Understanding the land carbon cycle with space data: current status and prospects

Citation for published version:

Digital Object Identifier (DOI):
10.1007/s10712-019-09506-2

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Surveys in geophysics

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breached copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Understanding the land carbon cycle with space data: current status and prospects

Jean-François Exbrayat*, A. Anthony Bloom², Nuno Carvalhais³, Rico Fischer⁴, Andreas Huth⁴, Natasha MacBean⁵, Mathew Williams¹

¹ National Centre for Earth Observation and School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FF, UK
² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
³ Max Planck Institute for Biogeochemistry, Department Biogeochemical Integration, Hans-Knoell-Str. 10, 07745 Jena, Germany
⁴ Helmholtz Centre for Environmental Research – UFZ, Department of Ecological Modelling, Permoserstrasse 15, 04318 Leipzig, Germany
⁵ School of Natural Resources and the Environment, University of Arizona, 1064 E Lowell St., 85721, Tucson, AZ, USA

*correspondence to: j.exbrayat@ed.ac.uk / ORCID ID: 0000-0002-3671-8626

University of Edinburgh
Crew Building
Alexander Crum Brown Road
Edinburgh EH9 3FF
United Kingdom
Abstract

Our understanding of the terrestrial carbon cycle has been greatly enhanced since satellite observations of the land surface started. The advantage of remote-sensing is that it provides wall-to-wall observations including in regions where in situ monitoring is challenging. This paper reviews how satellite observations of the biosphere have helped improve our understanding of the terrestrial carbon cycle. First, it details how remotely-sensed information of the land surface has provided new means to monitor vegetation dynamics, estimate carbon fluxes and stocks. Second, we present examples of studies which have used satellite products to evaluate and improve simulations from global vegetation models. Third, we focus on model-data integration approaches ranging from bottom-up extrapolation of single variables to carbon cycle data assimilation system able to ingest multiple types of observations. Finally, we present an overview of upcoming satellite missions which are likely to further improve our understanding of the terrestrial carbon cycle and its response to climate change and extremes.
1. Introduction

Terrestrial ecosystems help offset climate change by absorbing 25–30% of anthropogenic emissions of carbon dioxide (CO$_2$) each year (Canadell et al., 2007; Le Quéré et al., 2018). The global land carbon sink is calculated as the residual between reported fossil-fuel emissions, measurements of the atmospheric CO$_2$ growth and constrained estimates of ocean carbon uptake (see e.g. Le Quéré et al. 2018 for more details). The spatiotemporal distribution of the land carbon sink is estimated using an ensemble of process-based terrestrial ecosystem models (TEMs; Sitch et al. 2015). However, these largely unconstrained models exhibit significant differences in the location, magnitude and sign of the land carbon balance. This lack of agreement leads to large uncertainties in Earth system models’ projections of the response of terrestrial ecosystems to future climate change (Friedlingstein et al. 2006, Arora et al. 2013). This is further complicated by the interplay of vegetation CO$_2$ uptake and emissions from land use and land cover change (Arneth et al. 2017) such as deforestation (van der Werf et al. 2009).

A key issue for TEMs has been the challenge of integrating global observations to calibrate process parameters. Model spread seems to emerge from the lack of understanding of processes that control carbon allocation, turnover and mortality (Friend et al. 2014). Although networks of eddy-covariance towers (e.g. FLUXNET, Baldocchi et al. 2001) provide useful information to improve models (Williams et al. 2009, Kuppel et al. 2014), their distribution is highly skewed toward temperate regions of the northern hemisphere which challenges their applicability to other ecosystems. However, in recent years the multiplication of continuous Earth Observation (EO) has allowed the production of observational datasets relevant to the biosphere. While satellites do not measure carbon stocks and fluxes directly, they provide covariates for the extrapolation of in situ data to global gridded products related to ecosystem carbon fluxes (e.g. Jung et al. 2009) and biomass stocks (e.g. Saatchi et al. 2011; Baccini et
al. 2012; Avitabile et al. 2016) through machine-learning. Integrating EO to constrain process-based models has led to breakthrough in our understanding of the terrestrial carbon cycle, allowing a robust attribution of the increasing atmospheric CO$_2$ amplitude to the response of high latitude productivity (Forkel et al. 2016).

In this paper we review studies in which satellite-driven datasets have been used to improve our understanding of the terrestrial carbon cycle. We first focus on studies in which EO products have been used to monitor vegetation dynamics, carbon fluxes and stocks. Then, we illustrate how EO products can be used to evaluate TEMs, understand their biases and improve projections of the carbon cycle. Third, we present model-data integration approaches in which EO products are used to constrain TEMs using automated model-data fusion approaches. Finally, we review the foreseeable improvements future satellite missions are likely to generate.

