*, *5*, 4819–4838, https://doi.org/10.3390/rs5104819.

Myneni, R. B., S. Hoffman, Y. Knyazikhin, J. L. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G. R. Smith, A. Lotsch, M. Friedl, J. T. Morisette, P. Votava, R. R. Nemani, and S. W. Running (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, *83*, 214–231.

Myneni, R., Y. Knyazikhin, T. Park. (2015), MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD15A2H.006, Accessed, 13 November 2019.

Myneni, R., Y. Knyazikhin, T. Park. (2015), MYD15A2H MODIS/Aqua Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006, distributed by NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MYD15A2H.006, Accessed, 13 November 2019.

Oleson, K. W., D. Lawrence, G. Bonan, B. Drewniak, M. Huang, C. Koven, S. Levis, F. Li, W. Riley, Z. Subin, S. Swenson, P. Thornton, A. Bozbiyik, R. Fisher, E. Kluzek, J.-F. Lamarque, P. Lawrence, L. Leung, W. Lipscomb, S. Muszala, D. Ricciuto, W. Sacks, Y. Sun, J. Tang, and Z.-L. Yang (2013), Technical description of version 4.5 of the community land model (CLM), *Tech. Rep. Ncar Technical Note NCAR/TN-503+STR*, National Center for Atmospheric Research, Boulder, CO, https://doi.org/10.5065/D6RR1W7M.

Olin, S., M. Lindeskog, T. a. M. Pugh, G. Schurgers, D. Wårlind, M. Mishurov, S. Zaehle, B. D. Stocker, B. Smith, and A. Arneth (2015), Soil carbon management in large-scale Earth system modelling: implications for crop yields and nitrogen leaching, *Earth System Dynamics*, *6(2)*, 745–768, https://doi.org/10.5194/esd-6-745-2015.

Park, T., S. Ganguly, H. Tømmervik, E. S. Euskirchen, K.-A. Høgda, S. R. Karlsen, V. Brovkin, R. R. Nemani, and R. B. Myneni (2016), Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data, *Environmental Research Letters, 11 (8)*, 084001, https://doi.org/10.1088/1748-9326/11/8/084001

W. J. Parton and D. S. Schimel and C. V Cole and D. S. Ojima, Analysis of factors controlling soil organic-matter levels in Great-Plains grasslands, *Soil Sci. Soc. Am. J., 51*, 1173–1179

Peano, D., S. Materia, A. Collalti, A. Alessandri, A. Anav, A. Bombelli, and S. Gualdi (2019), Global variability of simulated and observed vegetation growing season, *JGR: Biogeosciences*, *124*, https://doi.org/10.1029/2018JG004881.

Piao, S., X. Wang, T. Park, C. Chen, X. Lian, Y. He, J. W. Bjerke, A. Chen, P. Ciais, H. Tømmervik, R. R. Nemani and R. B. Myneni (2020), Characteristics, drivers and feedbacks of global greening., *Nat Rev Earth Environ*, *1*, 14–27, https://doi.org/10.1038/s43017-019-0001-x.

Pongratz, J., C. Reick, T. Raddatz, and M. Claussen (2008), A re-construction of global agricultural areas and land cover for the last millennium, *Global Biogeochem. Cy.*, *22*, GB3018, https://doi.org/10.1029/2007GB003153.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler, V. (2013), Representation of natural and anthropogenic land cover change in MPI-ESM, *J. Adv. Model. Earth Syst.*, *5*, 459–482, https://doi.org/10.1002/jame.20022.

Reick, C., Gayler, V., Goll, D., Hagemann, S., Heidkamp, M., Nabel, J., Raddatz, T., Roeckner, E., Schnur, R. and Wilkenskjeld, S (2021), JSBACH 3 - The land component of the MPI Earth System Model: documentation of version 3.2., *Berichte zur Erdsystemforschung*, *240*.

Richardson, A.D., R.S. Anderson, M.A. Arain, A.G. Barr, G. Bohrer, G. Chen, J.M. Chen, P. Ciais, K.J. Davis, A.R. Desai, M.C. Dietze, D. Dragoni, S.R. Garrity, C.M. Gough, R. Grant, D.Y. Hollinger, H.A. Margolis, H. McCaughey, M. Migliavacca, R.K. Monson, J.W. Munger, B. Poulter, B.M. Raczka, D.M. Ricciuto, A.K. Sahoo, K. Schaefer, H. Tian, R. Vargas, H. Verbeeck, J. Xiao, and Y. Xue (2012), Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, *Glob. Change Biol.*, *18*, 566–584, https://doi.org/10.1111/j.1365-2486.2011.02562.x.

Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey (2013), Climate change, phenology, and phenological control of vegetation feedbacks to the climate system, *Agric. For. Meteorol.*, *169*, 156–173, https://doi.org/10.1016/j.agrformet.2012.09.012.

Richardson, A. D., K. Hufkens, T. Milliman, D. M. Aubrecht, M. Chen, J. M. Gray, M. R. Johnston, T. F. Keenan, S. T. Klosterman, M. Kosmala, E. K. Melaas, M. A. Friedl, and S. Frolking (2018), Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery, *Scientific Data*, *5*, 180028, https://doi.org/10.1038/sdata.2018.28.

Savoy, P., and D.S. Mackay (2015), Modeling the seasonal dynamics of leaf area index based on environmental constraints to canopy development, *Agric. For. Meteorol.*, *200*, 46–56, https://doi.org/10.1016/j.agrformet.2014.09.019.

Schwartz, M.D., R. Ahas, and A. Aasa (2006), Onset of spring starting earlier across the Northern Hemisphere, *Global Change Biology*, *12*, 343–351, https://doi.org/10.1111/j.1365-2486.2005.01097.x.

Smith, B., D. Wårlind, A. Arneth, T. Hickler, P. Leadley, J. Siltberg, and S. Zaehle (2014), Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, *11(7)*, 2027–2054, https://doi.org/10.5194/bg-11-2027-2014.

Seland, Ø., Bentsen, M., Seland Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., Debernard, J.B., Gupta, A.K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Schancke Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Hafsahl Karset, I.H., Landgren, O., Liakka, J., Onsum Moseid, K., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iverson, T., and M. Schulz (2020), The Norwegian Earth System Model, NorESM2 - Evaluation of the CMIP6 DECK and historical simulations, *Geosci. Model Dev. Discuss.*, https://doi.org/10.5194/gmd-2019-378.

Sitch, S., P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, A. Ahlström, S. C. Doney, H. Graven, C. Heinze, C. Huntingford, S. Levis, P. E. Levy, M. Lomas, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, A. Arneth, G. Bonan, L. Bopp, J. G. Canadell, F. Chevallier, P. Ciais, R. Ellis, M. Gloor, P. Peylin, S. L. Piao, C. Le Quéré, B. Smith, Z. Zhu, and R. Myneni (2015), Recent trends and drivers of regional sources and sinks of carbon dioxide, *Biogeosciences*, *12*, 643–679, https://doi.org/10.5194/bg-12-653-2015.

