

Terrestrial gross primary production inferred from satellite fluorescence and vegetation models

NICHOLAS C. PARAZOO^{1,2}, KEVIN BOWMAN^{1,2}, JOSHUA B. FISHER¹, CHRISTIAN FRANKENBERG¹, DYLAN B. A. JONES^{2,3}, ALESSANDRO CESCATTI⁴, ÓSCAR PÉREZ-PRIEGO⁵, GEORG WOHLFAHRT⁶ and LEONARDO MONTAGNANI^{7,8}

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA, ²Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, CA 90095, USA, ³Department of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada, ⁴European Commission, Joint Research Center, Institute for Environment and Sustainability, Ispra, I-21027, Italy, ⁵Departamento de Física Aplicada, Universidad de Granada, 18071 Granada, Spain, ⁶Institut für Ökologie, Universität Innsbruck, Sternwartestr. 15, 6020 Innsbruck, Austria, ⁷Forest Services, Autonomous Province of Bolzano, Via Brennero 6, 39100 Bolzano, Italy, ⁸Faculty of Science and Technology, Free University of Bolzano, Piazza Università 5, 39100 Bolzano, Italy

Abstract

Determining the spatial and temporal distribution of terrestrial gross primary production (GPP) is a critical step in closing the Earth's carbon budget. Dynamical global vegetation models (DGVMs) provide mechanistic insight into GPP variability but diverge in predicting the response to climate in poorly investigated regions. Recent advances in the remote sensing of solar-induced chlorophyll fluorescence (SIF) opens up a new possibility to provide direct global observational constraints for GPP. Here, we apply an optimal estimation approach to infer the global distribution of GPP from an ensemble of eight DGVMs constrained by global measurements of SIF from the Greenhouse Gases Observing SATellite (GOSAT). These estimates are compared to flux tower data in N. America, Europe, and tropical S. America, with careful consideration of scale differences between models, GOSAT, and flux towers. Assimilation of GOSAT SIF with DGVMs causes a redistribution of global productivity from northern latitudes to the tropics of 7–8 Pg C yr⁻¹ from 2010 to 2012, with reduced GPP in northern forests (~3.6 Pg C yr⁻¹) and enhanced GPP in tropical forests (~3.7 Pg C yr⁻¹). This leads to improvements in the structure of the seasonal cycle, including earlier dry season GPP loss and enhanced peak-to-trough GPP in tropical forests within the Amazon Basin and reduced growing season length in northern croplands and deciduous forests. Uncertainty in predicted GPP (estimated from the spread of DGVMs) is reduced by 40–70% during peak productivity suggesting the assimilation of GOSAT SIF with models is well-suited for benchmarking. We conclude that satellite fluorescence augurs a new opportunity to quantify the GPP response to climate drivers and the potential to constrain predictions of carbon cycle evolution.

Keywords: amazon, carbon cycle, climate change, flux towers, model benchmarking, water stress

Received 22 December 2013; revised version received 23 March 2014 and accepted 29 April 2014

Introduction

Terrestrial carbon sinks absorb approximately a quarter of anthropogenic emissions but there is much debate as to whether these sinks will continue to buffer the effects of increasing carbon emissions and climate change (e.g., Friedlingstein *et al.*, 2006; Sitch *et al.*, 2008; Knorr, 2009; Arora *et al.*, 2013; Le Quéré *et al.*, 2013). A major international effort to improve understanding of carbon cycle processes, under the auspices of the Regional Carbon Cycle Assessment and Processes (RECCAP; Sitch *et al.*, 2013), attempts to constrain regional terrestrial carbon budgets by reconciling flux estimates from top-down atmospheric inversions and bottom-up methods

(Canadell *et al.*, 2011). However, caution is needed in interpreting estimates of net carbon flux. Model inter-comparisons indicate that estimates of net flux have a much narrower range than the subcomponent estimates of gross productivity and respiration, which are large and uncertain (e.g., Huntzinger *et al.*, 2012). To better constrain net flux and take full advantage of efforts such as RECCAP, better evaluation of these gross fluxes is needed.

Carbon fixation by the terrestrial biosphere, or gross primary production (GPP), constitutes the largest flux in the terrestrial carbon balance (Beer *et al.*, 2010). GPP is controlled by many factors such as climate, disturbance history, management practices, water and nutrient availability, soil type, species composition, and community structure. Most dynamical global vegetation models (DGVMs) either do not account for all

Correspondence: Nicholas C. Parazoo, tel. +818 354 2973, fax +818 354 3221, e-mail: nicholas.c.parazoo@jpl.nasa.gov

these dependencies or have different but plausible representations of these underlying processes, leading to differences in the GPP response to the same climate forcing (Sitch *et al.*, 2008, 2013; Schaefer *et al.*, 2012).

Given the critical roles of GPP in the global carbon budget and current model limitations, there has been much effort to provide empirical based constraints (Myneni *et al.*, 2007; Suntharalingam *et al.*, 2008; Beer *et al.*, 2010; Frankenberg *et al.*, 2011; Luo *et al.*, 2012). Recent advances in the remote sensing of solar-induced chlorophyll fluorescence (SIF) have opened up a new possibility to measure the rate of planetary photosynthesis at increasing spatial resolution (down to ~10.5 km diameter in the Greenhouse gases Observing SATellite (GOSAT)), providing direct seasonal constraints on global GPP (Frankenberg *et al.*, 2011; Joiner *et al.*, 2011, 2013; Guanter *et al.*, 2012; Lee *et al.*, 2013; Parazoo *et al.*, 2013). SIF is light re-emitted from leaf chlorophyll during photosynthesis and therefore offers a direct probe into the photosynthetic process (Flexas *et al.*, 2002; Rascher *et al.*, 2009; Damm *et al.*, 2010; Daumard *et al.*, 2010). Near global retrievals of SIF from GOSAT correlate strongly ($r^2 = 0.80$) at global annual scale with flux measurements-based GPP extrapolated globally from the MPI-BGC model (Frankenberg *et al.*, 2011; Beer *et al.*, 2010; Jung *et al.*, 2011; henceforth G-mpi) indicating that, on average, most of the photosynthesis that occurs during emissions of SIF also leads to carbon uptake. Correlation of SIF with satellite fPAR ($r^2 = 0.46$) and NDVI ($r^2 = 0.46$) are, however, much weaker (Frankenberg *et al.*, 2011). Consequently, SIF has demonstrated further skill over classical vegetation indices in detecting large-scale GPP changes, including physiological effects of water limitation and temperature stress (Daumard *et al.*, 2010; Lee *et al.*, 2013).

While previous studies have demonstrated that satellite SIF provides a direct linear predictor of global GPP, the observed linear relationship is empirical, and further studies are needed to better quantify the exact relationship between SIF and GPP (Maxwell & Johnson, 2000; Lee *et al.*, 2013). This empirical constraint is further limited by infrequent sampling in space (from GOSAT) and systematic under-sampling in clouds. We therefore seek an integrated methodology in the estimate of GPP that takes advantage of complementary information in satellite retrievals and vegetation models. For this purpose, we use an optimal estimation framework, which minimizes a cost function that balances the difference between observed and predicted values of GPP subject to *a priori* uncertainty and observation precision error.

The *a priori* mean and uncertainty are derived from an ensemble of independent DGVMs, allowing integration of *a priori* knowledge over a wider range of

processes than is possible with a single model and improving *a priori* estimates of seasonal GPP. This methodology is compatible with a variety of model intercomparison studies (e.g., Sitch *et al.*, 2008, 2013; Huntzinger *et al.*, 2012; Schaefer *et al.*, 2012; Taylor *et al.*, 2012), and has potential to provide critical feedback for models. Here, we evaluate GPP from a consortium of eight DGVMs, herewith referred to as TRENDY models, used as part of the Global Carbon Project to examine global annual carbon budgets as well as regional trends in land carbon sinks over the period 1990–2009 (<http://dgvm.ceh.ac.uk/node/9>) (Sitch *et al.*, 2013).

We identify four primary objectives: (1) establish an optimal estimation framework that incorporates DGVM, semi-empirical upscaling (G-mpi) and SIF from existing and future satellites, (2) evaluate optimal estimates using multiple SIF scaling strategies, data-based GPP models, and available flux tower data, (3) quantify the information content of GOSAT data with respect to *a priori* uncertainty, and (4) identify biomes and regions where model predictions diverge from optimal estimates.

Materials and methods

GPP optimal estimation framework

The relationship between a set of GPP observations at a grid cell location j to a corresponding set of predictions can be expressed as:

$$\mathbf{y}_j = f(\beta_j, j) + \epsilon_j \quad (2.1)$$

where $\mathbf{y}_j \in \mathbb{R}^{N_j}$ is a vector of N_j GPP observations, $f(\beta_j, j) : \mathbb{R} \rightarrow \mathbb{R}^{N_j}$ a vector of predicted (*a priori*) GPP values evaluated at the times and location of observations, and ϵ_j is a vector of error estimates. We define $\beta_j \in \mathbb{R}$ as a grid cell GPP scaling factor to adjust *a priori* GPP according to observations and subject to *a priori* and observational uncertainties. We can then use a Bayesian optimal estimation framework that solves for grid-scale GPP by minimizing a cost function (C_{β_j}):

$$C_{\beta_j} = \frac{1}{2} [\mathbf{y}_j - f(\beta_j, j)]^T \mathbf{R}_j^{-1} [\mathbf{y}_j - f(\beta_j, j)] + [\beta_j - \beta_b]^2 \sigma_j^{-2} \quad (2.2)$$

where $\beta_b \in \mathbb{R}$ is the *a priori* scaling factor estimate, $\mathbf{R}_j \in \mathbb{R}^{N_j \times N_j}$ is the observation error covariance matrix, and σ_j^2 is a scalar *a priori* scaling factor error. This provides a general framework in which a collection of observations and uncertainty estimates can be used to inform an *a priori* distribution of GPP.

Here, we seek an estimate of monthly GPP that optimally accounts for uncertainties in predictions of GPP from DGVM's, estimates of GPP inferred from satellite observations of midday SIF, and relationships between SIF and GPP. In this case \mathbf{y}_j represents GPP inferred from satellite measurements of SIF in one month, $f(\beta_j, j)$ midday GPP

predictions from DGVM's and evaluated at the times and location of inferred GPP, and β_j a monthly constant scaling factor which varies seasonally and spatially. These inputs are described in more detail in the following sections. Examples of inferred and *a priori* midday GPP (y_j and $f(\beta_j, j)$, respectively) at a grid point in North America (120°W, 40°N) are shown in Fig. S1.

Diagonal terms in R_j are equal to midday observation precision errors derived from GOSAT SIF retrievals (Frankenberg *et al.*, 2011) and additional error estimates as described below. We assume that satellite observational errors are uncorrelated in time and consequently set the off-diagonal terms equal to zero. σ_j^2 , which represents the variance of scale factors predicted by an ensemble of models, is estimated as

$$\sigma_j^2 = \text{var}(\beta_{l,j}) = \text{var}\left(g_{l,j}/\bar{g}_j\right) \quad (2.3)$$

where $g_{l,j}$ is monthly GPP from model ensemble member l , $\bar{g}_j = \frac{1}{M} \sum_{l=1}^M g_{l,j}$ is monthly GPP averaged across the ensemble of models, and M is the number of models. \bar{g}_j is used throughout this study to represent *a priori* monthly GPP. For each grid point, the optimal scaling factor is found by minimizing the cost function

$$\hat{\beta}_j = \min_{\beta} C_{\beta_j} \quad (2.4)$$

which can be written in terms of Eqn (2.2) as

$$\hat{\beta}_j = \beta_b + (\mathbf{K}_j^T \mathbf{R}_j^{-1} \mathbf{K}_j + \sigma_j^{-2})^{-1} \mathbf{K}_j^T \mathbf{R}_j^{-1} (y_j - f(\beta_j, j)) \quad (2.5)$$

(Tarantola, 2005) where

$$\mathbf{K}_j = \frac{\partial f(\beta_j, j)}{\partial \beta_j} \quad (2.6)$$

represents the gradient or Jacobian of predicted GPP with respect to the monthly scaling factor.