2. Earth Observation to monitor vegetation dynamics, carbon fluxes and stocks

Satellite observations of the land surface allow a continuous monitoring of vegetation dynamics through the calculation of spectral indices (Myneni et al. 1995). One of the most common metrics is the Normalized Difference Vegetation Index (NDVI). It is calculated such as NDVI = (NIR+RED)/(NIR-RED) where RED and NIR are the spectral reflectance in the visible red and near-infrared region of the photosynthetically active radiation spectrum, respectively. Because chlorophyll strongly absorbs visible light during photosynthesis, active canopies have higher NDVI values.

NDVI has been used in numerous studies to characterize the response of plant phenology and productivity to climate trends and interannual variability. Bimonthly NDVI from the Advanced Very High Resolution Radiometer (AVHRR) spans the period since July 1981 until present. This long-term dataset has allowed describing a lengthening of the growing
season in temperate regions of the Northern Hemisphere due to an earlier disappearance of
snow in warming conditions (Myneni et al. 1997). Using the third generation of the AVHRR-
based Global Inventory Modelling and Mapping System (GIMMS) NDVI dataset (Pinzon
and Tucker, 2014), Buitenwerf et al. (2015) detected significant changes in phenological
cycles for more than half of the land surface between 1981 and 2012. This long-term dataset
has also been used to identify dominant climate modes driving the inter-annual variability in
the start of growing season across North America (Dannenberg et al. 2018).

Higher resolution NDVI data from the Moderate resolution Imaging Spectroradiometer
(MODIS) sensor (Huete et al. 2002) have been provided by the National Aeronautics and
Space Administration (NASA) since the year 2000. These data have been used to attribute the
anomalous greening of the Northern Hemisphere land surface in 2015 to a strong state of the
Pacific Decadal Oscillation (Bastos et al. 2017). The MODIS archive also includes an
Enhanced Vegetation Index (EVI) which performs well better in high biomass regions (Huete
et al. 2002). It has been used to describe leaf growth across the Amazon basin during the dry
season, which promotes primary productivity in sunnier conditions (Huete et al. 2006). EVI
has also been used to describe the phenological response of Australian ecosystems to climate
IAV (Broich et al. 2014), in particular the continental greening that followed the extremely
wet 2010/11 La Niña episode (Fasullo et al. 2013).

NDVI and EVI are useful proxies for vegetation activity but they are not biophysical
variables directly relatable to TEMs. However, two key state variables for these models can
also be derived from EO: the Fraction of Absorbed Photosynthetically Active Radiation
(FAPAR) and Leaf Area Index (LAI), the one-sided area of leaves per units of ground.
FAPAR and LAI are related to NDVI (Zhu et al. 2013) but can also be retrieved using
physically-based (Knyazhikin et al. 1998) or machine-learning (Baret et al. 2013) methods.
FAPAR is a key driver for light use efficiency models of primary productivity (e.g. Potter et al. 1993; Field et al. 1995; Prince and Goward 1995; Knorr 2000). Global MODIS-based FAPAR (Myneni et al. 2002) is used to produce high resolution (~1km) estimates of gross and net primary productivity across the global land surface (Running et al. 2004). This dataset has provided insights in a possible reduction of global net primary productivity because of a drying in the Southern Hemisphere (Zhao and Running 2010). It helped characterized the influence of the El Niño – Southern Oscillation on regional and global NPP (Bastos et al. 2013) and the impact of recent European heatwaves on productivity (Bastos et al. 2014).

LAI represents the physiologically active part of the vegetation which interacts with the atmosphere and is a key state variable for land surface and ecosystem models (Kala et al. 2014). Figure 1 presents mean annual LAI for the year 2015 derived from the European Space Agency (ESA)’s Proba-V satellite (Baret et al. 2013) as part of the European Union’s Copernicus programme and from NASA’s MODIS sensor. There is a good agreement in the spatial distribution of LAI between these datasets but the MODIS products reports higher LAI values for tropical which may lead to non-negligible differences in the calculation of energy, water and carbon fluxes (Kala et al. 2014). MODIS LAI data has been helpful to understand the seasonality of the Amazon (Myneni et al. 2007). Zhu et al. (2013) used the relationship between AVHHR NDVI and MODIS FAPAR and LAI products to create the GIMMS FAPAR3g and LAI3g dataset extending back to the 1980s. This long-term dataset exhibits a greening trend, i.e. an increase in LAI during the growing season (Zhu et al. 2016).

A recent advance in remote sensing has been the production of Solar-Induced Fluorescence (SIF) retrievals (Frankenberg et al. 2011). SIF is directly related to plant photosynthetic activity; therefore, SIF data provide a more direct measure of gross C uptake than reflectance-based indicators like NDVI or FAPAR (Porcar-Castell et al. 2014). There has
been an increase in the availability of global SIF products derived from space-borne instruments like GOSAT (Frankenberg et al. 2011; Guanter et al. 2012), GOME-2 (Joiner et al. 2013; Köhler et al., 2015), SCIAMACHY (Joiner et al. 2016) and OCO-2 (Sun et al., 2018). More details about SIF is provided in section 3.2.3 of the review by Scholze et al. (2017).