Stöckli, R., Rutishauser, T., Baker, I., Liniger, M. A., and Denning, A. S. (2011), A global reanalysis of vegetation phenology, *J. Geophys. Res.*, *116*, G03020, https://doi.org/10.1029/2010JG001545.

Swenson, S., and D. Lawrence, (2014), Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data, *Journal of Geophysical Research. Atmospheres*, *119(17)*, 10,299–10,312, https://doi.org/10.1002/2014JD022314

Sellar, A., J. Walton, C. G. Jones, N. L. Abraham, M. Andrejczuk, M. B. Andrews, T. Andrews, A. T. Archibald, L. de Mora, H. Dyson, M. Elkington, R. Ellis, P. Florek, P. Good, L. Gohar, S. Haddad, S. C. Hardiman, E. Hogan, A. Iwi, C. D. Jones, B. Johnson, D. I. Kelley, J. Kettleborough, J. R. Knight, M. O. Köhler, T. Kuhlbrodt, S. Liddicoat, I. Linova-Pavlova, M. S. Mizielinski, O. Morgenstern, J. Mulcahy, E. Neininger, F. M. O'Connor, R. Petrie, J. Ridley, J. -C. Rioual, M. Roberts, E. Robertson, S. Rumbold, J. Seddon, H. Shepherd, S. Shim, A. Stephens, J. C. Teixeira, Y. Tang, J. Williams, A. Wiltshire (2020), Implementation of UK Earth system models for CMIP, *J.ournal of Advances in Modeling Earth Systems*, *12*, e2019MS001946, https://doi.org/doi.org/10.1029/2019MS001946.

Soudani, K., G. le Maire, E. Dufrêne, C. François, N. Delpierre, E. Ulrich, and S. Cecchini (2008), Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data, *Remote Sensing of Environment*, *112*, 2643–2655, https://doi.org/10.1016/j.rse.2007.12.004.

Tang, H. and Dubayah, R. (2017), Light-driven growth in Amazon evergreen forests explained by seasonal variations of vertical canopy structure, *Proceedings of the National Academy of Sciences*, *114*, 2640–2644, https://doi.org/10.1073/pnas.1616943114.

Thornton, P. E., B. E. Law, H. L. Gholz, K. L. Clark, E. Falge, D. S. Ellsworth, A. H. Goldstein, R. K. Monson, D. Hollinger, M. Falk, J. Chen, and J. P. Sparks (2002), Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forest, *Agric. For. Meteorol.*, *113*, 185–222.

Verger, A., F. Baret, and M. Weiss (2011), A multisensor fusion approach to improve LAI time series, *Remote Sensing of Environment*, *115*, 2460–2470, https://doi.org/10.1016/j.rse.2011.05.006.

Verger, A., F. Baret, and M. Weiss (2019), Copernicus Global Land Operations "Vegetation and Energy", https://land.copernicus.eu/global/products/LAI, Copernicus Global Land operations –Lot 1, Date Issued: 11.03.2019, Issue: I1.41.

Viovy, N. (2018), CRUNCEP Version 7 - Atmospheric Forcing Data for the Community Land Model, http://rda.ucar.edu/datasets/ds314.3/, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo., Accessed, 15 July 2016.

Waha, K., L. G. J. van Bussel, C. Muller, and A. Bondeau, A. (2011), Climate-driven simulation of global crop sowing dates, *Global Ecol. Biogeogr.*, *21*, 247–259, https://doi.org/10.1111/j.1466-8238.2011.00678.x.

White, M. A., P. E. Thornton, and S. W. Running (1997), A continental phenology model for monitoring vegetation responses to interannual climatic variability, *Global Biogeochem. Cycles*, *11*, 217–234.

Wiltshire, A. J., M. C. Duran Rojas, J. M. Edwards, N. Gedney, A. B. Harper, A. J. Hartley, M. A. Hendry, E. Robertson, and K. Smout-Day (2020a), JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2, *Geosci. Model Dev.*, *13*, 483–505, https://doi.org/10.5194/gmd-13-483-2020.

Wiltshire, A. J. and Burke, E. J. and Chadburn, S. E. and Jones, C. D. and Cox, P. M. and Davies-Barnard, T. and Friedlingstein, P. and Harper, A. B. and Liddicoat, S. and Sitch, S. A. and Zaehle, S. (2020b), JULES-CN: a coupled terrestrial Carbon-Nitrogen Scheme (JULES vn5.1), *Geosci. Model Dev.*, 1–40, https://doi.org/10.5194/gmd-2020-205.

Wu, W.-B., P. Yang, H.-J. Tang, Q.-B. Zhou, Z.-X. Chen, and R. Shibasaki (2010), Characterizing Spatial Patterns of Phenology in Cropland of China Based on Remotely Sensed Data, *Agricultural Sciences in China*, *9*, 101–112, https://doi.org/10.1016/S1671-2927(09)60073-0.

Yan, K., T. Park, G. Yan, C. Chen, B. Yang, Z. Liu, R. R. Nemani, Y. Knyazikhin, and R. B. Myneni (2016), Evaluation of MODIS LAI/FPAR Product Collection 6. Part 1: Consistency and Improvements, *Remote Sens.*, *8*, 359, https://doi.org/10.3390/rs8050359.

Yan, K., T. Park, G. Yan, Z. Liu, B. Yang, C. Chen, R. R. Nemani, Y. Knyazikhin, and R. B. Myneni (2016), Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison, *Remote Sens.*, *8*, 460, https://doi.org/10.3390/rs8060460.

X. Yin and P.C. Struik, C3 and C4 photosynthesis models: An overview from the perspective of crop modelling, *NJAS - Wageningen Journal of Life Sciences, 57*, 27–38, https://doi.org/10.1016/j.njas.2009.07.001

Yool, A., J. Palmiéri, C. G. Jones, A. A. Sellar, L. de Mora, T. Kuhlbrodt, E. E. Popova, J. P. Mulcahy, A. Wiltshire, S. T. Rumbold, M. Stringer, R. S. R. Hill, Y. Tang, J. Walton, A. Blaker, A. J. G. Nurser, A. C. Coward, J. Hirschi, S. Woodward, D. I. Kelley, R. Ellis, and S. Rumbold-Jones (2020), Spin-up of UK Earth System Model 1 (UKESM1) for CMIP6, *Journal of Advances in Modeling Earth Systems*, *12*, 460, https://doi.org/10.1029/2019MS001933.

Zeng, X. (2001), Global vegetation root distribution for land modeling *J. Hydrometeor.*, *2*, 525–530.

Zeng, F.-W., G. J. Collatz, J.E. Pinzon, and A. Ivanoff (2013), Evaluating and Quantifying the Climate-Driven Interannual Variability in Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) at Global Scales, *Remote Sens.*, *5*, 3918-3950, https://doi.org/10.3390/rs5083918.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, F. Gao, B. C. Reed, and A. Huete (2003), Monitoring vegetation phenology using MODIS, *Remote Sens. Environ*, *84*, 471–475.