Uncertainty in the optimal scaling factor is

$$\hat{S}_j = (\mathbf{K}_j^T \mathbf{R}_j^{-1} \mathbf{K}_j + \sigma_j^{-2})^{-1} \quad (2.7)$$

with reduction of uncertainty estimated as

$$\text{err}_j = 1 - (\hat{S}_j)^{1/2} (\sigma_j^2)^{-1/2} \quad (2.8)$$

The uncertainty estimates in Eqn (2.7) are defined with respect to scale factors. We can recast those into uncertainties in prior and posterior GPP, respectively, as

$$Q_j = \bar{g}_j^2 \sigma_j^2 \quad (2.9)$$

$$\hat{Q}_j = (\bar{g}_j \beta_j)^2 \hat{S}_{\beta_j} \quad (2.10)$$

Finally, to examine patterns of uncertainty at zonal, regional, and seasonal scale we estimate the mean uncertainty from grid to regional scale as

$$\bar{Q} = \frac{1}{Nx} \sum_{j=1}^{j=Nx} Q_j \quad (2.11)$$

$$\hat{\bar{Q}} = \frac{1}{Nx} \sum_{j=1}^{j=Nx} \hat{Q}_j \quad (2.12)$$

where Nx the number of grid points in a given region. In the example in Fig. S1, our methodology derives a β_j that reduces

a priori GPP by a constant factor, with convergence toward inferred GPP weighted by *a priori* and observational uncertainties.

Prior mean and uncertainty

We use monthly averaged GPP output from the TRENDY project (<http://dgvm.ceh.ac.uk>) (Sitch *et al.*, 2013) to calculate *a priori* GPP and uncertainty ($f(\beta_j, j)$ and P_j respectively) (Piao *et al.*, 2013). TRENDY follows the studies of Le Quére *et al.* (2009) and Sitch *et al.* (2008) to examine spatial trends in net land-atmosphere carbon exchange over the period 1990–2009 using a consortium of DGVM's. Eight of the participating models provide monthly output, including TRIFID (Clark *et al.*, 2011), LPJ (Sitch *et al.*, 2003), LPJ-GUESS (Smith *et al.*, 2001; Ahlström *et al.*, 2012a,b), CLM4-CN (Thornton *et al.*, 2007), ORCHIDEE (Krinner *et al.*, 2005), OCN (Zaehle *et al.*, 2010), SDGVM (Cramer *et al.*, 2001), and VEGAS (Zeng *et al.*, 2005). The models are forced over the period 1860–2009 at 0.5° × 0.5° or coarser (Sitch *et al.*, 2013) using a factorial set of sensitivity simulations. Here, we use monthly mean GPP output from “S2” simulations forced with changing (CO₂) and climate information from the CRU-NCEP dataset, representing the combination of CRU TS.3.2 0.5° × 0.5° monthly climatology and NCEP reanalysis 2.5° × 2° 6 hour data (http://nacp.ornl.gov/thredds/fileServer/reccapDriver/cru_ncep/analysis/readme.htm).

The monthly *a priori* uncertainty in the GPP scaling factor is estimated from these data using Eqn (2.3). Deriving uncertainty from the spread of DGVMs also provides an estimate of the structural uncertainty in GPP arising from differences in model processes such as seasonal phenology and sensitivity to environmental forcing (e.g., light/water limitation) (Collins, 2007; O'Neill & Melnikov, 2008). In this way, the combination of prior and posterior uncertainty estimates will help to identify models and processes that are most consistent with optimal GPP and flux tower data (see below).

To estimate midday GPP at GOSAT overpass time from monthly TRENDY GPP, we follow a similar approach to Olsen & Randerson (2004) and distribute monthly GPP according to 3-h downward shortwave radiation flux derived from MERRA (Rienecker *et al.*, 2011). We then sample the 3-h GPP nearest to GOSAT overpass time. We note this technique does not account for important meteorological effects such as variations in water availability and that use of MERRA leads to inconsistencies with CRU-NCEP TRENDY forcing. However, MERRA provides improved resolution of diurnal variations relative to CRU-NCEP as well as coverage through 2012 (CRU-NCEP covers through 2010 at the time of this writing), providing a longer overlap with GOSAT SIF and therefore providing more robust estimates of seasonally averaged GPP.

To avoid noise driven by interannual variability of GPP in response to climate (Piao *et al.*, 2013), we average monthly output from the eight TRENDY models over the period 2000–2009 to create seasonal climatology. This has little influence on global GPP, which increases only slightly from 123.59 Pg C yr⁻¹ in 2000 to 125.05 Pg C yr⁻¹ in 2009 (124.77 Pg C yr⁻¹ in 2000–2009 average). Henceforth, the

ensemble average *a priori* monthly GPP, $\overline{g_{j,t}}$ will be denoted as G-pri. Optimal GPP, denoted as G-opt, is then G-pri multiplied by the optimal scale factor (Eqn 2.5). Zonal average mean and uncertainty of G-pri for the period 2000–2009 are shown in Fig. S2.

Midday GPP inferred from GOSAT chlorophyll fluorescence

We estimate midday GPP, denoted as G-sif, from GOSAT SIF data using the linear scaling technique described in Frankenberg *et al.* (2011), which infers GPP magnitude from the slope between SIF (collected \sim 1:30 pm local time, units of $\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$) and G-mpi. In this way, spatial and temporal GPP variability is constrained globally by SIF observations while annual magnitude is constrained by semi-empirical observations (G-mpi = $121.67 \text{ Pg C yr}^{-1}$, \sim 3% less than the average of TRENDY models).

We first calculate midday SIF from a single retrieval as the average of two bands (757 nm and 771 nm) and two polarizations (p and s), with SIF at 771 nm scaled by a factor of 1.8 since it falls farther away from peak emission on the SIF spectrum (e.g., Joiner *et al.*, 2011). Single retrievals are aggregated to $2.5^\circ \times 2^\circ$ grid cell mean ($\overline{\text{SIF}}$, where brackets indicate spatial averages) for each day with 1 or more retrievals. Since instantaneous measurements of SIF cannot be directly compared with GPP at global scale due to variability in local overpass time, length of day, and variations of the solar zenith angle (SZA), we approximate daily grid cell averaged SIF ($\overline{\text{SIF}}$, where overbar represents time averages) by normalizing $\overline{\text{SIF}}$ as the cosine of the solar zenith angle at overpass time t_0 and integrating over fractional length of day $t_0 + 1$ (Frankenberg *et al.*, 2011). We note this scaling assumes a linear dependence of GPP with the incoming radiation during the daily cycle and does not account for saturation trends of GPP with radiation,

which may lead to overestimates of GPP in pixels with local overpass close to noon.

Guanter *et al.* (2012) show that the relationship between SIF and GPP is sensitive to vegetation type and climate. We account for this variability by calculating biome-specific regressions using a modified IGBP classification following Frankenberg *et al.* (2011) consisting of needleleaf forests (NF), evergreen broadleaf (EB), deciduous broadleaf (DB), shrubland (SH), savannah (SV), grassland (GR), and cropland (CR). A map of biome distribution is shown in Fig. 1. We calculate slope (m) and y-intercept (b) by regressing annual averages of $\overline{\text{SIF}}$ and G-mpi for points with annual $1-\sigma$ measurement precision $< 0.04 \text{ W m}^{-2} \mu^{-1} \text{sr}^{-1}$. SIF precision is estimated as $\sigma_{\text{SIF}}/\sqrt{N-1}$, where $\sigma_{\text{SIF}} = 0.5 \text{ W m}^{-2} \mu^{-1} \text{sr}^{-1}$ is the retrieval measurement error (Frankenberg *et al.*, 2011) and N is the number of samples per $2.5^\circ \times 2^\circ$ grid cell overpass.

Scatter plots of G-mpi vs. $\overline{\text{SIF}}$ are shown for each biome in Fig. S3(a–g). High latitude biomes such as NF and DB have smaller sample size due to limited availability of winter observations. Still, the correlation is generally high for more productive biomes ($r \sim 0.82\text{--}0.89$) and low for the least productive biome ($r = 0.61$ for SH). The slope is greatest in CR ($17.08 \pm 0.83 \text{ gC m}^{-2} \text{d}^{-1}/\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$), smallest in SH ($6.44 \pm 0.82 \text{ gC m}^{-2} \text{d}^{-1}/\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$), and has a smaller range of variability across the remaining biomes ($13.46 \pm 0.91 \text{ gC m}^{-2} \text{d}^{-1}/\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$ in GR up to $15.82 \pm 1.1 \text{ gC m}^{-2} \text{d}^{-1}/\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$ in EB). Biome-specific slopes are smaller and less correlated compared to global regressions of all grid points, which produces 553 total samples with slope of $18.19 \pm 0.31 \text{ gC m}^{-2} \text{d}^{-1}/\text{W m}^{-2} \mu^{-1} \text{sr}^{-1}$, and correlation of 0.93 (Fig. S3 h). In particular, the SH slope is less than half that of other biomes, indicating SIF varies more strongly relative to G-mpi. The cause of this discrepancy is unclear from a physiological point of view, but it could suggest an underestimate of G-mpi for these plant

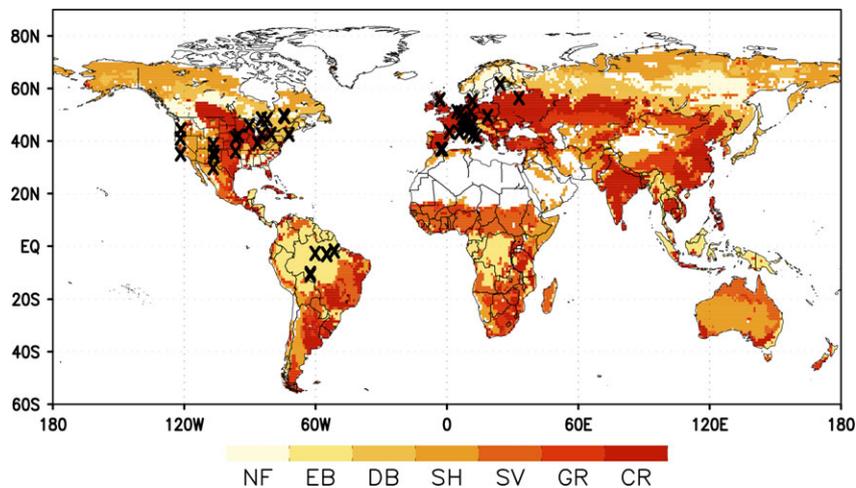


Fig. 1 Map of biome distribution and flux tower locations (indicated by X), including 49 sites in N. America and Europe and seven sites in S. America. Biomes are determined from a modified IGBP classification scheme following Frankenberg *et al.* (2011) and abbreviated as follows: NF, needleleaf forests; EB, evergreen broadleaf forests; DB, deciduous broadleaf forests; SH, shrublands; SV, savannahs; GR, grasslands; CR, croplands. (Color in the online version)

functional types due to undersampling of SH biomes by FLUXNET. Biome-specific slopes are used to convert $\overline{\text{SIF}}$ and uncertainty to daily GPP, which is scaled to midday GPP using $\cos(\text{SZA})$ and length of day. We note that our optimization algorithm is not sensitive to the use of daily vs. midday GPP, as long as the conversion from midday to daily average is consistent between *a priori* and inferred GPP. This final result, representing midday inferred GPP (G-sif), is used to calculate γ in Eqns (2.1) and (2.2).

Sensitivity to SIF scaling and observation error

Final estimates of G-opt for this study are based on biome-specific scaling against G-mpi and three estimates of observation uncertainty: (1) SIF measurement error, (2) SIF conversion error, and (3) errors in G-mpi. We acknowledge that estimates of G-opt are sensitive to SIF scaling strategy, including the set of grid points chosen for least-squares regressions (e.g., biome-specific vs. global points), choice of empirical GPP model, and observation error. We therefore examine sensitivity to a set of five different scaling strategies, which are summarized in Table 1 and referred to as SIF1-5 (for reference, our “best” estimate for this study is SIF4). Most of these details can be found in Text S2, but we provide a brief analysis here. Later, in the main results section, we examine the impact of these choices on G-opt and posterior uncertainty.

In general, global scaling (SIF1) leads to increased magnitude of midday G-sif compared to biome-specific scaling (SIF2-4), consistent with higher regression slopes. The exception is in the tropics where biome-based estimates are enhanced due to high γ -intercept in EB biomes, which reflects high empirical GPP (G-mpi) at low values of SIF.

We find a 5–10% increase in midday G-sif uncertainty due to scaling errors between SIF and GPP when root mean squared errors in slope and γ -intercept (resulting from scatter in the least squared regression) are added in quadrature to

measurement precision errors (for G-opt, these errors are incorporated into the SIF observation error covariance matrix). This enhances midday uncertainty (SIF3) by 0–0.5 $\text{gC m}^{-2} \text{d}^{-1}$ (~5–10%). Uncertainties in G-mpi, which arise from preprocessing of tower CO_2 flux measurements, tower representativeness, flux partitioning into GPP, uncertainties of climate and remote sensing datasets, and structural uncertainties of diagnostic models going into G-mpi calculations [Beer *et al.* (2010) and Fig. S2], lead to global standard deviation of 8 Pg C yr^{-1} . This uncertainty, which is included in our final estimate (SIF4), is similar in magnitude to rmse regression errors (5–10%).