Additionally to plant productivity, satellite datasets have been used for over three decades to monitor fire, a fundamental component of the terrestrial carbon cycle which accounts for a large degree of the inter-annual variability of the terrestrial land sink (van der Werf et al., 2010; Le Quéré et al., 2018). Observations of decadal trends in burned area (Flannigan et al., 1986; Giglio et al., 2013; Andela et al., 2017) have been used to establish the role of fires as a key component of the long-term C balance evolution (Le Quéré et al., 2018; Arora & Melton, 2018, amongst others). A range of satellite-based observations of fire have also radically advanced insight into continental-scale fire characteristics and processes, including understory fires (Morton et al., 2013); fire radiative power (Freeborn et al., 2014), and interactions between fire and species distribution (Rogers et al., 2015). Archibald et al. (2013) identified 5 dominant types of fire regimes using remotely-sensed observation of fire frequency, radiative power and size. Active fire detection also offers a crucial constraint on land C fire losses, with current observing system such as VIIRS offering new possibilities to detect boreal fires (Waigl et al. 2017).

Beyond monitoring vegetation dynamics and inferring land-atmosphere fluxes, satellite observations have allowed the creation of high resolution maps of above-ground biomass (AGB) covering large regions such as the pantropics (e.g. Saatchi et al., 2011; Baccini et al. 2012). Pantropical maps were created using allometric equations (e.g. Chave et al. 2014) relating tree height measured by NASA’s Geoscience Laser Altimeter System sensor onboard the Ice, Cloud, and land Elevation Satellite (Zwally et al. 2002) to AGB. These wall-to-wall
maps allow a first-order approximation of remotely-sensed deforestation (e.g. Hansen et al. 2013) on AGB stocks in the tropics (Harris et al., 2012; Baccini et al., 2012). Recently, maps by Saatchi et al. (2011) and Baccini et al. (2012) have been fused with additional in situ measurements to create a third map, currently considered as the state-of-the-art (Avitabile et al. 2016).

Thurner et al. (2014) created a map for the northern boreal and temperate forests paper using retrievals of Growing Stock Volume from Envisat Advanced Synthetic Aperture Radar (Santoro et al. 2011, 2015), databased information about wood density (Chave et al. 2009) and allometric equations. Thurner et al. (2016) used this map in combination with MODIS NPP to evaluate gradients in turnover dynamics across these regions.

While these previous studies relied on a single AGB map to produce estimates of gross emissions from deforestation, more recent studies have produced annually-resolved AGB maps which allow tracking the counteracting impact of regrowth and derive net changes of biomass globally. For example, Liu et al. (2015) used the correlation between Vegetation Optical Depth and AGB from Saatchi et al. (2011) to produce annual biomass maps for 1993–2012 at a 0.25° spatial resolution. They concluded to a loss of global AGB driven by a loss of tropical forests partially compensated by gains over boreal, temperate and savannah regions. More recently, Brandt et al. (2018) used a similar approach to describe a net carbon loss across sub-Saharan Africa for the period 2010–2016. Baccini et al. (2017) also concluded that the tropics are a net source of atmospheric CO₂ based on annual maps they constructed from 2003–2014.

EO is useful to identify land cover change (e.g. Hansen et al. 2013) and, by extension, intact forest landscapes (Potapov et al., 2008). Potential AGB maps, representative of the hypothetical undisturbed landscape, can be reconstructed using the relationship between AGB (Saatchi et al. 2011; Baccini et al. 2012; Liu et al. 2015) and climate (New et al. 2002).
in these intact regions. This approach attributes ~1.5% of the recent increase in atmospheric 
CO₂ to the Amazonian deforestation (Exbrayat and Williams 2015) while climate change has 
reduced the capacity of these forests to recover (Exbrayat et al. 2017).

International policy efforts such as the Paris agreement on climate change and the Bonn 
challenge for forest restoration have raised the interest of countries to produce country-scale 
maps for monitoring and reporting. For example Rodriguez-Veida et al. (2016) used local 
information from the Mexican forestry. Similarly, Xu et al. (2017) produced a biomass map 
for the Democratic Republic of Congo using additional data which were not available to 
pantropical maps. Both studies presented measurable increase in mapping quality and 
uncertainty quantification.

3. Evaluating terrestrial ecosystem models

Land surface models are key components of Earth system models that simulate energy and 
mass transfers between the land and the atmosphere, hence, these are key components in the 
prediction of climate variations from short to long time scales (Pitman 2003). EO of the 
biosphere provides unprecedented means to evaluate vegetation dynamics, carbon fluxes and 
biomass stocks simulated by land surface models in a temporally and spatially-explicit 
manner. The evaluation strategies have been largely focusing on aspects related to: the timing 
of seasonal vegetation development, and long term trends in vegetation greenness; the 
seasonal and spatial variations in photosynthesis patterns; the spatial variations in plant 
carbon stocks; and EO-derived estimates of carbon turnover times on land.