Zhang, X., M. A. Friedl, C. B. Schaaf, and A. H. Strahler (2005), Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments, *Journal of Geophysical Research*, *110*, D12103, https://doi.org/10.1029/2004JD005263.

Zhang, L., H. Lei, H. Shen, Z. Cong, D. Yang, and T. Liu (2019), Evaluating the representation of vegetation phenology in the Community Land Model 4.5 in a temperate grassland, *Journal of Geophysical Research: Biogeosciences*, *124*, https://doi.org/10.1029/2018JG004866

Zhang, Y., Commane, R., Zhou, S., Williams, A. P., and Gentine, P. (2020), Light limitation regulates the response of autumn terrestrial carbon uptake to warming, *Nature Climate Change*, *10*, 739–743, https://doi.org/10.1038/s41558-020-0806-0.

Zhu, Z., J. Bi, Y. Pan, S. Ganguly, A. Anav, L. Xu, A. Samanta, S. Piao, R. R. Nemani, and R. B. Myneni (2013), Global data sets of vegetationLeaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FAPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011, *Remote Sens.*, *5*, 927–948, https://doi.org/10.3390/rs5020927.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N. (2016), Greening of the Earth and its drivers, *Nature Climate Change*, *6*, 791–795, https://doi.org/10.1038/nclimate3004.

Table 1. Grid spatial resolution used for each land surface model (LSM) and brief summary of their main features. PFT stands for Plant Functional Type and CFT stands for Crop Functional Type. The second part gives further details on the temperature and moisture thresholds for start (S) and end (E) of the growing season in the phenology schemes.

LSM	Original Resolution	PFT	Soil level	CFT	Phenology scheme	Phenology drivers	Root zone
CLM 4.5	1.25° x 0.9375°	15	15	1 C3	evergreen; seasonal-deciduous; stress-deciduous	Soil temperature; soil moisture; day-length	*Zeng* (2001)
CLM 5.0	0.5° x 0.5°	15	20	2 C3	evergreen; seasonal-deciduous; stress-deciduous	Soil temperature; moisture day-length; precipitation	*Jackson et al.* (1996)
JULES-ES	1.875° x 1.25°	13	4	1 C3, 1 C4	Deciduous trees	Surface temperature	*Wiltshire et al.* (2020a)
JSBACH	1.9° x 1.9°	12	5	1 C3, 1 C4	evergreen; summergreen; raingreen; grasses; tropical crops; extra-tropical crops	air temperature; soil temperature; soil moisture; NPP	*Kleidon* (2004)
LPJ-GUESS	0.5° x 0.5°	25	2	3 C3, 2 C4	evergreen; seasonal-deciduous; stress-deciduous	Air temperature; soil moisture	Root in top soil layer†: Herbaceous PFTs 90%, Woody PFTs 60%
ORCHIDEE	0.5° x 0.5°	15	11	1 C3, 1 C4	deciduous; dry and semi-arid; grasses and crops grasses and crops	Air temperature; soil moisture soil moisture	Exponential profile within the 2m soil *Krinner et al.* (2005)
ISBA-CTRIP	1° x 1°	16	14	1 C3, 1 C4	leaf biomass	Leaf biomass	*Canadell et al.* (1996)

LSM	Temperature variable	Temperature threshold	Moisture variable	Moisture Threshold	Reference
CLM 4.5	Third soil layer* temperature	0 °C (S,E)	Third soil layer* water potential	-2 MPa (S,E)	*Oleson et al.* (2013)
CLM 5.0	Third soil layer** temperature	0 °C (S,E)	Third soil layer** water potential	-0.6 MPa (S), -2 MPa (E)	*Lawrence et al.* (2018)
JULES-ES	mean daily surface temperature	depending on PFT and whether S/E: 5 °C, 6.85 °C (S,E)	None	N/A	*Clark et al.* (2011)
JSBACH	depending on the phenology: air, pseudo-soil temperature	depending on the phenology and whether S/E: 4°C, 10°C, or critical heat sum (S,E)	soil moisture in the root zone	wilting point of 0.35 m/m (S,E)†	*Mauritsen et al.* (2019) *Reick et al.* (2021)
LPJ-GUESS	Mean daily air temperature	Sum above 5 °C (S)	water stress scalar (ω)	minimum of ω ($\omega_{\min}$) (S,E)	*Smith et al.* (2014)
ORCHIDEE	mean daily air temperature weekly temperature	Sum above -5 °C, 0°C (S) depending on PFT: 2, 5, 7, 10, 12 °C (E)	soil moisture in root zone weekly relative soil moisture	5 days moisture increase (S) depending on PFT: 0.2, 0.3 m^3/m^3 (E)	*Botta et al.* (2000) *Krinner et al.* (2005)
ISBA-CTRIP	no phenology model: LAI is deduced from leaf biomass through Specific Leaf Area, which depends on nitrogen content				*Delire et al.* (2020)

* about 6cm depth. ** 9cm depth. †relative soil moisture content above which growth is possible and below which growth stops and shedding sets in.