Guanter *et al.* (2012) show that the relative distribution of slopes for different biomes is consistent for data-driven GPP models [G-mpi and MODIS MOD17A2 (Myneni *et al.*, 2007), denoted G-mod], but with slope more variable and correlation weaker in G-mod compared to G-mpi. We find a similar pattern, including smaller absolute slopes, consistent with weaker global GPP (106.5 Pg C yr^{-1} in 2010 in G-mod compared to 121.2 Pg C yr^{-1} in G-mpi). Frankenberg *et al.* (2011) and Guanter *et al.* (2012) speculate that different slopes are related to differences in the structure of each model (in particular, the use of a biome-specific look-up table for light use efficiency (LUE) in G-mod) and reduced correlation in G-mod to remote sensing issues in high cloud regions. The reason for these discrepancies is beyond the scope of this study; the important point is that scaling of SIF against G-mod (SIF5) reduces midday GPP and uncertainty, which propagates to G-opt and reduces the overall magnitude of constrained GPP. For this study, we do not account for uncertainty due to choice of GPP model; rather, we include estimates of G-opt based on SIF scaling to G-mod as a reference to evaluate results based on G-mpi.

Temporal sampling biases in GOSAT

SIF retrievals are screened for scenes with high optical depth due to difficulty in measuring SIF in the presence of high scattering of sunlight. While this procedure eliminates low quality data, regions with high frequency of clouds and aerosol loading have lower sampling yield (see Fig. 2). Optimal estimation will find the appropriate balance between models and SIF data in these regions, but we have to be aware of possible sampling biases introduced into optimized GPP related to systematic sampling of clear-sky conditions, which is expected to cause differences in observed and expected time averaged GPP due to sensitivity to total incoming radiation.

We examine clear sky sampling biases by comparing time-averaged model GPP to satellite subsets based on GOSAT footprints for each grid cell. We refer to these estimates of time-averaged GPP as “true” and “observed” GPP (G_T and G_O , respectively) to illustrate how clear sky satellite derived observations may differ from the true mean. Biases are calculated at monthly scale using 3-hour data from the CASA-GFED3 vegetation model, a LUE model that estimates GPP using relationships between LUE and fPAR (van der Werf *et al.*, 2010; data downloaded from <http://nacp-files.nacarbon.net>).

Table 1 Description of SIF scaling strategies. The first row represents a particular strategy (SIF1-5) based on subsampling (row 2, biome or global), semi-empirical GPP product (row 3, G-mpi or G-mod), and observation errors (row 4–5) including SIF conversion error as quantified by root-mean-squared error of regression slope and γ -intercept (row 4) and uncertainty in GPP product (row 5) SIF measurement error is included in all strategies. G-mpi represents flux tower based data from the MPI-BGC semi-empirical model (Beer *et al.*, 2010). G-mod represents represent remote sensing data from the MODIS MOD17A2 product (Myneni *et al.*, 2007).

Scaling strategy	SIF1	SIF2	SIF3	SIF4	SIF5
Subsampling	Global	Biome	Biome	Biome	Biome
Semi-empirical GPP model	G-mpi	G-mpi	G-mpi	G-mpi	G-mod
Scaling error	NO	NO	YES	YES	YES
	NO	NO	NO	YES	NO

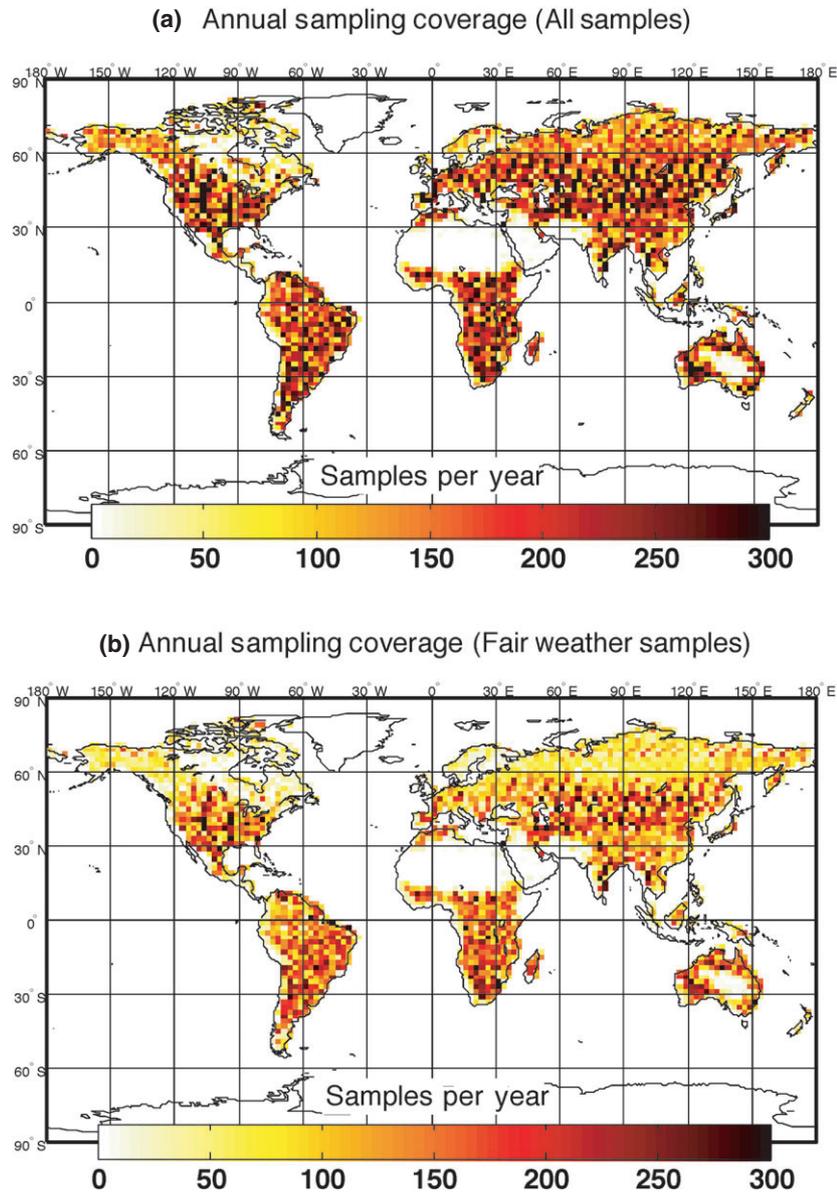


Fig. 2 Maps of total annual GOSAT SIF sampling coverage in 2010, aggregated to $2.5^\circ \times 2^\circ$ pixels, for (a) all available samples that have passed quality control (including cloudy scenes with high aerosol optical depth) and (b) fair weather samples only (excluding cloudy scenes). Cloudy scenes are flagged using simultaneous retrievals of the oxygen A-band. (Color in the online version)

org/nacp-kawa-01/). G_T is taken as the monthly average of all midday values and G_O as the average of midday GOSAT footprints.

G_T and G_O have high correlation and slope of fit for all biomes ($R^2 > 0.92$, slope ranges from 1.01 to 1.08). Sampling bias, defined as $100 * (\overline{G_O} - G_T / \overline{G_T})$, are shown as grid-scale and zonal averages in Figs. S9 and S10 respectively. We estimate potential grid scale biases during boreal summer (Jun–Aug 2010) ranging from 0% to 10% in the tropics (~3.7% in the zonal average), 5–15% in mid-latitudes (~4.5% in the zonal average), and 5–20% at high latitudes (9.8% in the zonal average). Sampling biases are reduced by 3.2%, 3.7%, and 7.2% in the tropical, mid-latitude, and high latitude zonal average,

respectively, when retrievals flagged as cloudy according to simultaneous retrievals of the oxygen A-band are included in the GOSAT observations. This suggests that these biases are driven in part by differences in light limitation between clear and cloudy days.

Low bias in the tropics is consistent with simulations of SIF in the Soil Canopy Observation of Photochemistry and Energy Flux (SCOPE) (Van der Tol *et al.*, 2009) photosynthesis model, which indicate that GPP is weakly sensitive to changes in light under high light conditions typical of tropical regions (Lee *et al.*, 2013). In contrast, high latitude vegetation is temperature- and light-limited, causing increased GPP sensitivity to synoptic fluctuations. Requiring a minimum threshold of

sampling days for data assimilation can reduce these biases. For example, sampling bias in NF biomes in mid-latitudes decreases from 7.3% using a minimum coverage of 1-day per month to 6.2% using 5 days and 5.6% using 10 days. This benefit is small, however, compared to loss of sampling coverage, which decreases by 25% in the 5-day minimum and 85% with 10 days. We therefore use a 1-day minimum requirement to optimize GPP.

Observing system simulation experiments (OSSE)

We use OSSE's to demonstrate proof-of-concept of our optimal estimation strategy, which aims to retrieve the true state of monthly GPP and uncertainty from *a priori* estimates of GPP (from DGVMs) constrained by midday clear sky observations and subject to temporal sampling biases and uncertainty in measurements and models. Before interpreting this satellite constrained GPP, it is imperative to first understand sensitivity to changes in model and observational constraints. To this end, we examine sensitivity of optimal GPP estimates to (a) *a priori* constraints of diurnal GPP, (b) observational coverage, and (c) inclusion of cloudy retrievals. These experiments are summarized in Table 2.

As in the real world case, *a priori* GPP and uncertainty are taken as the ensemble average and standard deviation of TRENDY models. The true state of GPP, unknown in the real world, is prescribed artificially in OSSE's, in this case using simulated GPP from CASA-GFED3. We then create synthetic GPP observations by sampling midday GPP from CASA-GFED3 according to GOSAT v3.3 footprints that have passed through prior and posterior retrieval screening. This represents a greatly simplified version of the real world since we directly observe GPP rather than make inferences from SIF, and therefore does not require SIF/GPP scaling assumptions and consequently is not subject to the errors associated with these assumptions. The synthetic measurement error is taken from the GOSAT v3.3 SIF retrieval errors. The SIF retrieval error, $0.5 \text{ W m}^{-2} \mu^{-1} \text{ sr}^{-1}$, is approximately equal to the average of global SIF measurements; we therefore set the synthetic measurement error to the average of global midday GPP samples. Global OSSE's are examined for the period 2010 in the tropics and northern mid-latitudes.

Flux tower data

Eddy covariance flux tower data are described in detail in Text S5. This analysis uses estimates of GPP from 49 eddy covari-

ance flux tower sites in N. America and Eurasia and 7 sites in S. America, shown in Fig. 1 and described in more detail in Table S1. In N. America and Europe, we select only sites that have at least 12 calendar months of data from 2009 to 2010. Flux tower data in tropical S. America are limited to the period 1999–2006 and therefore these comparisons are based on climatological seasonal cycles. GPP is inferred from observations of net ecosystem exchange (NEE) and modeled ecosystem respiration (R_{eco}) as $\text{GPP} = R_{\text{eco}} - \text{NEE}$ (assuming NEE is negative for carbon uptake) using partitioning techniques based on models of temperature sensitivity (Reichstein *et al.*, 2005), artificial neural networks (Papale *et al.*, 2006), and/or light response curves (Lasslop *et al.*, 2010).

For comparison to G-pri and G-opt, we average over multiple flux towers and years to examine seasonal averages. For N. America and Europe, we group towers by biome, which yields 12 NF sites, 14 DB sites, 11 GR sites, 8 CR sites, and 4 SH sites spread throughout N. America and Europe from 2009 to 2010. For S. America, we group all available towers together, and compare to the average from 5 EB and 2 SV sites from 1999 to 2006. We then compare flux tower data to the 2000–2009 average for G-pri and the 2009–2012 average for G-opt. Our main findings are not sensitive to averaging period.