The representation of phenology in land surface models is a major source of uncertainty for 
the calculation of energy, water and carbon fluxes (Kala et al. 2014). Many studies have 
focused on in-situ evaluation of modelled LAI (e.g. Richardson et al. 2012; Migliavacca et al. 
2012) but long-term EO-derived products have also been used. Zhu et al. (2013) described a
systematic overestimation of LAI by 18 Earth system models compared to the GIMMS LAI3g dataset. This was accompanied by a shift toward earlier peak in LAI in boreal regions. Similarly, Anav et al. (2013) evaluated models participating in the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) which underpinned the fifth Assessment Report of the Intergovernmental Panel on Climate Change. They described a tendency for CMIP5 models to overestimate LAI, although most models captured LAI trends. The poor performance of models to represent phenology has led several intercomparison projects to impose EO-derived LAI dataset to all participating models (Huntzinger et al. 2013; Haarsma et al. 2016). EO-derived products have also been used to assess vegetation productivity simulated by ecosystem models. Kolby-Smith et al. (2015) created a long-term NPP dataset based on MODIS NPP algorithm (Running et al. 2004; Zhao and Running 2010) driven by long-term GIMMS FAPAR3g (Zhu et al. 2013). They compared this new dataset with five CMIP5 models which exhibited a much stronger trend of increasing NPP than the EO-based product. They concluded that models’ sensitivity to increasing atmospheric CO$_2$ was too high, probably owing to the lack of representation of nutrient limitation on productivity. Ito et al. (2017) showed that spatial and seasonal variations of GPP simulated by eight ecosystem models were in agreement with the MODIS GPP product. However, they also showed that models failed to simulate GPP anomalies in response to extreme events such as the 1997–1998 El Niño or the eruption of Mount Pinatubo in 1991. Slevin et al. (2017) identified an underestimation of GPP in the tropics when comparing the Joint UK Land Environment Simulator (JULES; Clark et al. 2011) with EO-derived GPP products. EO-derived products of GPP and NPP now allow skill-based ensemble averaging studies to be applied to ecosystem models. These post-processing procedures have been used in atmospheric sciences for many decades (e.g. Krishnamurti et al. 1999) and can be applied to
ecosystem models in a spatially-explicit way. Schwalm et al. (2015) applied the Reliability
Ensemble Averaging method (Giorgi and Mearns 2012) constrained by FLUXNET MTE-
GPP and biomass estimates to ten ecosystem models participating to the MsTMIP
(Huntzinger et al. 2013). Exbrayat et al. (2018) used a similar method to constrain projections
of 21st century change in NPP predicted by 30 simulations from the ISIMIP ensemble (Friend
et al. 2014; Warszawski et al. 2014). They showed that the uncertainty in global change in
NPP could be reduced by two-third using a skill-based ensemble averaging whilst gaining
confidence on the sign of the change for more than 80% of the global land surface.

A recent emphasis has been put on the need to move beyond the separate evaluation of pools
and fluxes by terrestrial land models. For many years, global models have been initialized
using a spin-up procedure from which biomass stocks would emerge as a result of input
fluxes and turnover times at steady-state (Exbrayat et al. 2014). However, models perform
poorly to simulate vegetation carbon stocks in agreement with observation-based products.

For example, Figure 2 presents a comparison of the recent pantropical biomass map from
Avitabile et al. (2016) with models from the ISIMIP ensemble (Friend et al. 2014;
Warszawski et al. 2014). There is a large uncertainty represented by the inter-model spread
while they tend to overestimate biomass stocks in regions of the Americas and Africa located
north of 10°N and south 15°S. Friend et al. (2014) clearly demonstrated that the highest
disagreement between models resides in the internally modelled residence times of carbon
which can be inferred from the ratio between observable fluxes and stocks (Friend et al.
2014; Sierra et al. 2017). Under the same future changes in environmental and climate
conditions models alternatively predict longer or shorter turnover times of carbon in
vegetation. This mismatch reflects a disagreement in the sign of the terrestrial carbon cycle
feedback on future changes in climate and atmospheric CO₂. Current EO-based estimates
suggest a pervasive control of hydrology on whole ecosystem apparent turnover times of
carbon, which are not captured by current Earth system models (Carvalhais et al. 2014). In particular, the spatial patterns of vegetation C turnover times in forests suggest strong climatic controls in mortality patterns associated to drought and heat, but also extreme winter cold temperatures which could expand plant mortality, or reduce it, via reductions in herbivore activity (Thurner et al. 2016). An across model comparison also revealed that most state of the art global vegetation models do not reflect the direct effects of climate induced mortality (Thurner et al. 2017), emphasizing the present challenge of understanding mortality induced by climate extreme (Hartmann et al. 2015). Furthermore, recent results have also emphasize the role of land use, in addition to land cover, as a substantial factor for an overall reduction in carbon residence times in terrestrial vegetation (Erb et al. 2016). In general, all of these works have been emphasizing the mismatch between model and observation-derived ecosystem dynamics, and hypothesizing on the missing, or misrepresented, underlying mechanisms that drive carbon dynamics. Downstream, a full body of research has also been focusing on formally integrating these observations into model-data-assimilation frameworks to maximize information transfer from observation to models.