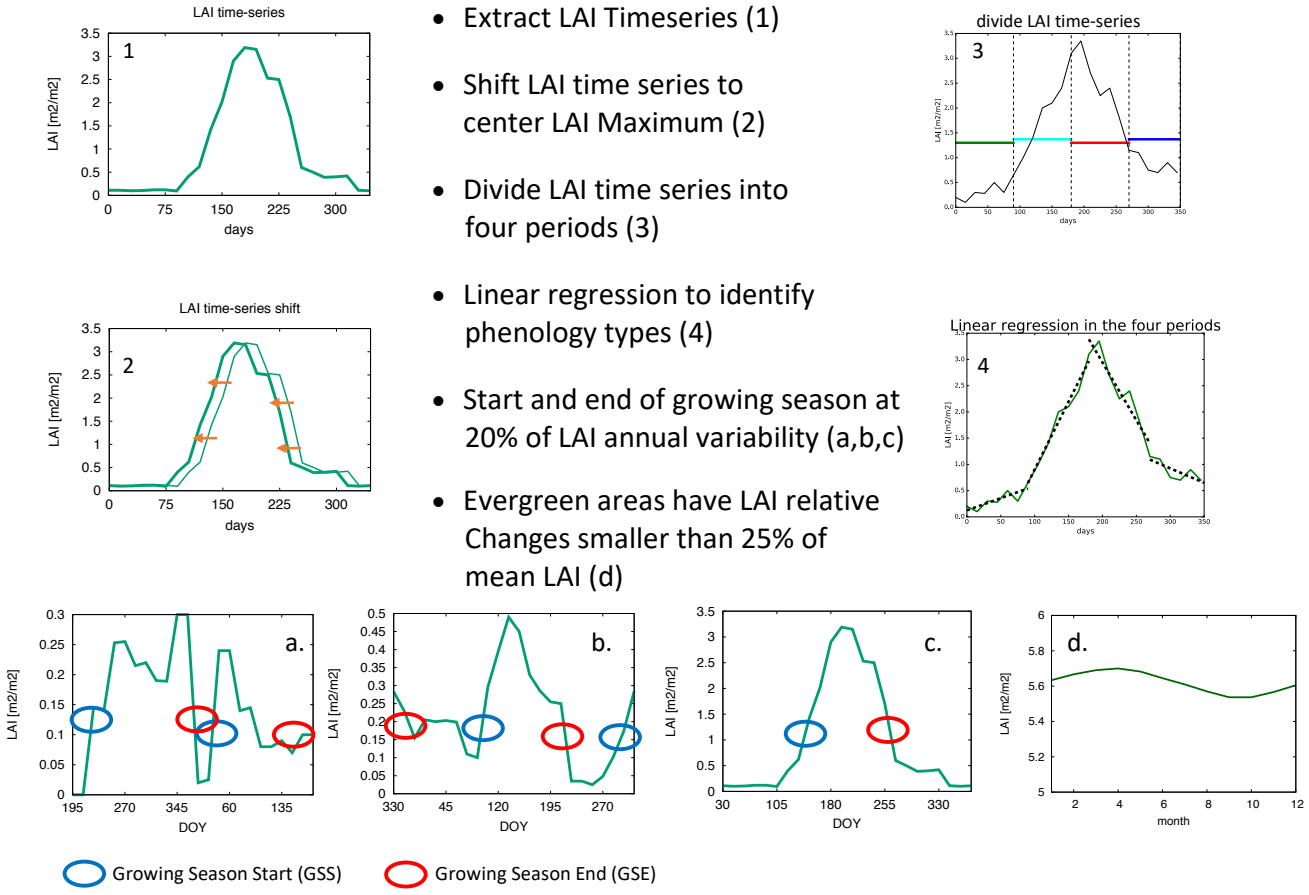

Figure 1. Scheme of the Four Growing Season Type method used in evaluating start and end of the growing season.

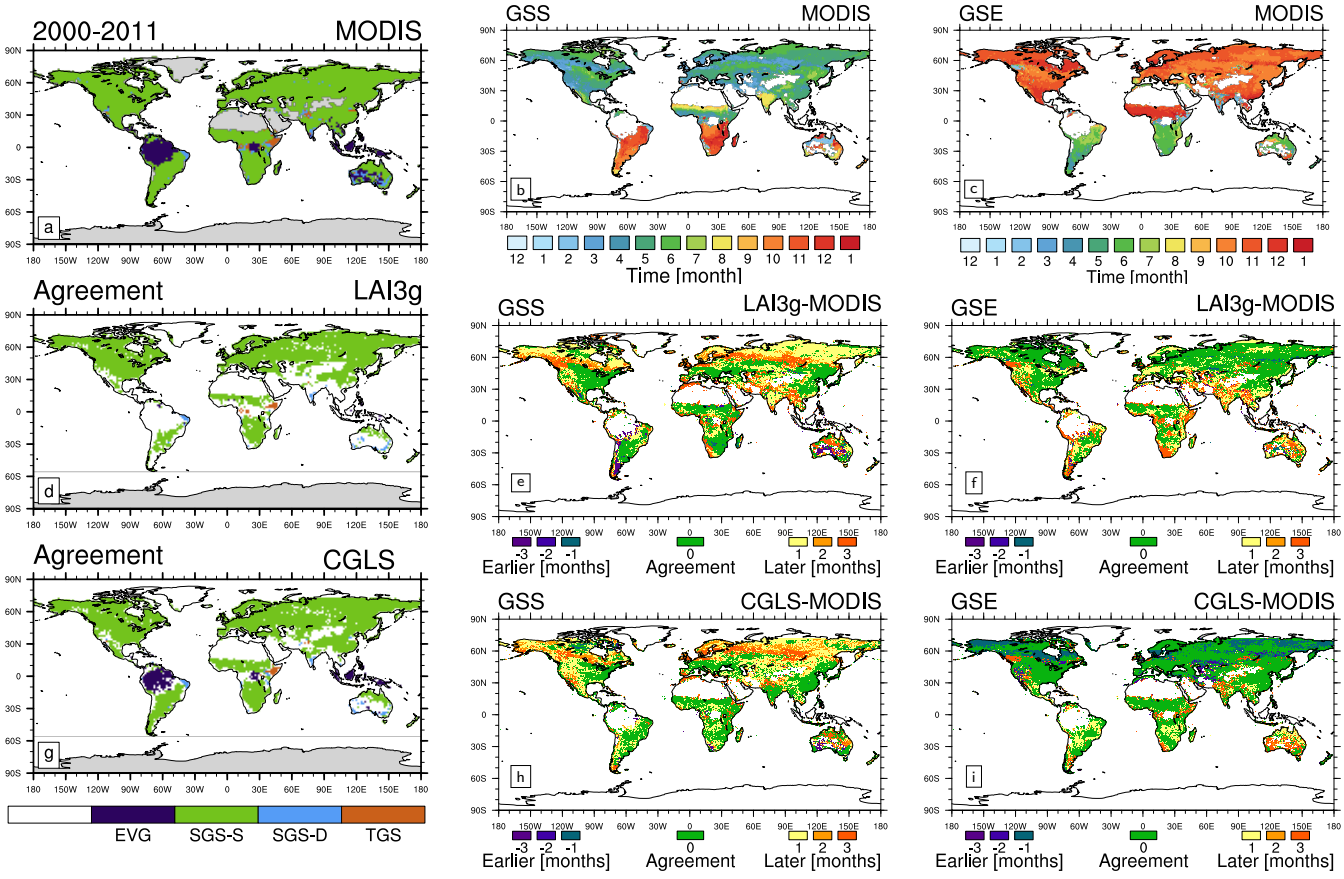

Figure 2. Global climatological (averaged over 2000-2011) distribution of (a) the four main growing season modes, (b) growing season start (GSS) timings, and (c) growing season end (GSE) timings for MODIS version 6. The other panels show the comparison between MODIS and LAI3g (second row), and CGLS (third row). In particular, panels (d) and (g) show the areas characterized by the same phenology types in both MODIS and LAI3g and CGLS, respectively; panels (e) and (h) exhibit the difference in GSS timings while panels (f) and (i) display the differences in GSE timings. In panels (d) and (g) white areas represent non-vegetated areas and regions of disagreement between MODIS and LAI3g and CGLS. White areas in panels (b),(c),(e),(f),(h), and (i) show evergreen and non-vegetated areas.

Table 2. The fraction of land grid-cell in agreement with MODIS for each satellite product and each land surface model in the four growing season types. Values are reported in percentage, and refer to the colored regions in Figures 2d,g and 3.

