Agricultural and Forest Meteorology 214–215 (2015) 416–429

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Improving the performance of remote sensing models for capturing intra- and inter-annual variations in daily GPP: An analysis using global FLUXNET tower data

M. Verma[a,b,*], M.A. Friedl[a], B.E. Law[c], D. Bonal[d], G. Kiely[e], T.A. Black[f], G. Wohlfahrt[g,h], E.J. Moors[i], L. Montagnani[j,k], B. Marcolla[l], P. Toscano[m], A. Varlagin[n], O. Roupsard[o,p], A. Cescatti[q], M.A. Arain[r], P. D'Odorico[s]

[a] Department of Earth and Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA
[b] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA
[c] Earth Systems Science Division, Oregon State University, Corvallis, OR 97331, USA
[d] INRA Nancy, UMR EEF-Université de Lorraine/INRA, 54280 Champenoux, France
[e] Environmental Research Institute, Civil and Environmental Engineering Department, University College, Cork, Ireland
[f] Faculty of Land and Food Systems, 135-2357 Main Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4
[g] Institute of Ecology, University of Innsbruck, Sternwartestr. 15, 6020 Innsbruck, Austria
[h] European Academy of Bolzano, Drususallee, 1, 39100 Bolzano, Italy
[i] Climate Change & Adaptive Land and Water Management, Alterra Wageningen UR, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[j] Faculty of Science and Technology, Free University of Bolzano-Bozen, Italy
[k] Forest Services, Autonomous Province of Bolzano, Bolzano, Italy
[l] Sustainable Agro-ecosystems and Bioresources Department, Fondazione Edmund Mach, via Mach, 1 I-38010 S. Michele all'Adige (TN), Italy
[m] Institute of Biometeorology (IBIMET – CNR), via G.Caproni 8, 50145 Firenze, Italy
[n] A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Lenisky pr.33, Moscow 119071, Russia
[o] CIRAD, UMR Eco&Sols (Ecologie Fonctionnelle & Biogéochimie des Sols & Agroécosystèmes), 34000 Montpellier, France
[p] CATIE (Tropical Agricultural Centre for Research and Higher Education), 93-7170 Turrialba, Costa Rica
[q] European Commission, Joint Research Center, Institute for Environment and Sustainability, Ispra, Italy
[r] School of Geography and Earth Sciences, McMaster Centre for Climate Change, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
[s] Institute of Agricultural Sciences, ETH Zurich, LFW C56 Universitatsstr. 2, 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 28 May 2015
Received in revised form 6 September 2015
Accepted 7 September 2015

Keywords:
Gross primary productivity
Remote sensing
Modeling
FLUXNET
Seasonal
Lagged effects

ABSTRACT

Accurate and reliable estimates of gross primary productivity (GPP) are required for monitoring the global carbon cycle at different spatial and temporal scales. Because GPP displays high spatial and temporal variation, remote sensing plays a major role in producing gridded estimates of GPP across spatiotemporal scales. In this context, understanding the strengths and weaknesses of remote sensing-based models of GPP and improving their performance is a key contemporary scientific activity. We used measurements from 157 research sites (~470 site-years) in the FLUXNET "La Thuile" data and compared the skills of 11 different remote sensing models in capturing intra- and inter-annual variations in daily GPP in seven different biomes. Results show that the models were able to capture significant intra-annual variation in GPP (Index of Agreement = 0.4–0.80) in all biomes. However, the models' ability to track inter-annual variation in daily GPP was significantly weaker (IoA < 0.45). We examined whether the inclusion of different mechanisms that are missing in the models could improve their predictive power. The mechanisms included the effect of sub-daily variation in environmental variables on daily GPP, factoring-in differential rates of GPP conversion efficiency for direct and diffuse incident radiation, lagged effects of environmental variables, better representation of soil-moisture dynamics, and allowing spatial variation in model parameters. Our analyses suggest that the next generation remote sensing models need better representation of soil-moisture, but other mechanisms that have been found to influence GPP in site-level studies may not have significant bearing on model performance at continental and global scales. Understanding

* Corresponding author at: Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA.
E-mail address: Manish.K.Verma@jpl.nasa.gov (M. Verma).

http://dx.doi.org/10.1016/j.agrformet.2015.09.005
0168-1923/

the relative controls of biotic vis-a-vis abiotic factors on GPP and accurately scaling up leaf level processes to the ecosystem scale are likely to be important for recognizing the limitations of remote sensing model and improving their formulation.

1. Introduction

Terrestrial gross primary productivity (GPP) is the largest component flux of the global carbon cycle (Solomon et al., 2007; Beer et al., 2010), and along with respiration, drives fluctuation in atmospheric CO_2 concentration (Keeling et al., 1995; Bonan, 1995; Schimel, 2007). Because GPP shows high spatial and temporal variation, remote sensing plays an important role in modeling gridded, temporally frequent estimates of GPP. Light-use efficiency (LUE) remains the most widely used approach (Law and Waring, 1994; Running et al., 2004; Yuan et al., 2014a) to model GPP from remote sensing, but other models have also been proposed in the last decade (Yang et al., 2007; Sims et al., 2008). Availability of eddy covariance data has facilitated extensive evaluation of remote sensing models and several studies have compared GPP estimated from remote sensing with tower-derived GPP. These studies can be broadly classified into two types.