4. Model-data integration

While EO can be used to benchmark models, they do not measure all aspects of the terrestrial carbon cycle. Therefore, models are needed to fill gaps but benchmarking studies reviewed in section 3 have generally pointed to poor performances and systematic biases in forward models. Model-data integration aims to synergize data and models through an interactive process. There exists multiple forms of model-data integration in which EO has been used to provide global covariates for the extrapolation of in situ data (e.g. Jung et al. 2011), retrieve state variables such as LAI from reflectance through a complex radiative transfer scheme
(Lewis et al. 2012), constrain productivity and phenology model parameters in terrestrial Carbon Cycle Data Assimilation Systems (CCDAS; e.g. Knorr et al. 2010) or even provide initial conditions to detailed forest models (e.g. Rödig et al. 2017, 2018). The global coverage of EO has allowed the development of a range of “bottom-up” approaches to upscale data-driven in situ models to spatially explicit gridded estimates. One major development in this area has relied on training machine-learning algorithm to reproduce local ecosystem fluxes as a function of climate and vegetation properties available from EO (e.g. NDVI, FAPAR). The first example of this approach was reported by Papale and Valentini (2003) who used and Artificial Neural Network trained at 16 European sites to generate continental maps of forest productivity using information about land cover, seasonal temperatures and maximum NDVI from AVHRR. Jung et al. (2009) developed the Model Tree Ensemble (MTE) algorithm to create gridded products of GPP and latent heat fluxes. They first demonstrated the potential for their approach in a synthetic example using LPJmL simulations as training data. Using this method driven by fluxes measured at several FLUXNET site has allowed the first data-driven description of the distribution of global GPP between biomes and an attribution of dominant regional climate drivers for the period 1998-2005 (Beer et al. 2010). Building on this approach, monthly gridded estimates of GPP extending back to 1982 have been created using GIMMS FAPAR data (Jung et al. 2011). They identified semiarid and sub-humid as experiencing a high inter-annual variability in productivity due to rainfall variations. These data, used in combination with process-based models, further helped identify the response of savannahs ecosystems to ENSO as the dominant driver of the variability in the land carbon sink (Poulter et al. 2014). Multiple machine-learning approaches, relying on various algorithms, have been compared by Tramontana et al. (2016). The evaluation of different approaches was performed as part of the FLUXCOM initiative (http://www.fluxcom.org). The conclusion of this comparison was
that machine-learning approaches were skilled at reproducing heat and productivity fluxes but may be biased to predict net ecosystem carbon fluxes due to the lack of feedback representation and knowledge of historical disturbance regimes. Nevertheless, bottom-up and top-down approaches estimates of GPP are in good agreement. Figure 3 presents mean annual GPP estimated by FLUXCOM during 2000-2013 and compares it to the MODIS GPP product. Both approaches are partly driven by the same estimates of MODIS FAPAR which yields a high spatial correlation ($r = 0.80; p << 0.001$). FLUXCOM indicates a mean annual GPP of 124.7 Pg C y$^{-1}$ while MODIS estimates a 136.4 Pg C y$^{-1}$, a 9% relative difference. Beyond empirical approaches, the increasing availability of EO has played a key role in the development of more “top-down” terrestrial Carbon Cycle Data Assimilation Systems (CCDAS). Unlike “bottom-up” approaches which consist in extrapolating in situ models, “top-down” CCDAS are centred around using EO, including non-carbon variables (Scholze et al. 2017), to constrain process-based models in a spatially explicit way. One of the first CCDAS was based on the Bethy ecosystem model (Knorr 2000). It has been incrementally improved with additional processes such as dynamic phenology (Knorr et al. 2010). An interesting aspect of CCDAS studies has been to focus on the development and comparison of inversion strategies (e.g. Ziehn et al. 2012) to reduce the computational cost of the assimilation. We refer the reader to a detailed review of the evolution of CCDAS by Kaminski et al. (2013) for more information about this particular framework. One strategy introduced by Peylin et al. (2016) has also been to use a stepwise approach to first constrain parameters related to phenology in the ORCHIDEE model before assimilating fluxes in a subsequent step. We focus the following paragraphs on example of new knowledge derived from CCDAS applications, and point the readers to the recent review of Scholze et al. (2017) for more technical information about the type of assimilation techniques and EO used. The advantage
of EO is that model-data integration is performed globally and CCDAS framework such as CARDAMOM (Bloom et al. 2016) and CASA-GFED (van der Werf et al. 2010) provide compelling methodologies for reconciling land-surface and atmospheric constraints on the terrestrial C balance, through which major uncertainties in process representation such as phenology (e.g. Stöckli et al. 2011, Forkel et al. 2014), allocation (Bloom et al. 2016), combustion and emission dynamics (Bloom et al., 2015; Worden et al., 2017) can be explicitly constrained.