	LAI3g	CGLS	CLM 4.5	CLM 5.0	JULES-ES	JSBACH	LPJ-GUESS	ORCHIDEE	ISBA-CTRIP	MME
EVG	10.8	78.3	58.1	72.9	95.4	36.0	55.2	72.3	75.3	51.0
SGS-S	84.4	89.9	64.3	44.3	47.9	71.5	73.8	79.5	74.1	77.2
SGS-D	68.3	80.9	47.6	36.3	7.1	50.6	33.2	33.1	58.8	23.8
TGS	37.2	35.9	19.1	16.9	0.4	13.9	15.4	7.9	12.9	0.7
Total	75.4	86.4	61.2	45.4	48.3	65.2	68.1	74.0	71.1	69.5

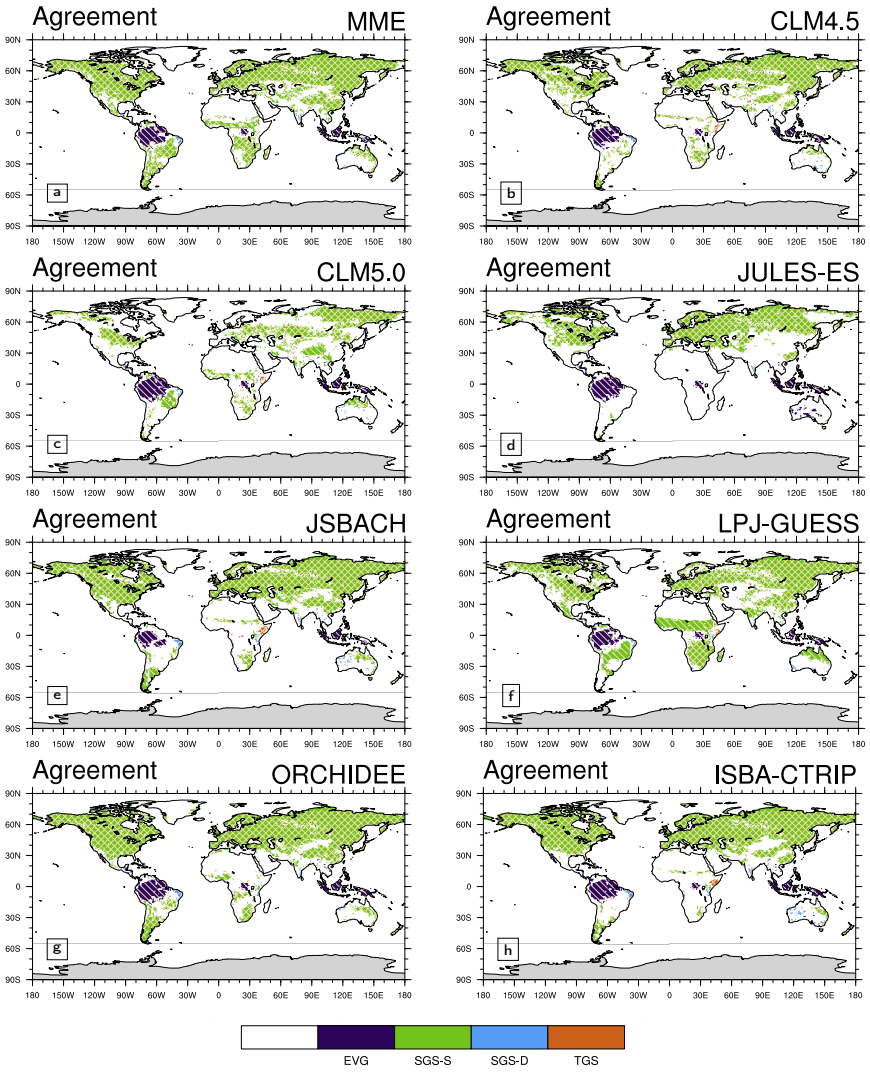

Figure 3. Global climatological (averaged over 2000-2011) distribution of the four main growing season modes for (a) MME; (b) CLM 4.5; (c) CLM 5.0; (d) JULES-ES; (e) JSBACH; (f) LPJ-GUESS; (g) ORCHIDEE; (h) ISBA-CTRIP. The areas characterized by the same type of LSMs and MODIS (Figure 2a) are coloured. These common areas are called agreement regions. Index values: (purple) evergreen; (green) single season with summer LAI peak; (cyan) single growing season with summer dormancy; (orange) two growing seasons type. White regions are for disagreement areas. Above this selection, areas of agreement between satellite products are shaded with a different hatching pattern: MODIS and LAI3g (Figure 2d) slash hatching (/); MODIS and CGLS (Figure 2g) backslash hatching (\); MODIS, CGLS, and LAI3g crossed hatching (X).

Table 3. The fraction of land grid-cell in agreement with MODIS for LAI3g, CGLS, each land surface model, and multi-model ensemble mean (MME) in growing season start (GSS) and growing season end (GSE) timings. Values are reported in percentages for global, North Hemisphere (NH), and South Hemisphere (SH). Green shaded areas in Figures 4 and 5. The last row reports the fraction of land grid-cell in agreement with MODIS for each land surface model in growing season length. The values in brackets give the percentage of global, NH, and SH with a maximum difference of one month.

	LAI3g	CGLS	CLM 4.5	CLM 5.0	JULES-ES	JSBACH	LPJ-GUESS	ORCHIDEE	ISBA-CTRIP	MME
GSS	37.9	43.7	14.6	6.5	2.7	16.7	17.3	19.1	16.0	15.8
	(74.0)	(74.2)	(36.1)	(21.6)	(15.0)	(41.3)	(45.1)	(44.2)	(43.1)	(31.3)
GSE	45.1	34.0	23.5	6.5	4.9	12.6	20.6	26.4	19.9	25.1
	(80.1)	(70.5)	(44.8)	(30.0)	(14.7)	(38.5)	(60.5)	(63.8)	(62.5)	(44.6)
GSS NH	39.2	39.7	15.1	5.7	3.2	17.1	17.6	18.9	16.5	16.3
	(76.4)	(72.6)	(37.8)	(20.6)	(17.7)	(43.4)	(44.5)	(43.9)	(44.5)	(32.9)
GSE NH	47.3	29.6	25.4	6.5	6.0	14.4	19.9	29.9	21.7	27.4
	(83.6)	(68.1)	(46.9)	(31.2)	(17.7)	(43.5)	(61.0)	(69.2)	(68.1)	(47.5)
GSS SH	32.6	60.7	12.6	10.0	0.7	14.8	16.0	19.8	14.1	13.3
	(63.8)	(81.0)	(28.8)	(25.9)	(3.4)	(32.4)	(47.7)	(45.6)	(36.8)	(24.8)
GSE SH	35.9	52.7	15.0	6.6	0.2	5.1	23.4	11.6	12.0	15.5
	(64.8)	(80.9)	(35.7)	(24.9)	(1.8)	(17.5)	(58.2)	(40.4)	(38.7)	(32.1)
Length	26.7	28.9	7.9	10.3	9.5	17.4	9.2	12.7	13.7	16.7
	(60.6)	(54.2)	(28.4)	(33.7)	(27.0)	(45.3)	(31.9)	(39.1)	(35.2)	(35.2)

Table 4. Average difference between MODIS and LAI3g, CGLS, each land surface model, and multi-model ensemble mean (MME) in Growing Season Start (GSS) and Growing Season End (GSE) timings. Values are reported in months for global, North Hemisphere (NH), and South Hemisphere (SH). Positive values stand for later timings and negative values correspond to earlier timings.