Results

Optimal estimation using synthetic GPP (OSSE's)

Meridional cross sections of simulated (G-sim, black), prior (G-pri, blue), and optimal (G-opt, red) GPP and uncertainty are shown along with mean absolute error (MAE) for G-pri and G-opt (dashed and solid grey lines, respectively) for July 2010 in Fig. 3. Seasonal distributions of uncertainty reduction and MAE are shown for the tropics and northern midlatitudes in Figs. S11 and S12. Since the true state of GPP is prescribed from simulations of a vegetation model, and therefore perfectly known, MAE provides a metric to evaluate temporal sampling bias resulting from the assimilation of clear sky satellite observations. G-sim and G-pri have similar spatial distributions, with peak GPP in the tropics and northern extra-tropics, but different amplitudes, with *a priori* amplitude overestimated in the tropics and subtropics and underestimated in the extra-tropics. High spread among TRENDY models used in the calculation

Table 2 Description of Observation System Simulation Experiment (OSSE) test cases. Rows indicate test case (Test 1–4). Columns indicate differences in experimental design. Diurnal refers to estimate of diurnal variability (explicitly with CASA-GFED3 or down-scaled from monthly using MERRA SW). Sampling coverage refers to percent of actual GOSAT retrievals used for assimilation. Cloud refers to whether retrievals flagged as cloudy are included or not. OSSE results are shown in Fig. 3

OSSE case	Test 1	Test 2	Test 3	Test 4
Diurnal	CASA-GFED3	CASA-GFED3	CASA-GFED3	MERRA SW
Sampling coverage	10%	100%	100%	100%
Cloud	NO	NO	YES	YES

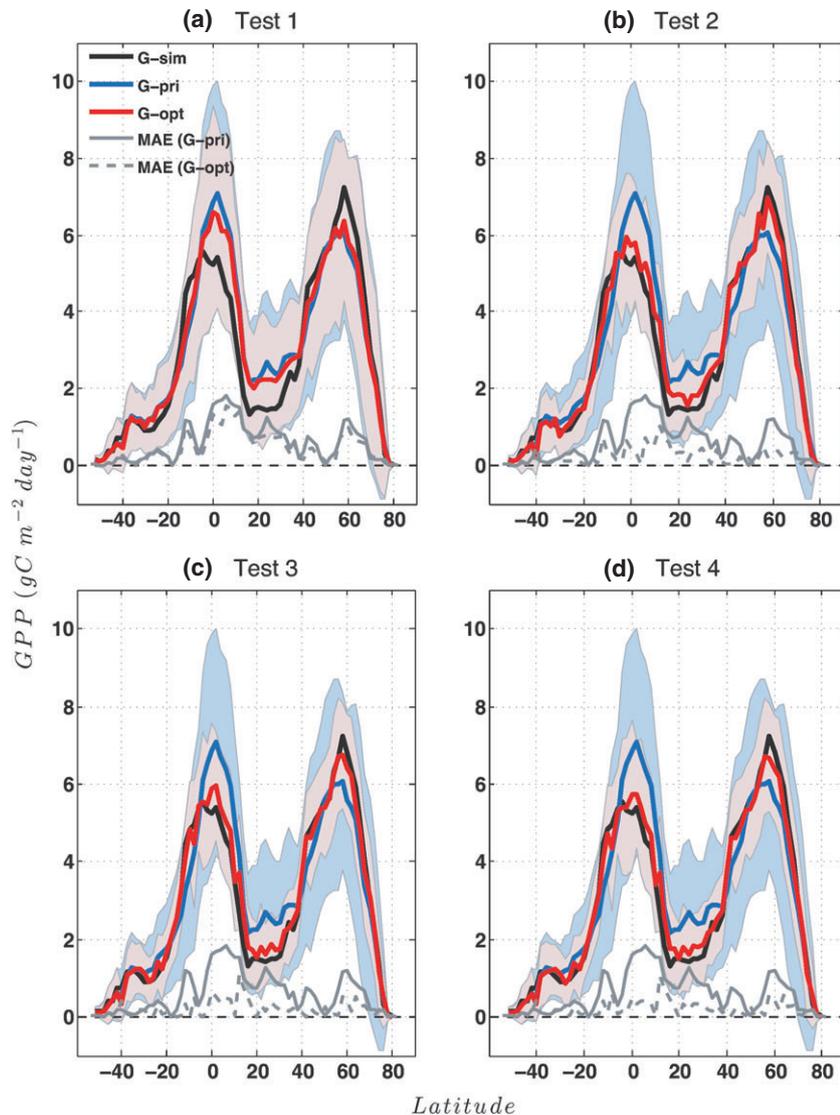


Fig. 3 Zonal average GPP and uncertainty for July 2010 based on an observing system simulation experiments (OSSE), which tests the optimal estimation methodology discussed in the main text using simulated GPP and observations. Here, simulated GPP (G-sim, black) is prescribed from the CASA-GFED3 vegetation model, observed GPP from midday samples of CASA-GFED3 collected from GOSAT satellite footprints, and prior GPP (G-pri, blue) from the ensemble average of TRENDY models. Optimal estimates (G-opt) are shown in red. Plots a–d represent four test cases, with each defined according to: (a) *a priori* estimate of diurnal GPP, (b) percentage of synthetic GPP observations retained, and (c) inclusion of cloudy retrievals (see Table 1). Grey lines show mean absolute errors of G-pri (solid) and G-opt (dashed). Blue and red shading represent prior and posterior uncertainty respectively. (Color in the online version)

of G-pri (1-sigma standard deviation among models is ~30–50% in July) creates high *a priori* uncertainty such that G-sim lies within the *a priori* constraint. G-opt and uncertainty, which are calculated from Eqns (2.4) and (2.6), are described below.

Test 1. In the first two tests we estimate diurnal GPP from 3-hour CASA-GFED3, and therefore assume diurnal GPP is perfectly known. In Test 1, however, we reduce the total number of observations by randomly selecting 10% of available synthetic data, effectively

reducing the possible GOSAT sampling constraint by an order of magnitude. As a result, we see only a slight reduction (<10%) of MAE of G-opt (dashed grey line in Fig. 3 and dark blue lines in Figs. S11b and S12b) and little reduction of uncertainty, ranging from 5% to 10% in the tropics and subtropics and up to 15% in northern latitudes.

Test 2. Retaining all available synthetic samples reduces MAE by ~50–75% in the tropics and subtropics (highest percent reduction in boreal winter when MAE

is smallest) and 50–100% in the northern extra-tropics. The magnitude of MAE reduction is similar throughout the year, with higher percent reductions when MAE is low (e.g., tropics during boreal winter), highlighting the importance of an accurate prior. Posterior uncertainty is reduced by 30–45% in the tropics year round, 20–40% in the subtropics with peak reduction in the summer hemisphere, 30–50% in the extra-tropics up to 40°N during winter and 60°N during summer.

Test 3. In this case, we add samples previously flagged as cloudy based on high optical depth to examine their potential impact on the estimate. We do this for two reasons: (1) to test potential information loss by ignoring GPP under thick clouds, and (2) in preparation for future instruments such as the Orbital-Carbon-Observatory 2 (OCO-2) that will have higher resolution sampling footprint and therefore increased potential to observe through cloudy scenes. Adding these retrievals in July 2010 increases sampling coverage by ~10–20% in the tropics, 20–40% in the northern extra-tropics up to 50°N, and effectively doubles coverage at high latitudes. The overall impact of cloudy pixels on G-opt is fairly small in July 2010, with additional 5–10% uncertainty reduction throughout the year in the tropics and mid-latitudes. The degree of GPP bias reduction (MAE) varies between tropical and midlatitudes. It is generally small in the tropics during late dry season and throughout the wet season (July–April), but increases to ~30% from May–June. In northern midlatitudes, it peaks at ~50% during summer but is closer to 10–30% during shoulder season. These results suggest that information loss due to clouds is potentially significant. This has important implications for high-resolution satellites such as OCO-2 which will have greater odds of sampling in-between clouds and therefore reduced regional and temporal GPP sampling biases.

This test also reveals a change in the amount of uncertainty reduction possible after August 2010, when GOSAT switched from 5-point to 3-point sampling mode (Crisp *et al.*, 2012). While this switch increases the number of retrievals per sample, it effectively reduces sampling coverage in space. As a result, uncertainty reduction, which averages 35–40% in Amazonia from January–July 2010 in Test 3, drops to 25–30% from August–December 2010.

Test 4. Test 4 is identical to Test 3 except diurnal GPP is estimated by downscaling TRENDY monthly GPP using shortwave radiation from MERRA. Results show little change in uncertainty reduction at global scale and GPP bias in midlatitudes. However, mean absolute error is reduced by 20–50% in the tropics, indication that our optimization methodology is sensitive to esti-

mates of midday GPP. This issue is particularly relevant in poorly investigated regions as the tropics where measurement signal-to-noise ratio (SNR) is low and model estimates of the timing and magnitude of diurnal GPP are difficult due to sensitivities to precipitation, temperature, humidity, and soil moisture stress, and circadian rhythms operating in vegetation (Baker *et al.*, 2013), whereas here we have only accounted for effects due to solar radiation. We do not explore this issue in any more detail in this article, but caution the reader that future investigation is needed.

Optimal estimates using GPP inferred from GOSAT SIF

Estimates of G-opt in this section are based on SIF4, with occasional references to other scaling techniques. Figure 4 shows characteristic annual maps of prior and posterior GPP, with high annual GPP in the tropics and reduced GPP toward the poles. GPP patterns in G-opt are similar to G-pri except with enhanced spatial variability and pixelation due to assimilation of SIF data. Figure 5 shows the corresponding reduction of uncertainty in GPP scaling factor, which is estimated according to Eqn (2.7). Because sunlight drives photosynthetic activity and increases sampling coverage, uncertainty reduction is correlated with sampling coverage and GPP magnitude. Indeed, comparison with Fig. 2b indicates that *a priori* uncertainty is reduced in all regions where SIF data are collected. Uncertainty reduction ranges from 30% to 50% in most productive tropical and temperate regions and exceeds 50% during the growing season in southeast N. America and Europe and during dry months throughout the tropics. Uncertainty reduction is much weaker in regions of low SNR where GPP is weak or zero and SIF error/scatter are high, including tundra regions (Arctic N. America), deserts (Australia, southwestern S. and N. America, and Gobi, Arabian, and Sahara), and mountains (Himalayas). Data are generally discarded or not retrieved in ice-covered regions and during winter at high latitudes, when reflected sunlight is weak or zero, explaining why annual uncertainty reduction is weaker in extra-tropical latitudes.

Uncertainty reduction in GPP scale factor is reflected in plots of zonal mean GPP and uncertainty (Fig. 6), where grid scale uncertainty is aggregated zonally according to Eqns (2.11) and (2.12). The global magnitude and spatial distribution of G-opt is highly consistent with G-pri (Global GPP is 127.42 Pg C yr⁻¹ in 2010 in G-opt compared to 124.77 Pg C yr⁻¹ in G-pri in 2009) with high GPP year-round in the tropics and during the growing season in northern latitude temperate regions, indication that the ensemble of TRENDY DGVM's provides a reasonable *a priori* constraint on GPP spatial distribution.

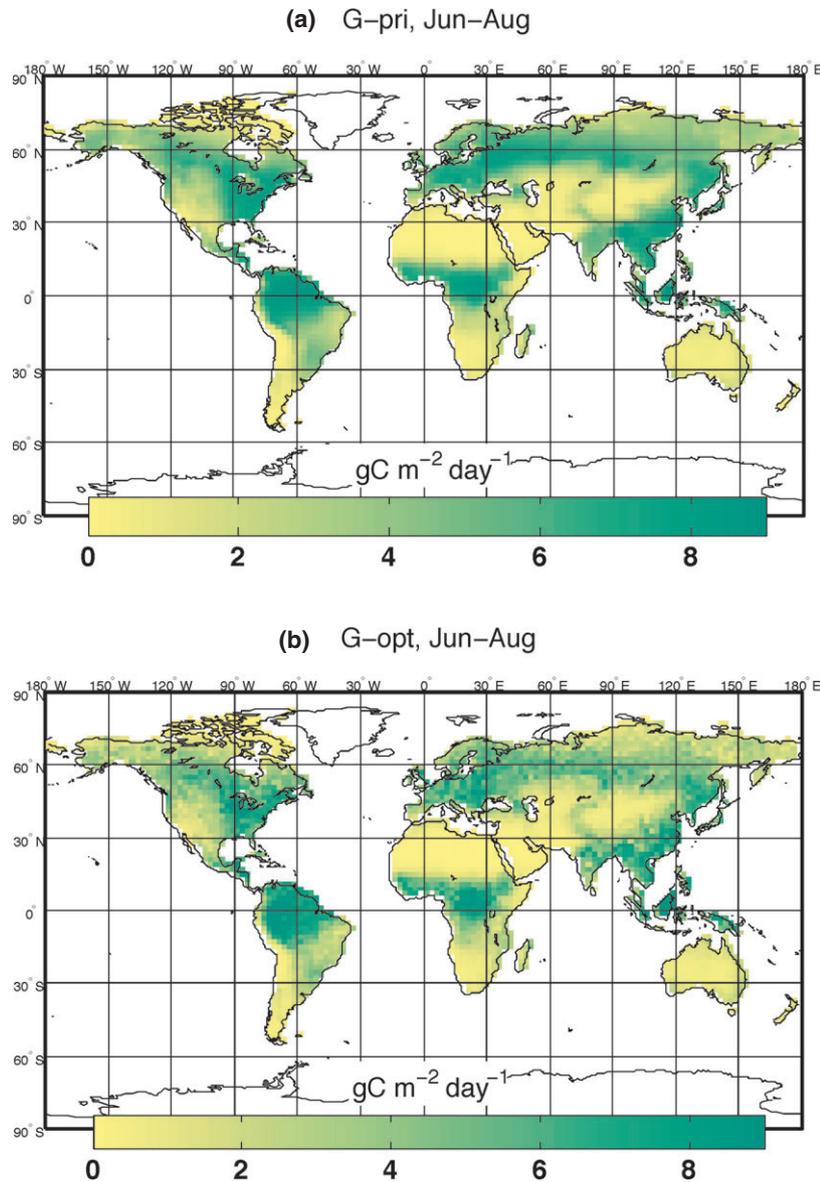


Fig. 4 Annual maps of GPP in 2010 based on (a) the average from eight TRENDY models (G-pri) and (b) the optimal combination of TRENDY models and GOSAT SIF (G-opt). (Color in the online version)