Phenology is a poorly represented process and assimilating reflectance-based EO of NDVI, FAPAR and LAI has allowed m development and validation of new global models. For example, Knorr et al. (2010) assimilated daily FAPAR at seven sites in a generic phenological model. Quaife et al. (2008) demonstrated that assimilating reflectance from the MODIS sensor in the Data-Assimilation Linked Ecosystem Carbon model (DALEC; Williams et al. 2005) led to an improvement of simulated carbon fluxes at a coniferous forest site in Oregon, US. Stöckli et al. (2011) assimilated 10 years of MODIS LAI and FAPAR data in a phenological model based on the Growing Season Index concept (GSI; Jolly et al. 2005). They identified used the constrained model to produce a 50-year re-analysis of LAI and FAPAR. Forkel et al. (2014) implemented a modified version of the GSI model in the LPJmL dynamic global vegetation model. They retrieved dominant controls of phenology by assimilating 30 years of GIMMS FAPAR, highlighting the codominant role of moisture stress on the variability in phenology (Forkel et al. 2015) which contrasts with classical temperature-based parameterizations. The importance of moisture availability was also found by MacBean et al. (2015) based on the biases in the temperature-driven phenology of the ORCHIDEE model. While these previous studies have relied on plant functional types, Caldararu et al. (2014) successfully fitted a phenological model to pixel-wise MODIS LAI
data. Their approach based on carbon optimality concluded that leaf age was also a limiting factor for phenology in evergreen tropical regions.

EO of fire has also been used to constrain emission estimates. Top-down estimates of surface CO emissions amount to a robust constraint on continental-scale fire C emissions:

measurements of atmospheric CO – including those from ESA’s IASI instrument, NASA’s MOPITT, TES and AIRS instruments – have been used to constrain atmospheric chemistry and transport models in data assimilation frameworks (Jiang et al., 2015; Gonzi et al., 2011; Krol et al., 2013; Kopacz et al., 2010, amongst many others). Subsequently, estimates of fire CO:CO₂ ratio (Andreae & Merlet, 2001; Akagi et al., 2011) have been used to quantify continental-scale fire C fluxes (Gatti et al., 2014; Bowman et al., 2017). However, CO:CO₂ have been identified as a potential source of error in extreme fire events (Krol et al., 2013; Bloom et al., 2015), where CO:CO₂ values and their uncertainty characteristics are poorly known. Overall, estimates of fire emissions derived from an array of bottom-up and top-down constraints are invaluable for obtaining a spatially-explicit estimates of fire C fluxes (Bowman et al., 2017; Liu et al., 2017). Ultimately, satellite-based estimation of fire C emissions, characteristics, and trends are key to advance process-level understanding of fires as a dynamic component of the Earth System.

Although a number of satellite-derived products (NDVI, LAI, FAPAR, biomass and XCO₂) have been used to constrain both modelled leaf phenology, biomass, and net CO₂ fluxes (Kaminski et al., 2013; Forkel et al., 2015; MacBean et al., 2015; Peylin et al., 2016; Bloom et al., 2016), these data only provide indirect information on gross C uptake. For some vegetation types, even ground-based net CO₂ fluxes derived from eddy covariance towers only provide limited capacity in constraining the gross C fluxes (Kuppel et al., 2014).