	LAI3g	CGLS	CLM 4.5	CLM 5.0	JULES-ES	JSBACH	LPJ-GUESS	ORCHIDEE	ISBA-CTRIP	MME
GSS	0.25	0.30	0.54	0.81	1.23	0.35	0.37	0.64	0.44	0.56
GSE	-0.16	-0.31	-0.30	-1.15	-2.26	-0.28	0.14	-0.10	-0.30	-0.49
GSS NH	0.41	0.38	0.95	1.47	1.42	0.95	0.44	0.69	0.97	0.98
GSE NH	-0.31	-0.48	-0.47	-1.69	-2.58	-0.69	0.10	-0.29	-0.59	-0.84
GSS SH	-0.42	-0.01	-1.31	-2.18	-1.80	-2.48	0.11	0.41	-2.04	-1.23
GSE SH	0.46	0.39	0.42	1.34	2.94	1.66	0.34	0.82	1.03	1.00

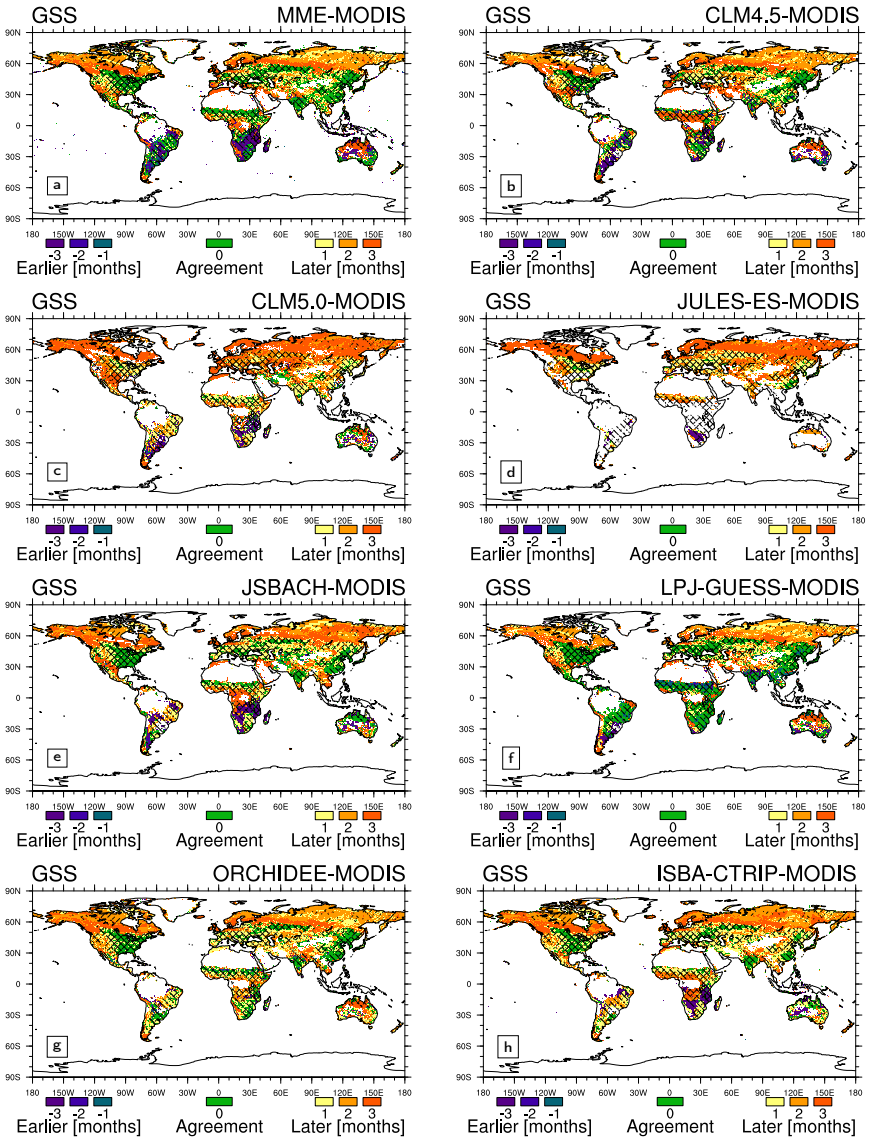

Figure 4. Global climatological (averaged over 2000-2011) differences in growing season start timings (GSS) between (a) Multi-Model Ensemble mean (MME); (b) CLM 4.5; (c) CLM 5.0; (d) JULES-ES; (e) JSBACH; (f) LPJ-GUESS; (g) ORCHIDEE; (h) ISBA-CTRIP and MODIS (Figure 2b). The green regions represent areas of agreement between MODIS and LSMs. Yellow-red colors correspond to areas where models timings are later compared to MODIS, while blue-violet colors correspond to areas where models timings are earlier compared to MODIS. Regions where GSS timings are not computed, such as non-vegetated and evergreen areas, are in white. Above this selection, areas of agreement between satellite products are shaded with a different hatching pattern: MODIS and LAI3g (Figure 2d) slash hatching (/); MODIS and CGLS (Figure 2g) backslash hatching (\); MODIS, CGLS, and LAI3g crossed hatching (X). Note that the GSS in the TGS regions corresponds to the GSS of the first growing season cycle.