Despite similar spatial patterns, there is a substantial shift in global production from high to low latitudes compared to DGVMs, with GPP enhancement in the tropics of 10% in the annual mean and GPP damping by 10–20% in midlatitudes during the growing season. This translates to an average increase of production in the tropics (20°S–20°N) of 3.56 Pg C yr⁻¹ from 2010 to 2012 (3.96 in 2010, 3.65 in 2011, and 3.34 in 2012) and decrease in northern latitudes (20°N–90°N) of 3.65 Pg C yr⁻¹ (3.81 in 2010, 3.50 in 2011, and 3.39 in 2012). This global shift is independent of model prior year (i.e., GPP shifts from high to low latitudes for all year-specific TRENDY priors from 2000 to 2010) but

with magnitude slightly sensitive to model prior year, with an increasing trend of GPP reduction in high latitudes (3.7 Pg C in 2000 vs. 4.6 Pg C in 2010) and decreasing trend of GPP enhancement in the tropics (4.0 Pg C vs. 3.3 Pg C). These results, which require further investigation, suggest that partitioning of anthropogenic carbon uptake from northern latitudes to the tropics may be underestimated by DGVMs.

Comparison to additional SIF scaling strategies listed in Table 2, and summarized in Fig. S6, shows negligible sensitivity in the zonal average to changes in observation uncertainty (SIF2-4), weak sensitivity to subsampling (SIF1-2; global GPP increases from

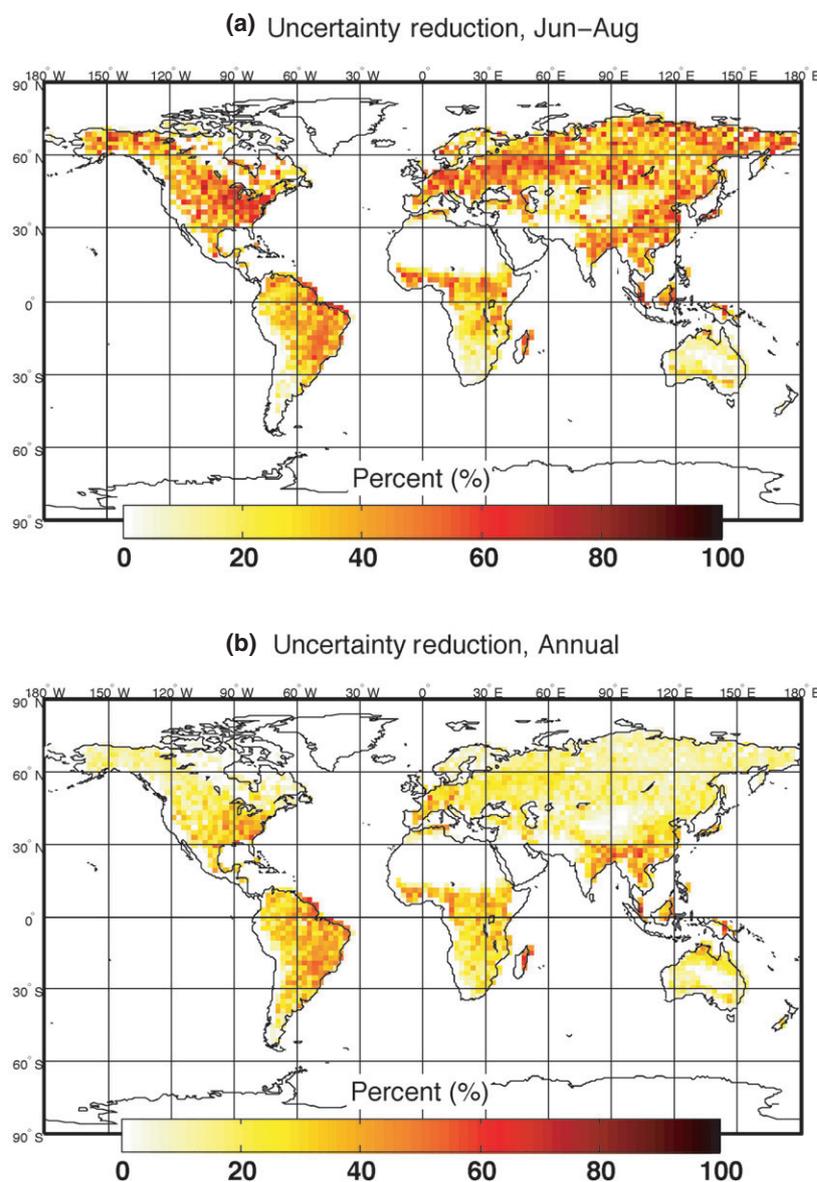


Fig. 5 Map of uncertainty reduction (expressed as percentage) of monthly GPP scale factor in 2010 during (a) boreal summer (June–August) and (b) in the annual average. (Color in the online version)

127.41 Pg C yr⁻¹ for biome-specific scaling to 131.82 Pg C yr⁻¹ for global scaling), and high sensitivity to choice of data-driven GPP (SIF3 vs. SIF5; global GPP decreases to 119.42 Pg C yr⁻¹). Specifically, scaling SIF to G-mod instead of G-mpi reduces GPP magnitude and uncertainty across all latitudes. The meridional distribution of GPP and uncertainty is maintained in all cases.

Comparison of individual TRENDY models to prior and posterior GPP and uncertainty helps identify models that may be outliers in specific regions and/or times. Comparison to prior uncertainty, for example, suggests that CLM4CN and LPJ-GUESS systematically over- and under-estimate tropical GPP, respectively,

throughout the year relative to the TRENDY ensemble and G-opt, while Orchidee and LPJ overestimate growing season GPP in the northern extra-tropics. Optimizing with SIF data and subsequent uncertainty reduction tightens the probable range even further, bringing additional models (e.g., LPJ-GUESS and VEGAS in the extra-tropics) to the edge of our optimal constraint. Models will inevitably have different strengths and weaknesses depending on time of year, region, and biome type, which is addressed in some detail with respect to GPP in the Amazon Basin (below). However, detailed analysis of models and underlying processes is beyond the scope of this study. The primary goal here is to present a framework to evaluate these models.

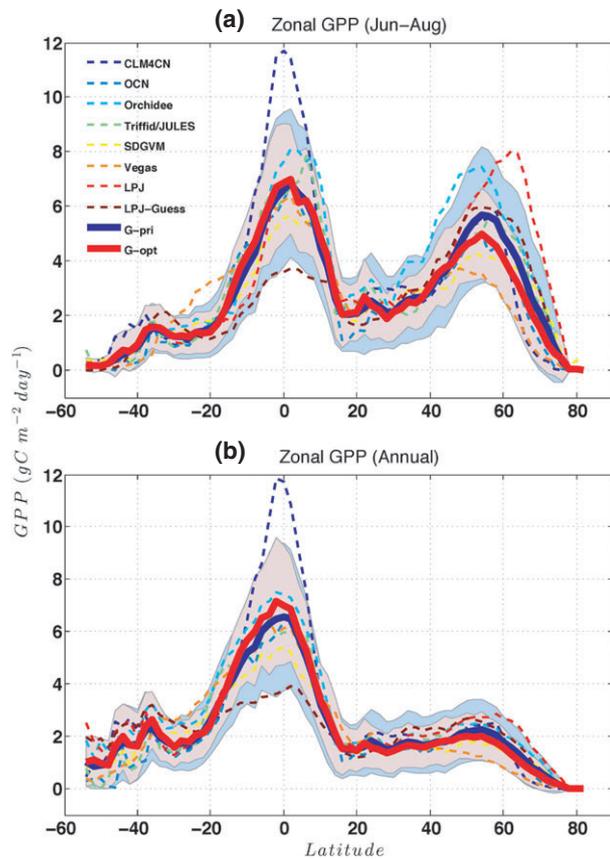


Fig. 6 Zonal average GPP and uncertainty in 2010 for individual TRENDY models (dashed), ensemble average of TRENDY models (G-pri, solid blue), and the optimal combination of TRENDY models and GOSAT SIF (G-opt, solid red), shown for (a) boreal summer (June–August) and (b) the annual average. Shading represents prior (blue) and posterior (red) uncertainty. (Color in the online version)

Figure 7 shows anomalies of the optimal scale factor (deviation from the *a priori* value of unity) during boreal summer (June–August) at grid and biome scale. A general pattern emerges at global scale with increased range of spatial GPP variability compared to models which correlates to some extent with productivity, with reduced GPP in northern latitudes in low productivity tundra (SH, ~20% decrease at biome scale) and medium productivity temperate (DB and NF, ~5–10%) regions and enhanced GPP in high productivity tropical regions (EB, ~20%). The exception is that GPP is actually significantly enhanced in low productivity desert, mountains, and subtropical regions (including SV, SH, and GR, ~10–20% increase at grid scale).

We also find a tendency for reduced GPP in all biomes with increasing productivity. This is illustrated in scatterplots of scale factor anomaly and G-pri (Fig. 8). Although scatter is high, all biomes show tendency for a decrease in scale factor anomaly with increasing GPP,

with regressions indicating negative anomalies at high GPP on average in NF, GR, DB, and SH biomes (isolated here to tundra ecosystems north of 50°N).

Prior and posterior uncertainty, also plotted in Fig. 8, show reduced uncertainty with increasing GPP, indication corrections are not statistically significant at low values of GPP. Posterior uncertainty is also high at high GPP in SH biomes, but the scale factor anomaly is so high at increasing GPP that this correction becomes increasingly significant.

High uncertainty in SH biomes is also evident in seasonal plots (Fig. 9), but so is the statistical significance of this correction, including consistency across all SIF scaling strategies except SIF1 (global regression, which is least likely to apply in SH considering the small number of samples). Corrections at high GPP in NF biomes are also statistically meaningful (Fig. 8a), but the number of corrected points is limited. Consequently, seasonal plots (Fig. 9a), which also show a decrease in G-opt during the growing season, show no significant difference from G-pri. Seasonal estimates of G-pri and G-opt are in good agreement in SV and GR biomes (Fig. 9e, f respectively), with low scatter particularly at medium and high values of GPP and lack of bias in seasonal averages (Fig. 8e, f respectively). This suggests low *a priori* uncertainty due to seasonal agreement among TRENDY models; nevertheless, the optimal combination with GOSAT SIF reduces uncertainty in this estimate.

Seasonal estimates in CR biomes are also in good agreement (Fig. 9g), but there is significant grid scale variability at high GPP (Fig. 8g), suggesting TRENDY models capture the mean behavior of crops but the combination with SIF data improves spatial and sub-grid variability. This is especially encouraging considering most TRENDY models do not explicitly represent crop functional types, and suggests that crude representation of crops may be sufficient in terms of the carbon cycle. We note that estimates based on G-mod are 10–20% lower but not statistically significant.