Consequently, SIF products have been used in a variety of ways to assess and improve land surface model (LSM) simulations: i) to benchmark GPP and SIF temporal dynamics
simulated for a range of sites (Lee et al., 2015; Thum et al., 2017); ii) to optimize global-scale
GPP estimates from a LSM inter-comparison *a posteriori* (Parazoo et al., 2014); and to
optimize parameters of both fluorescence and photosynthesis models at local to global scales
(Zhang et al., 2014; MacBean et al., 2018; Norton et al., 2018; Norton et al., in review). The
latter studies have demonstrated considerable potential for SIF to constrain both in situ and
global scale GPP simulations. MacBean et al. (2018) and Norton et al. (in review) show
strong reductions in both the spatio-temporal misfit (increased correlation and decreased bias)
across vegetation types between modelled and observed GPP and SIF, and in the simulated
global-scale GPP uncertainty. The reduction in GPP uncertainty is a result of constraining
both fluorescence, photosynthesis, and phenology-related parameters. In many of the
abovementioned modelling studies, an explicit formulation of the relationship between
photosynthesis and fluorescence has been developed – largely based on the SCOPE (Soil
Canopy Observation Photochemistry and Energy Fluxes) model (van der Tol et al., 2009) –
and implemented within each respective LSM (Lee et al., 2014; Thum et al., 2017; Norton et
al., 2018; Norton et al., in review). However, SIF has been shown to be linearly correlated
with GPP at a range of spatial and temporal scales (Frankenberg et al., 2011; Guanter et al.,
2012; Joiner et al., 2014; Yang et al., 2015; Zhang et al., 2016; Wood et al., 2017; Yang et
al., 2017). This assumed linear relationship also allows a relatively simple and
straightforward means by which modelled GPP and SIF can be compared with, and
constrained by, remote sensing SIF estimates at large spatial scales (MacBean et al., 2018).
Continental-scale temporal variability of the terrestrial land sink can be robustly observed
through atmospheric CO₂ measurements from satellites - most notably from SCIAMACHY
instrument onboard ENVISAT, JAXA’s GOSAT and the NASA OCO-2 missions (Buchwitz
& Burrows, 2004; Yokota et al., 2009; Eldering et al., 2017). Terrestrial CO₂ fluxes can be
quantitatively retrieved through assimilation of these observations into inverse modeling
frameworks (Houweling et al., 2015; Deng et al., 2016), although we note that absolute CO\textsubscript{2} flux estimates are susceptible to a number of model and observation biases (Feng et al., 2016; Miller et al., 2017; Worden et al., 2017b, Basu et al., 2018). Notable insights into terrestrial C cycle processes from satellite-constrained estimates of land C fluxes include multi-year constraints on the Australian C balance (Detmers et al., 2015) regional constraints on the seasonal and inter-annual Amazon C fluxes (Parazoo et al., 2013; Bowman et al., 2017) and Indonesia fire C emissions during the 2015 ENSO (Heymann et al., 2017). Pan-tropical continental-scale estimates of inter-annual CO\textsubscript{2} flux variations by Liu et al., (2017) demonstrated the synergistic capacity of GOSAT and OCO-2 CO\textsubscript{2} measurements – along with ancillary constraints from solar-induced fluorescence – to disentangle the processes regulating the temporal variability of the terrestrial C sink.

While previous studies have focused on fluxes, reconciling stocks is key as these are responsible for uncertainty in residence times (Carvalhais et al., Friend et al. 2014). Bloom et al. (2016) have used pantropical biomass estimates from Saatchi et al. (2011) to constrain global retrievals of carbon allocation and residence times in the CARDAMOM framework. Figure 4 shows an updated version of these data based on the assimilation of biomass estimates from Avitabile et al. (2016) and MODIS LAI for the period 2000-2015. These simulations, limited to the area covered by Avitabile et al. (2016), provide EO constrained estimated of NPP (43.3 Pg C y\textsuperscript{-1}), Rh (40.2 Pg C y\textsuperscript{-1}), fire emissions (1.2 Pg C y\textsuperscript{-1}) and Net Biome Exchange (-1.7 Pg C y\textsuperscript{-1}, corresponding to a sink).

Furthermore, an accurate representation of stocks in ecosystem models is required to robustly estimate emissions related to land use change. Li et al. (2017) used estimates of current biomass compiled by Carvalhais et al. (2014) to constrain centennial emissions from land use change in an ensemble of nine models. While global numbers of cumulative emissions from land use change were similar between the unconstrained and constrained models, regional
differences appeared. For example, the data-constrained estimates yielded larger emissions from land use change in the tropics, and smaller in temperate regions, compared to the unconstrained estimates. In a more recent study Lienert and Joos (2018) also used biomass data from Carvalhais et al. (2014) to constrain emissions from land use change using alternative representation of emissions due to net and/or gross land use transitions.

Studies presented in previous paragraphs have focused on using EO data to constrain fluxes and state variables in conceptual models. However, remotely-sensed information of vegetation structure can be connected to highly detailed forest models to provide mechanistic estimates of forest biomass and productivity (Shugart et al. 2015, Knapp et al. 2018a, Knapp et al. 2018b). Forest structure is indeed an important element to describe the state of forests. Precise estimates of forest structure need to consider small-scale variations resulting from local disturbances, on the one hand, and require large-scale information on the state of the forest that can be detected by remote sensing, on the other hand (Rödig et al. 2017). Local forest models can simulate and analyse different kinds of local disturbances and thus small-scale changes in forest structures more accurately than global ecosystem models. Remote sensing has the potential to provide global high-resolution measurements of the structure of a forest (e.g. forest height by Lidar or interferometric Radar measurements).