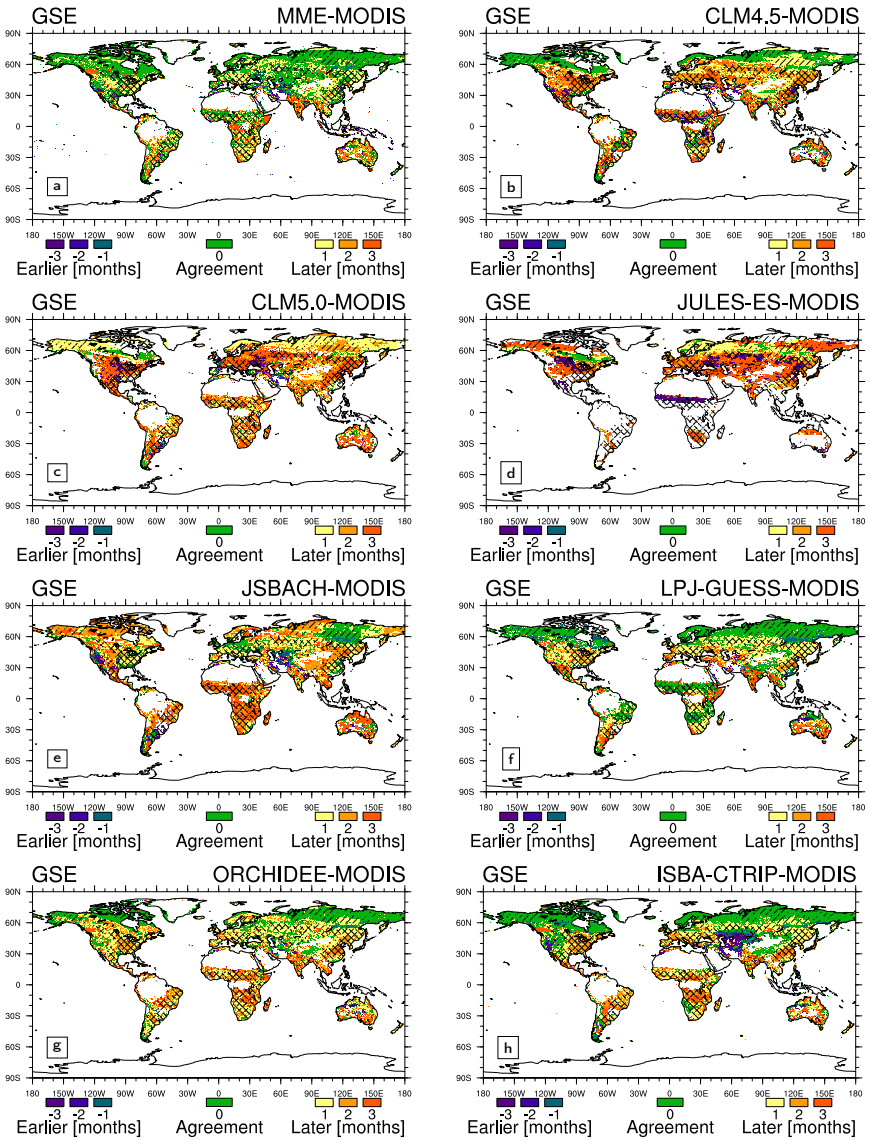

Figure 5. As Figure 4 but for growing season end (GSE) timings. Note that the GSE in the TGS regions corresponds to the GSE of the second growing season cycle.

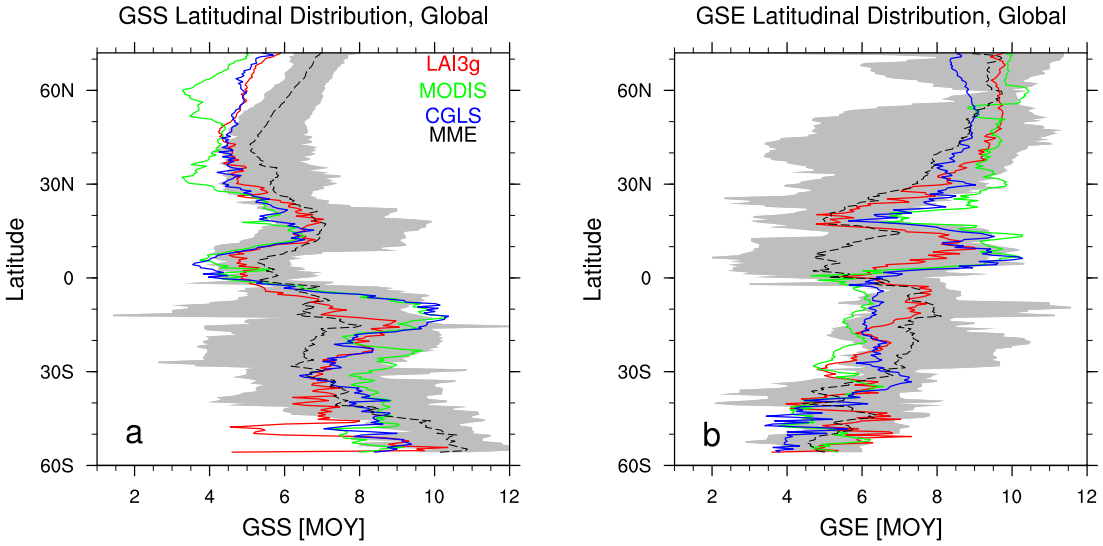

Figure 6. Zonal mean (a) growing season start (GSS) and (b) growing season end (GSE) timings for LAI3g (red lines), MODIS (green lines), CGLS (blue lines), and multi-model ensemble mean (black dashed line). The grey regions show the multi-model ensemble spread. Values are reported as month of the year (MOY), and the latitudinal coverage goes from 56°S to 75°N, which is the range covered by CGLS.