The same is true for EB biomes (Fig. 8b), which represent more productive moist tropical forests in S. America, Africa, and Tropical Asia. We find a similar range of variability at high GPP from Dec–Feb (data not shown) but with high positive bias, explaining enhanced GPP in January (Fig. 9b). Although still not statistically different from the prior in the regional average, corrections of mean GPP, seasonal amplitude, and seasonal phase in EB biomes represents the most significant divergence of optimal estimates from TRENDY models of all biomes analyzed here (including scaling based on G-mod, though to a lesser extent), and represents the only case in which the amplitude of temporal variability implied by models is enhanced by SIF. Given

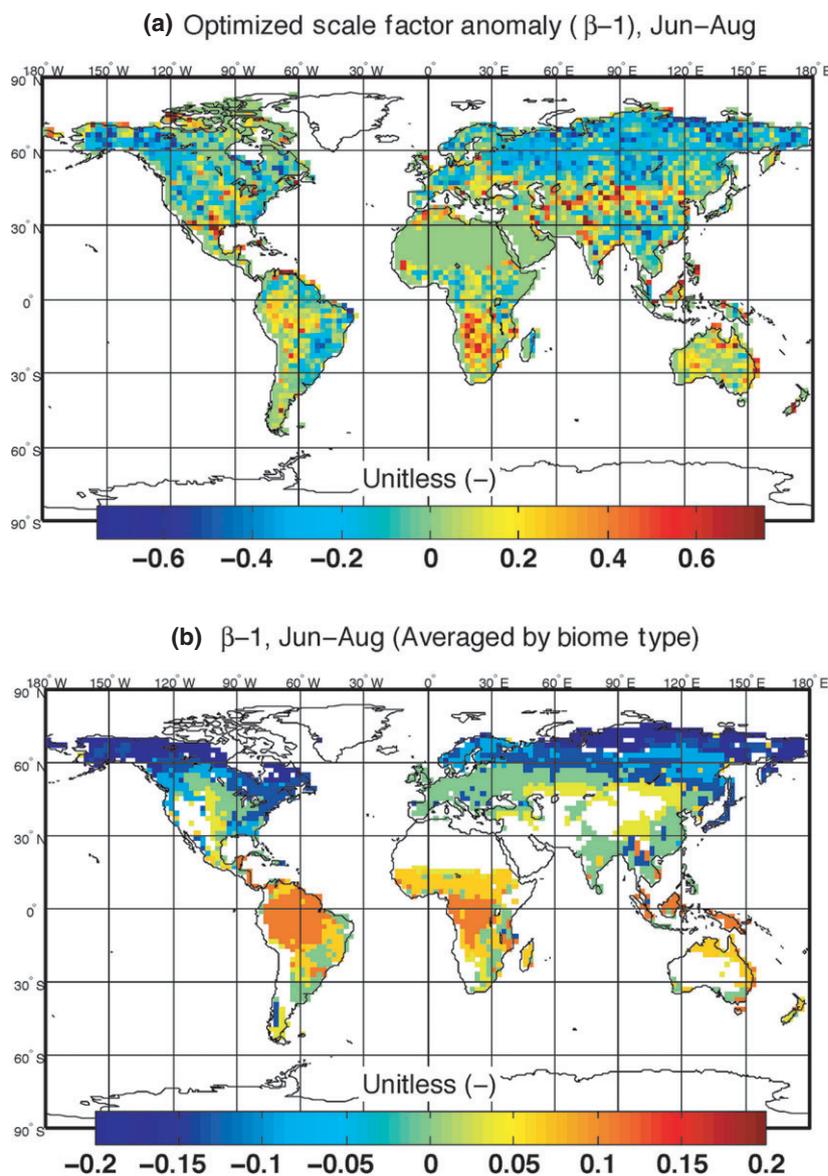


Fig. 7 Map of optimized scale factor anomaly (unitless, expressed as deviation from initial value of 1) averaged from June–August 2010 at (a) grid scale and (b) biome scale (according to IGBP biomes in Fig. 1). Negative values indicate that the ensemble average of TRENDY models overestimate GPP compared to SIF on average (vice-versa for positive values; e.g., -0.1 means models overestimate GPP by 10%). (Color in the online version)

that EB biomes are the most productive biomes in the world, these potential errors in model seasonal cycles have important implications for interpreting the global carbon budget. It is therefore critical to evaluate these estimates against flux tower data, as discussed below.

Comparison to flux tower data

N. America and Europe. Seasonal averages of G-pri and G-opt are compared against flux tower data (G-ft) from 49 sites in N. America and Europe in Fig. 10. GPP is

grouped by biome type, with flux tower biomes defined in Table S1. Here we present results in which G-pri and G-opt are sampled at flux tower pixels since flux tower sites are generally not representative of the large scale (e.g., Cescatti *et al.*, 2012). G-pri represents the mean from 2000 to 2009, G-opt from 2009 to 2012, and G-ft from 2009 to 2010. We also weigh grid-scale estimates by fractional biome coverage to help account for sub-grid heterogeneity, which can be significant especially in more developed and managed regions such as Europe. More details on the analysis of flux tower data and

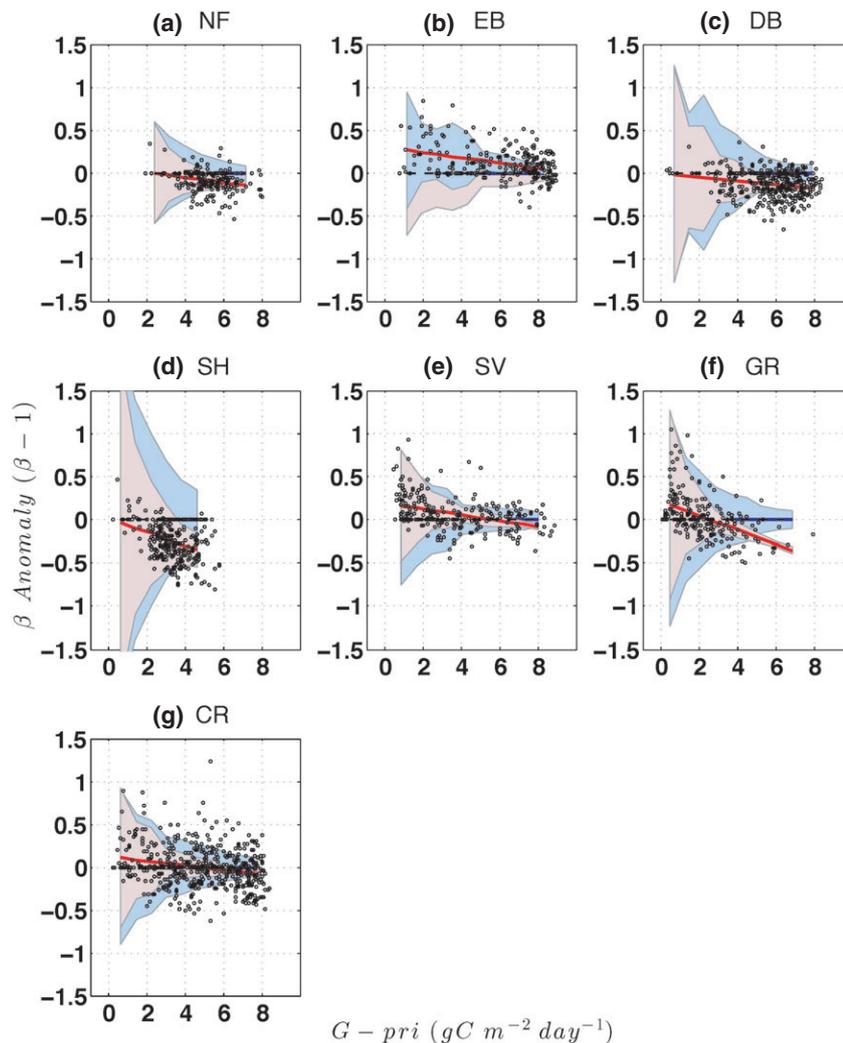


Fig. 8 Scatter plots of optimized scale factor anomaly vs. ensemble average GPP from TRENDY models ($G\text{-pri}$), as a function of biome type (panels a–g, based on dominant biome type in Fig. 1). Blue and red lines indicate prior and posterior least squares regression lines respectively (blue line sits of x-axis since prior scale factor is unity everywhere). Blue and red shading indicate aggregated prior and posterior uncertainty. (Color in the online version)

comparison to $G\text{-pri}$ and $G\text{-opt}$ at different scales is included in Text S5.

$G\text{-pri}$ and $G\text{-opt}$ are generally well correlated with $G\text{-ft}$ ($R > 0.95$ in NF, DB, GR and CR and ~ 0.60 in SH) despite the different scales of comparison (e.g., ~ 1 km vs. ~ 200 km), indicating seasonal phase is well represented by models. Growing season width is reduced in $G\text{-opt}$ in DB and CR biomes and in better agreement with $G\text{-ft}$. This difference is most noticeable in autumn as an enhanced rate of decreasing GPP, suggesting models underestimate the rate of leaf senescence compared to SIF and $G\text{-ft}$.

In general, seasonal amplitude of $G\text{-pri}$ and $G\text{-opt}$ are statistically similar to each other and, due to high variability across flux towers, to $G\text{-ft}$. However, there is a clear tendency to underestimate amplitude of $G\text{-ft}$ (with

the exception of SH biomes), particularly in GR biomes where $G\text{-ft}$ exceeds $G\text{-pri}$ and $G\text{-opt}$ by nearly a factor of 3 (although site-to-site flux tower variability is high; see Fig. S14) and in DB biomes where site-to-site variability is low. Amplitude differences scale to some extent with fractional biome coverage, in particular in GR biomes where amplitude errors are large but grid scale areal coverage is only 7.4% on average. Amplitude errors are smaller and less statistically significant in other biomes, whose fractional coverage range from 25% to 65%. In general subgrid heterogeneity is high such that differences are attributable primarily to representation error.

Divergence of $G\text{-opt}$ magnitude and uncertainty from $G\text{-pri}$ is related to sampling coverage and SNR of SIF data. $G\text{-opt}$ has improved seasonality in DB and CR bi-

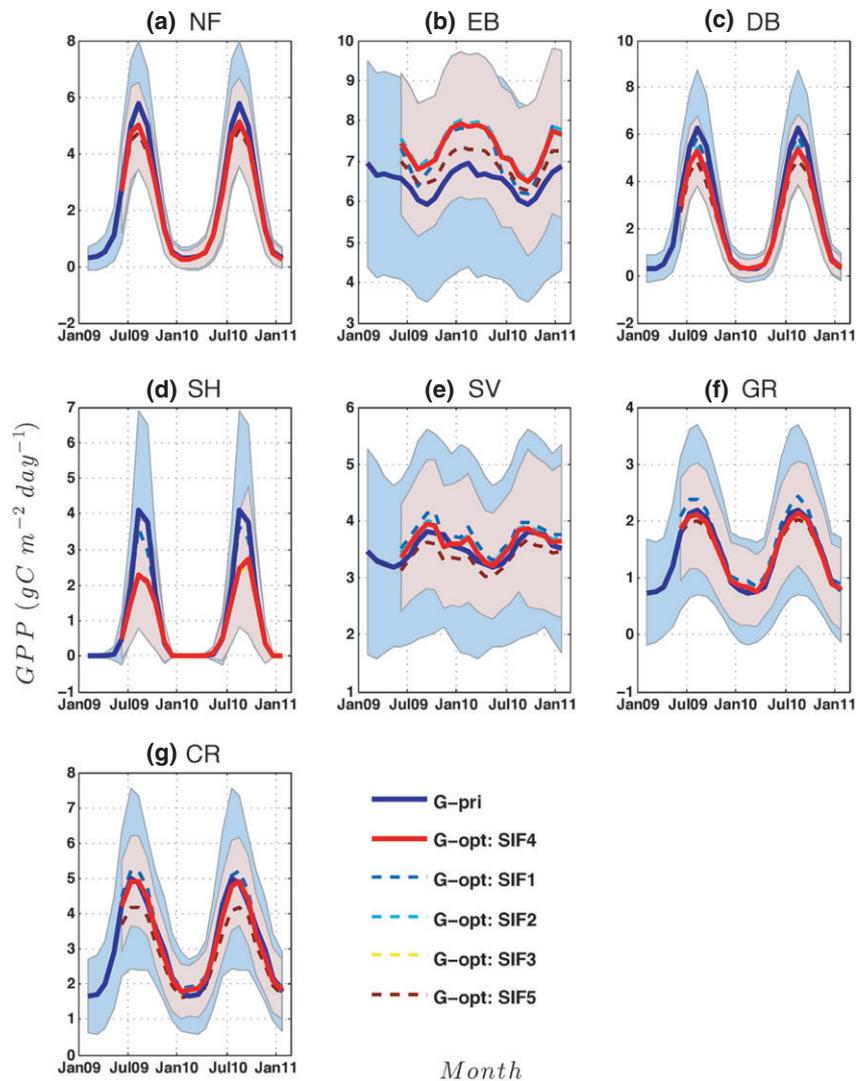


Fig. 9 Seasonal GPP from 2009–2010 for the ensemble average of TRENDY models (G-pri, solid blue) and optimal combination of TRENDY models and GOSAT SIF (G-opt, solid red) as a function of biome (panels a–g, based on dominant biome type in Fig. 1). Blue and red shading indicate aggregated prior and posterior uncertainty. Dashed lines represent estimates of G-opt from different SIF scaling strategies (SIF1–5; see Table 1). (Color in the online version)

omes likely a result of high sampling density, which averages to about 91.4 samples per site/year/biome in DB biomes and 278 samples per site/year/biome in CR biomes. Strong uncertainty reduction in these biomes is driven in part by high SNR (smaller error bars in seasonal plots of SIF in Fig. S18). Conversely, unchanged seasonality and low uncertainty reduction in NF, GR, and SH biomes stems from relatively low sampling density (56.9, 86.5, and 81.6 samples per year per site, respectively) and low SNR (larger error bars in Fig. S18).