As an example, Rödig et al. (2017) used remote sensing data with a resolution of 1km² (i.e. canopy height map derived from ICESat) to establish large-scale applications of a local forest gap model (i.e. FORMIND, Fischer et al. 2016). Forest gap models (Shugart et al. 2018) simulate forest succession at the individual tree level. The advantage of using a local forest model at the large scale is that it brings along information on many different forest attributes (e.g. productivity, carbon sequestration, water fluxes) in a very fine resolution. In combination with remote sensing, this enables the derivation of high-resolution maps of carbon stocks and fluxes – which was conducted for the whole Amazon (Rödig et al. 2017,
By this approach, it was possible to simulate each tree in the Amazon. Finally growth of more than 410 billion trees was analysed. According to this study, forests in the Amazon store high amounts of aboveground biomass (76 Gt of carbon) and are an important sink of 0.56 Gt C a\(^{-1}\) under current conditions (Rödig et al. 2018).

5. Outlook

EO has made an essential contribution to our understanding of the terrestrial carbon cycle since the 1980s. It ranges from the continuous monitoring of vegetation activity through NDVI, FAPAR and LAI to providing wall-to-wall constraints for model-based estimates of land-atmosphere carbon fluxes. Nowadays, multiple dedicated missions and services, such as the Sentinel satellites of the European Union’s Copernicus programme, provide almost real-time observations with a high level of quality. For example, the ESA TROPOMI instrument of Sentinel-5P provides CO measurements with an unprecedented spatiotemporal coverage (Bordsorff et al., 2018). In the next few years, multiple sensors will be launched to complement the existing constellation of Sentinels and provide coincident observations of several aspects of the biosphere (Stavros et al. 2017).

NASA’s Global Ecosystem Dynamics Investigation (GEDI) mission, a LIDAR system on-board the International Space Station (ISS), will provide a global coverage of canopy height and foliage vertical profiles. It will provide updated and more detailed structural information for integration with forest models (e.g. Rödig et al. 2017, 2018). ESA’s 7\(^{th}\) Earth Explorer BIOMASS (Le Toan et al. 2011) will provide repeated measurements of tropical biomass with a P-band synthetic aperture radar. In contrast with currently available biomass maps, temporally-resolved information losses from deforestation and gain from regrowth will be especially useful to reduce the uncertainty and correct bias in CCDAS framework (Smallman et al. 2017).
Recent studies have demonstrated the potential for satellite CO$_2$ observations to constrain land-atmosphere exchange (Liu et al. 2017). NASA’s upcoming OCO-3 will replace OCO-2 on-board the ISS, while the geostationary GeoCARB will focus on the Americas. Both systems, and the dedicated ESA’s Fluorescence Explorer (FLEX) mission, will provide measurements of SIF which has a great potential to constrain models of ecosystem productivity (MacBean et al. 2018), especially following the implementation of mechanistic representation of leaf physiology in CCDAS (Norton et al. 2018).

Finally, the carbon cycle is tightly linked to the energy and water cycles semi-arid areas in particular have been pointed as key ecosystems to understand the global land carbon sink (Poulter et al. 2014). Therefore, the development of new non-carbon EO and their assimilation in CCDAS frameworks plays a major role in simulating the carbon cycle (Scholze et al. 2017). NASA’s ECOSTRESS will measure evapotranspiration (Stavros et al. 2017) which will be used to obtain estimates of water-use efficiency, the ratio of productivity to evapotranspiration, which will be useful to drive process-based models of the biosphere.

Overall, the next few years will see an increase in the amount of observing systems with a ever-increasing spatial resolution and higher frequency. One of the key challenges for the modelling community is to build systems able to ingest all this information in an efficient way to provide high confidence retrievals of the terrestrial carbon cycle.
References


Science 356:1356–1362. doi: 10.1126/science.aal4108


land-use changes are possibly larger than assumed. Nat Geosci 10:79–84. doi: 10.1038/ngeo2882


Arora VK, Melton JR (2018) Reduction in global area burned and wildfire emissions since 


Zhu Z, Bi J, Pan Y, et al (2013) Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3G) for the period 1981 to 2. Remote Sens 5:927–948. doi: 10.3390/rs5020927


Figure 1. Mean LAI in 2015 according to a) Copernicus and b) NASA’s MODIS. Datasets were resampled at 0.25° spatial resolution for plotting purpose.
Figure 2. Biomass density from Avitabile (top) and comparison of zonal means simulations (bottom, as indicated) from 6 ISIMIP ecosystem models. Results indicate an overestimation of stocks by models in the Americas and Africa.
Figure 3. Mean annual GPP during 2000-2013 as reported by a) the FLUXCOM product and b) MODIS MOD17 GPP product. FLUXCOM corresponds to eddy-covariance data upscaled at 0.5° using machine-learning. MODIS product is based on a light use efficiency model and was regridded from 30” (~1km) to 0.5°.
Figure 4. CARDAMOM retrievals of land-atmosphere C fluxes, averaged over 2000-2015.

All fluxes are in g C m$^{-2}$. In d) NBE is calculated as $\text{NBE} = -\text{NPP} + R_h + \text{fire}$, hence $<0$ values correspond to a sink (in blue) and $>0$ correspond to a source of carbon (in red).