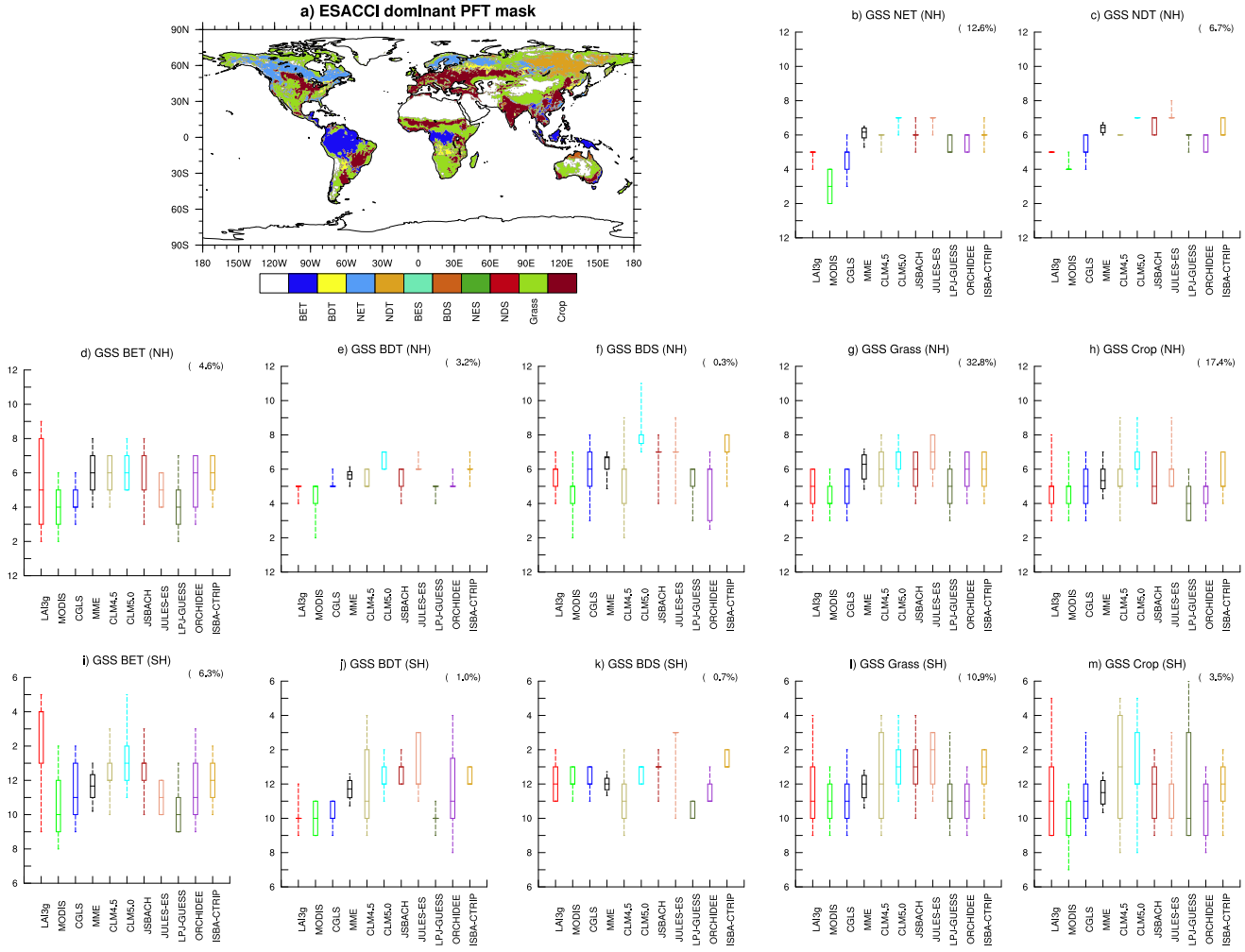

Figure 7. (a) Global distribution of the main land cover types for the 2000-2011 period based on ESA-CCI data (Li et al., 2018). Comparison in growing season start (GSS) timings between satellite products (LAI3g, red; MODIS, green; CGLS, blue) and land surface models (LSMs: MME, black; CLM4.5, dust; CLM5.0, cyan; JSBACH, dark red; JULES, pink; LPJ-GUESS, dark green; ORCHIDEE, purple; ISBA-CTRIP, dark yellow) in (b) Needle-leaf Evergreen Trees (NET) in the North Hemisphere (NH); (c) Needle-leaf Deciduous Trees (NDT) in the NH; Broad-leaf Evergreen Trees (BET) in the (d) NH and (i) SH; Broad-leaf Deciduous Trees (BDT) in the (e) NH and (j) SH; Broad-leaf Deciduous Shrubs (BDS) in the (f) NH and (k) SH; grass-covered areas (Grass) in the (g) NH and (l) SH; crop-covered areas (Crop) in the (h) NH and SH (m). Note that no area is dominated by Broad-leaf Evergreen Shrubs (BES), Needle-leaf Evergreen Shrubs (NES), or Needle-leaf Deciduous Shrubs (NDS) biome. The boxplots represent the median, 25/75th percentile, and 10/90th percentile of the distribution of grid-points belonging to each biome illustrated in panel a. Each panel shows in parenthesis the percentage of global vegetated area covered by each biome. Note that the y-axis is different in NH and SH panels, but, in both cases, the summer season is central along the axis.

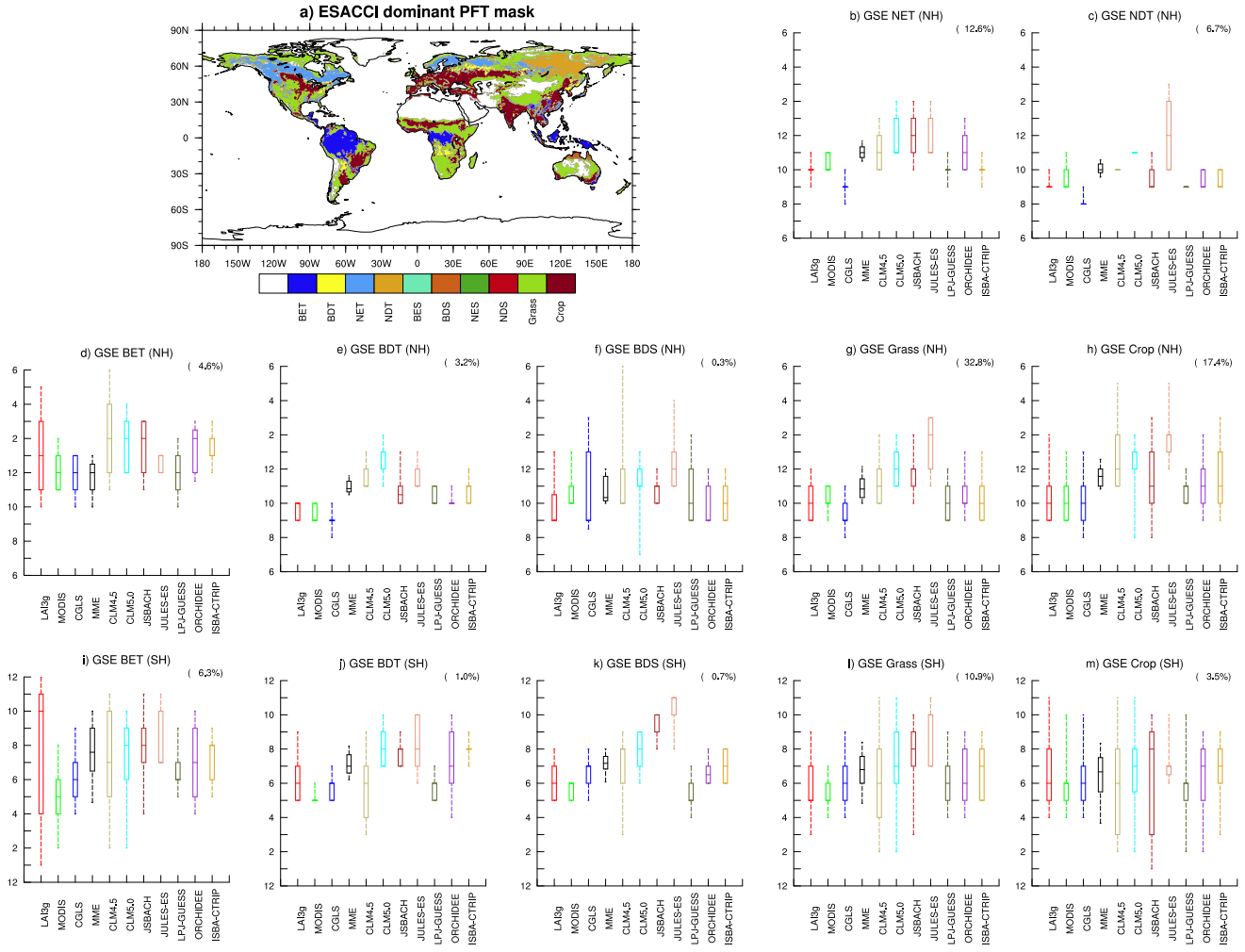

Figure 8. As Figure 7 but for growing season end (GSE) timings. In this case, the winter season is central along the y-axis.