Overall, these results are consistent with our understanding of the key environmental controls for seasonal phase and amplitude in extra-tropical forests. The seasonal phase, which is characterized by high GPP in the

middle of the growing season (~July) and low GPP outside the growing season (~January), is driven primarily by large-scale seasonal climate drivers such as solar insolation. Since insolation is coherent over large spatial scales it has similar effects at the local scale of flux towers and large scale of models. Amplitude, in contrast, is determined more by local factors such as plant functional type, soil characteristics, topography, and management practices – all of which have large subgrid variability which tend to average out at scales predicted by models and sampled by GOSAT. This explains low G-pri and G-opt amplitude relative to flux towers, giving increased confidence that remote sensing and model based approaches provide realistic estimates of landscape-scale flux and uncertainty.

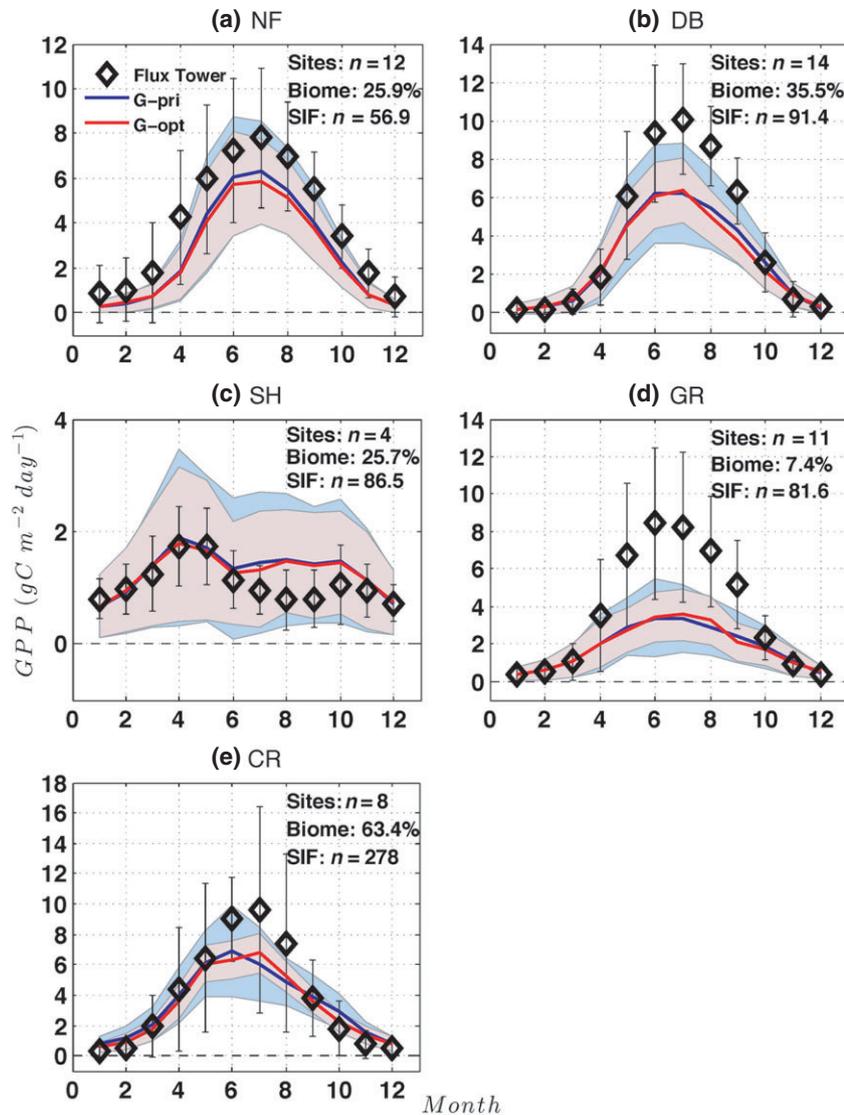


Fig. 10 Seasonal cycles of GPP from flux towers (G-ft, diamonds), ensemble average of TRENDY models (G-pri, blue), and optimal estimates of TRENDY models and GOSAT SIF (G-opt, red), grouped by biome type for (a) needleleaf forest (NF), (b) deciduous broadleaf forest (DB), (c) shrublands (SH), (d) grasslands (GR), and (e) croplands (CR). Grid scale estimates are based on sampling at flux tower locations and weighting by fractional biome coverage. G-ft is based on seasonal averages from 2009 to 2010, G-pri from 2000 to 2009, and G-opt from 2009 to 2012. Error bars on flux tower data equal 1-sigma standard deviations across sites. Labeling on top right of panels indicates number of flux towers used per biome (Sites), average percent coverage of biomes within flux tower pixels (Biome), and number of SIF samples per year/site/biome used for data assimilation in G-opt (SIF). Blue and red shading represent prior and posterior uncertainty. (Color in the online version)

These comparisons highlight some of the reasons why an optimal estimation approach is preferable to separate estimates by models or SIF. While GOSAT provides near global coverage of SIF and a direct measurement of subgrid scale variability at high resolution, there is often unrealistic variability and/or negative values in cases when sampling coverage is reduced and precision error is high, especially during winter and in cases where GPP signals are weak. Model ensembles provide a strong constraint on

seasonal phase in northern latitudes due to spatial coherence of large-scale environmental factors such as insolation and temperature, which are relatively straightforward to model, but typically underestimate subgrid scale variability due to unresolved and/or poorly understood processes.

Amazon Basin. Here, we compare climatological GPP estimates and flux tower data in the Amazon Basin. GPP records range from 2000–2009 in G-pri, 2009–2012

in G-opt, and 1999–2006 for 7 flux tower sites (see Table S1). We use fractional biome weighting as before but average G-pri and G-opt across all available grid points in Amazonia to further avert the issue of non-temporal overlap. Results are shown in Fig. 11 (seasonal averages using only flux tower pixels can be found in Fig. S19b).

G-ft indicates peak GPP from November–February, roughly coinciding with the wet season in Amazonia, with a mean seasonal amplitude of $\sim 2.5 \text{ g C m}^{-2} \text{ d}^{-1}$. G-pri and G-opt are easily within the statistical uncertainty of G-ft, which is very high. Seasonal amplitude is underestimated by nearly a factor of 2 in G-pri, with seasonal phase (on average) delayed by ~ 1 month and GPP recovery at the end of the dry season (\sim September) too slow. Low GPP during the dry season is aligned with G-ft but peak wet season GPP is substantially underestimated. These findings are generally true for individual TRENDY models as well (also shown in Fig. 11), although in the case of LPJ-GUESS seasonality is reversed relative to flux towers while OCN is relatively flat. The TRENDY seasonal average is unchanged when using a shorter overlapping time period (i.e., 2000–2006).

Subgrid sampling by SIF leads to substantial improvement in both the seasonal phase and amplitude of GPP, including faster rates of seasonal transition from wet-to-dry season (\sim May) and from dry-to-wet season (\sim September) and uncertainty reductions of 30–40% throughout the year. With the exception of LPJ, most models don't reproduce the seasonal phase of G-ft. Similar to evidence from G-ft and SIF that models underestimate rates of leaf senescence in temperate DB forests, poor timing in the tropics may be partly an issue of models unable to capture the rate and timing of leaf-flush just before the onset of the wet season (Zeri *et al.*, 2014). These are key examples of how SIF can help inform DGVMs.

Discussion

We demonstrate an optimal estimation technique for constraining seasonal and spatial GPP variability at global scale that exploits information from DGVM ensembles and satellite measurements of SIF. We use an ensemble of eight state-of-the-art DGVMs and global estimates of GPP inferred from GOSAT SIF retrievals. The model ensemble provides an improved estimate of GPP compared to individual models as well as a process-based estimate of uncertainty. SIF data augments this information by observing the rate of global photosynthetic activity at subgrid scale, which improves estimates of seasonal phase and amplitude with respect to flux towers and semi-empirical data in

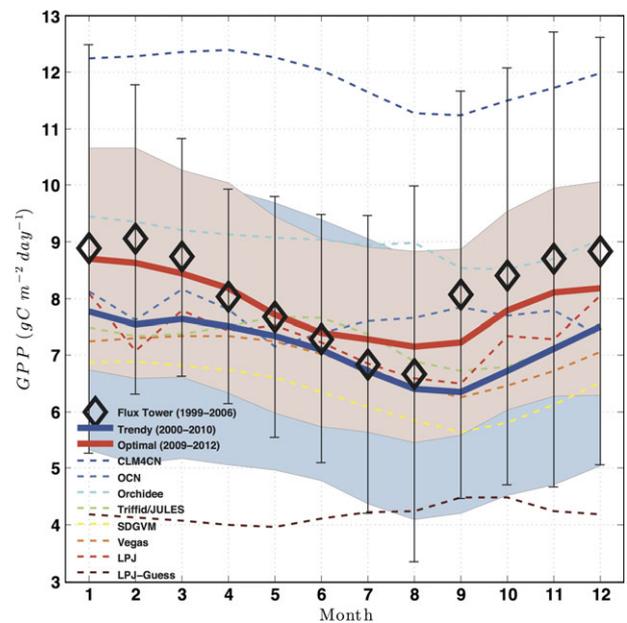


Fig. 11 Estimates of seasonal GPP averaged across the Amazon Basin, representing evergreen broadleaf (EB) and savannah (SV) biomes. Flux tower observations (diamonds) represent the average across seven flux tower sites (see Table S1). Grid scale estimates represent the average over the entire Amazon Basin and based on seasonal averages from the ensemble average of TRENDY models from 2000 to 2009 (G-pri, blue), optimal estimates from TRENDY models and GOSAT SIF from 2009 to 2012 (G-opt, solid red), and individual TRENDY models from 2000 to 2009 (dashed). Error bars on flux tower data equal 1-sigma standard deviations across sites. Blue and red shading represents grid scale averages of prior and posterior uncertainty. (Color in the online version)

N. America, Europe, and S. America, reduces *a priori* uncertainty of process models by up to 40–70% in high productivity tropical and temperate regions, and redistributes global productivity from northern latitudes to the tropics through reduced GPP in northern forests ($\sim 3.6 \text{ Pg C yr}^{-1}$) and enhanced GPP in tropical forests ($\sim 3.7 \text{ Pg C yr}^{-1}$).

Theory and experiments lead to a prediction of CO_2 fertilization scaled to GPP, with peak productivity in the tropics (Norby & Zak, 2011; see Fig. 6). Our findings suggest that the partitioning of GPP from northern to tropical forests is underestimated in DGVMs, and that tropical CO_2 fertilization is stronger than currently predicted. These results have important implications for our understanding of the global carbon cycle and its response to climate and environmental changes, and warrants further investigation.

The assimilation of satellite SIF to reduce uncertainty of predicted GPP provides an empirical constraint that helps benchmark models, which may lead to process understanding. For example, models that overestimate

growing season GPP (relative to SIF) in the extra-tropics (e.g., Fig. 6a) may be overly sensitive to environmental factors such as (CO₂) fertilization and nitrogen deposition, climate forcing such as temperature and precipitation (Poulter *et al.*, 2011). In contrast, systematic underestimates of seasonal amplitude and delayed timing of dry season GPP loss/recovery cycles in Amazonia (e.g., Fig. 11) relates to the way models represent seasonal interactions of GPP, precipitation, and sunlight (Baker *et al.*, 2008; Restrepo-Coupe *et al.*, 2013; Harper *et al.*, 2014). As seasonal interactions of water stress and SIF in southern Amazonia are correlated with column (CO₂) variability (Lee *et al.*, 2013; Parazoo *et al.*, 2013), it is important for models to correctly represent these processes in the context of the global carbon cycle.

Despite these key benefits of assimilating SIF data there are several opportunities for improvement. In particular, we note the absence of any correction for biases associated with undersampling of diffuse light under cloudy conditions, which enhances sampling biases at high latitudes due to increases photosynthetic efficiency (e.g., Mercado *et al.*, 2009), correction for GPP response to light saturation at subdaily scale, or rigorous assessment of errors related to estimates of diurnal GPP. We are working to constrain nonlinearity of light response through use of data from flux towers. Furthermore, we expect cases of low SNR to be reduced with future satellites such as OCO-2, which are expected to enhance the quantity and quality of these data, improving subgrid scale sampling and precision that we have shown is critical for reducing GPP uncertainty.

Acknowledgements

We thank M. Reichstein for providing the MPI-BGC dataset. We acknowledge all MODIS land product science team members for providing an invaluable public dataset. The improved MOD17 GPP data were provided by the Numerical Terradynamic Simulation Group at the University of Montana. We thank the Global Carbon Project and TRENDY modelers for contributing model output: S. Sitch and C. Huntingford (TRIFID/JULES), B. Poulter (LPJ), A. Ahlström (LPJ-GUESS), M. Lomas (SDGVM), S. Levis (CLM4CN), S. Zaehle (OCN), N. Viovy (Orchidee), and N. Zeng (VEGAS). We also give special thanks to S Sitch and P Friedlingstein for organizing TRENDY and for valuable discussions regarding this work. This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), CarboEuropeIP, CarboItaly, CarboMont, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCAN), GreenGrass, KoFlux, LBA (Partial funding provided by FAPESP-0858120-3), NECC, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO,

iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada, and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California – Berkeley, University of Virginia. Funding for US-PFa is provided by NSF Award #0845166. We acknowledge the receipt of data from the Groundhog River site (CA-OMW) collected under the auspices of the Fluxnet-Canada Research Network (FCRN, 2002-07) and the Canadian Carbon Program (CCP, 2007-11) with funding support from the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), the Canadian Natural Sciences and Engineering Research Council (NSERC), the Canadian Forest Service (CFS), and the Ontario Ministry of Environment. We give special thanks to H. McCaughey (CA-OMW), R. Huc (FR-FBN), N. Buchmann (CH-Lae, CH-Oe2, CH-Dav), N. Restrepo-Coupe (LBA), G. Bohrer (US-UMd), A. Desai (US-PFa) for contributing flux tower data and providing valuable discussions. This research was funded by the NASA Atmospheric CO₂ Observations from Space (ACOS) program (grant number NNX10AT42G), and carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA © 2013.

References

- Ahlström A, Miller PA, Smith B (2012a) Too early to infer a global NPP decline since 2000. *Geophysical Research Letters*, **39**, L15403.
- Ahlström A, Schurgers G, Arneth A, Smith B (2012b) Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environmental Research Letters*, **7**, 044008.
- Arora VK, Boer GJ, Friedlingstein P *et al.* (2013) Carbon-concentration and carbon-climate feedbacks in CMIP5 earth system models. *Journal of Climate*, **26**, 5289–5314.
- Baker IT, Prihodko L, Denning AS, Goulden M, Miller S, Da Rocha HR (2008) Seasonal drought stress in the Amazon: reconciling models and observations. *Journal of Geophysical Research*, **113**, G00B01, 10 pp.
- Baker IT, Harper AB, da Rocha HR (2013) Surface ecophysiological behavior across vegetation and moisture gradients in tropical South America. *Agricultural and Forest Meteorology*, **182–183**, 1770188.
- Beer C, Reichstein M, Tomelleri E *et al.* (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, **329**, 834–838.
- Canadell JG, Ciais P, Gurney K, Le Quééré C, Piao S, Raupach MR, Sabine CL (2011) An international effort to quantify regional carbon fluxes. *Eos, Transactions American Geophysical Union*, **92**, 81–82.
- Cescatti A, Marcolla B, Santhana Vannan SK *et al.* (2012) Intercomparison of MODIS albedo retrievals and *in situ* measurements across the global FLUXNET network. *Remote sensing of environment*, **121**, 323–334.
- Clark DB, Mercado LM, Sitch S *et al.* (2011) The joint UK land environment simulator (JULES), model description. II. Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, **4**, 701–722.
- Collins M (2007) Ensembles and probabilities: a new era in the prediction of climate change. *Philosophical Transaction of the Royal Society A*, **365**, 1957–1970.
- Cramer W, Bondeau A, Woodward FI *et al.* (2001) Global response of terrestrial ecosystem structure and function to CO₂ and climate change: results from six dynamic global vegetation models. *Global Change Biology*, **7**, 357–373.
- Crisp D, Fisher BM, O'Dell C *et al.* (2012) The ACOS CO₂ retrievals algorithm – Part II: global XCO₂ data characterization. *Atmospheric Measurement Techniques*, **5**, 687–707.
- Damm A, Elbers J, Erler A *et al.* (2010) Remote sensing of sun-induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP). *Global Change Biology*, **16**, 171–186.
- Daumard F, Champagne S, Fournier A, Goulas Y, Ounis A, Hanocq J-F, Moya I (2010) A field platform for continuous measurement of canopy fluorescence. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 3358.
- Fleixas J, Escalona J, Evain S, Gulias J, Moya I, Osmond C, Medrano H (2002) Steady-state chlorophyll fluorescence (Fs) measurements as a tool to follow variations of

- net CO₂ assimilation and stomatal conductance during water-stress in C₃ plants. *Physiologia Plantarum*, **114**, 231–240.
- Frankenberg C, Fisher JB, Worden J *et al.* (2011) New global observations of the terrestrial carbon cycle from GOSAT: patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, **38**, L17706. doi:10.1029/2011GL048738.
- Friedlingstein P, Cox P, Betts R (2006) Climate-carbon cycle feedback analysis: results from the C4MIP Model Intercomparison. *Journal of Climate*, **19**, 3337–3353.
- Guanter L, Frankenberg C, Dudhia A *et al.* (2012) Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. *Remote Sensing of Environment*, **121**, 236–251.
- Harper A, Baker IT, Denning AS, Randall DA, Dazlich D, Branson M (2014) Impact of evapotranspiration on dry season climate in the Amazon forest. *Journal of Climate*, **27**, 571–591.
- Huntzinger DN, Post WM, Wei Y *et al.* (2012) North American Carbon Program (NACP) regional interim synthesis: terrestrial biospheric model intercomparison. *Ecological Modelling*, **232**, 144–157.
- Joiner JY, Yoshida Y, Vasilkov AP, Yoshida Y, Corp LA, Middleton EM (2011) First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences*, **8**, 637–651.
- Joiner J, Guanter L, Lindström R (2013) Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques Discussions*, **6**, 3883–3930.
- Jung M, Reichstein M, Margolis HA *et al.* (2011) Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, **116**, G00J07.
- Knorr W (2009) Is the airborne fraction of anthropogenic CO₂ emissions increasing? *Geophysical Research Letters*, **36**, doi: 10.1029/2009GL040613.
- Krinner G, Viovy N, de Noblet-Ducoudré N *et al.* (2005) A dynamic global vegetation model for studies of the coupled atmosphere biosphere system. *Global Biogeochemical Cycles*, **19**, 1–33.
- Lasslop G, Reichstein M, Papale D *et al.* (2010) Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biology*, **16**, 187–208.
- Le Quéré C, Raupach MR, Canadell JG *et al.* (2009) Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, **17**, 831–836.
- Le Quéré C, Andres RJ, Boden R *et al.* (2013) The global carbon budget 1959–2011. *Earth System Science Data*, **5**, 165–185.
- Lee JE, Frankenberg C, van der Tol C *et al.* (2013) Forest productivity and water stress in Amazonia: observations from GOSAT chlorophyll fluorescence. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 9.
- Luo YQ, Randerson JT, Abramowitz G *et al.* (2012) A framework for benchmarking land models. *Biogeosciences*, **9**, 3857–3874.
- Maxwell K, Johnson G (2000) Chlorophyll fluorescence – a practical guide. *Journal of Experimental Botany*, **51**, 659–668.
- Mercado LM, Bellouin N, Sitch S, Boucher O, Huntingford C, Wild M, Cox PM (2009) Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, **458**, 1014–1017.
- Myneni RB, Yang W, Nemani RR *et al.* (2007) Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Sciences*, **104**, 4820–4823.
- Norby RJ, Zak DR (2011) Ecological lessons from free-air CO₂ enrichment (FACE) experiments. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 181–203.
- Olsen SC, Randerson JT (2004) Differences between surface and column atmospheric CO₂ and implications for carbon cycle research. *Journal of Geophysical Research*, **109**, D02301, 11 pp.
- O'Neill B, Melnikov N (2008) Learning about parameter and structural uncertainty in carbon cycle models. *Climate Change*, **89**, 23–44.
- Papale D, Reichstein M, Aubinet M (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, **3**, 571–583.
- Parazoo NC, Bowman K, Frankenberg C *et al.* (2013) Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of XCO₂ and chlorophyll fluorescence from GOSAT. *Geophysical Research Letters*, **40**, 2829–2833.
- Piao S, Sitch S, Ciais P *et al.* (2013) Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO₂ trends. *Global Change Biology*, **19**, 2117–2132.
- Poulet B, Frank DC, Hodson EL, Zimmermann NE (2011) Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO₂ airborne fraction. *Biogeosciences*, **8**, 2027–2036.
- Rascher U, Agati G, Alonso L *et al.* (2009) CEFLES 2: the remote sensing component to quantify photosynthetic efficiency from the leaf to the region by measuring sun-induced fluorescence in the oxygen absorption bands. *Biogeosciences*, **6**, 1181–1198.
- Reichstein M, Falge E, Baldocchi D *et al.* (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, **11**, 1424–1439.
- Restrepo-Coupe N, da Rocha HR, Hutryra LR *et al.* (2013) What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network. *Agricultural and Forest Meteorology*, **182**, 128–144.
- Rienecker MM, Suarez MJ, Gelaro R *et al.* (2011) MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, **24**, 3624–3648.
- Schaefer K, Schwalm CR, Williams C *et al.* (2012) A model-data comparison of gross primary productivity: results from the North American Carbon Program site synthesis. *Journal of Geophysical Research*, **117**, G03010, doi:10.1029/2012JG001960
- Sitch S, Smith B, Prentice IC *et al.* (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161–185.
- Sitch S, Huntingford C, Gedney N *et al.* (2008) Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global Change Biology*, **14**, 2015–2039.
- Sitch S, Friedlingstein P, Gruber N (2013) Trends and drivers of regional sources and sinks of carbon dioxide over the past two decades. *Biogeosciences Discussions*, **10**, 1–65.
- Smith B, Prentice IC, Sykes MT (2001) Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Global Ecology and Biogeography*, **10**, 621–637.
- Suntharalingam P, Kettle AJ, Montzka SM, Jacob DJ (2008) Global 3-D model analysis of the seasonal cycle of atmospheric carbonyl sulfide: implications for terrestrial vegetation uptake. *Geophysical Research Letters*, **35**.
- Tarantola A. (2005) *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Pennsylvania.
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485–498.
- Thornton PE, Lamarque J-F, Rosenbloom NA, Mahowald NM (2007) Influence of carbon–nitrogen cycle coupling on land model response to CO₂ fertilization and climate variability. *Global Biogeochemical Cycles*, **21**, GB4018, doi: 10.1029/2006GB002868.
- Van der Tol C, Verhoef W, Timmermans J, Verhoef A, Su Z (2009) An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences*, **6**, 3109–3129.
- van der Werf GV, Randerson JT, Giglio L *et al.* (2010) Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics*, **10**, 11707–11735.
- Zaehle S, Friend AD, Friedlingstein P, Dentener F, Peylin P, Schulz M (2010) Carbon and nitrogen cycle dynamics in the O-CN land surface model. II. The role of the nitrogen cycle in the historical terrestrial C balance. *Global Biogeochemical Cycles*, **24**, GB1006, doi:10.1029/2009GB003522.
- Zeng N, Qian H, Roedenbeck C, Heimann M (2005) Impact of 1998–2002 midlatitude drought and warming on terrestrial ecosystem and the global carbon cycle. *Geophysical Research Letters*, **32**, L22709. doi:10.1029/2005GL024607.
- Zeri M, Sá LDA, Manzi AO *et al.* (2014) Variability of carbon and water fluxes following climate extremes over a tropical forest in southwestern Amazonia. *PLoS ONE*, **9**, e88130.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

File S1. Information about (1) SIF scaling technique, (2) satellite fluorescence sampling coverage, (3) observation system simulation experiments, and (4) flux tower data.