The first type of studies proposed new models (different for LUE-based) to estimate GPP from remote sensing and showed how well the models estimated GPP (Olofsson et al., 2008; Sims et al., 2008; Schubert et al., 2012). Some of the models are simpler than the LUE approach and have been justified based on the claim that one or more parametrizations used in the LUE approach are not required (Gitelson et al., 2006; Sims et al., 2008; Jung et al., 2008; Jahan and Gan, 2009; Ueyama et al., 2010; Wu et al., 2011; Sjöström et al., 2011; Sakamoto et al., 2011; Hashimoto et al., 2012). Other models follow more complex, data-driven neural network or model-tree ensemble to estimate GPP (Yang et al., 2007; Xiao et al., 2010). The underlying assumption here is that the nonlinear, flexible structure of machine learning algorithms is better suited than the physiologically based LUE approach to model relationship between remotely sensed predictor variables and GPP.

The second type of studies have focused specifically on the LUE method (Monteith, 1972; Xiao et al., 2004; Mahadevan et al., 2008), which is followed in the MODIS GPP product (MOD17; Running et al., 2004). These studies have noted disagreement between modeled and tower-derived GPP and have identified or postulated the following reasons for the disagreement: the assumption that parameters remain constant within biomes is not correct and especially maximum LUE shows large spatial variation (Turner et al., 2006; Lin et al., 2011; Sjöström et al., 2013); inaccuracies in upstream meteorological and remote sensing inputs cause errors in estimated GPP (Zhao et al., 2005, 2006; Heinsch et al., 2006; Kanniah et al., 2009; Propastin et al., 2012); current parameterizations do not capture the dynamics of soil moisture well (Leuning et al., 2005; Yuan et al., 2007); sub-pixel heterogeneity is not represented well (Chasmer et al., 2009); the effects of foliar N concentration and disturbance on productivity are not captured (Makela et al., 2008; Cook et al., 2008); and remotely sensed data does not provide accurate information of rapid phenological changes in spring and fall (Coops et al., 2007). However, the actual potential of these mechanisms in improving model performance at large spatial and temporal scales has not been investigated.

In this study, we synthesize findings reported across the different types of studies highlighted above and analyze the potential of "big-leaf" models for estimating intra- and inter-annual variation in daily GPP (Figure S1 shows examples of the two types of variability) by taking two steps. First, we resolve claims about the relative suitability of remote sensing models from a suite of models that include LUE-based and neural-network type models. Thus, in this step we test the validity of the conclusions, drawn in different studies highlighted above, at the global scale by asking "Which of the currently available remote sensing models provide the most accurate representation of intra- and inter-annual variation in daily GPP (Falge et al., 2002)?" To do this, we compare 11 different remote sensing models using data from 157 sites (~470 site-years) in seven different biomes. We specifically focus on model performance that is consistent both within and across biomes. The large dataset and the spectrum of models allow a systematic understanding of the power of these models.

Next, we determine the potential of different mechanisms that are missing in the current models, but are known to affect GPP, in improving the predictive power of the models. To do this, we choose the best-performing model from the step above and determine whether the inclusion of missing mechanisms improved the agreement between modeled and tower-derived GPP in intra- and inter-annual analysis. Of the several mechanisms suggested in the studies noted above, we focused on five that have been widely reported and can be examined with available data. We hypothesize that incorporation of each of the following five mechanisms will improve model performance and test the hypothesis by comparing model performance with and without the inclusion of the mechanism.

(i) *Model calibration at sub-daily time scale*: Photosynthesis can show non-linear rapid responses to sub-daily variations in environmental variables, such as photosynthetically active radiation (PAR) and vapor pressure deficit (VPD; Law et al., 2001). In the case of dynamic models it has been suggested that models perform better if run at sub-daily time scale (Schwalm et al., 2010; Medvigy et al., 2010).

(ii) *Separation of direct and diffuse radiation*: Light use efficiency for diffuse radiation is significantly higher than direct radiation (Gu et al., 2002; Dai et al., 2004; Alton et al., 2007). Variation in direct and diffuse radiation has been shown to be an important driver that affects day-to-day variability, especially during peak growing season (Jenkins et al., 2007) and it has been suggested that direct and diffuse radiation should be treated separately in models (McCallum et al., 2009; Yuan et al., 2014a).

(iii) *Lagged effects and model calibration at a 15-day scale*: Lagged effects have been postulated and empirically confirmed in site studies (Thomas et al., 2009; Zielis et al., 2014). Both environmental variables such as soil moisture, and biotic variables such as carbohydrate reserves are thought to produce lagged effects and influence productivity. It has been postulated that models may perform better at coarser temporal scale (e.g. 15-day) because fast variations get smoothed out and lagged effects are captured at longer time scales (Dietze et al., 2011).

(iv) *Better representation of soil-moisture dynamics*: Soil-moisture dynamics exerts an important control on GPP (Irvine et al., 2004). Remote sensing models use VPD or remotely sensed land surface water index as a surrogate for soil moisture. However, in many situations these metrics fail to capture actual variation in root-zone soil moisture.

(v) *Calibration of model parameters at each site*: Models assume that parameters are biome specific. However, for enzyme kinetic model it has been shown that parameters vary as much within

as among plant functional types (PFT) (Groenendijk et al., 2011) and efforts are being made to spatialize parameters of remote sensing models (Horn and Schulz, 2011; Madani et al., 2014).

2. Materials and methods

2.1. Data from FLUXNET

The FLUXNET 'La Thuile' data contains eddy-covariance measurements of net ecosystem CO_2 exchange (NEE) from more than 250 sites worldwide (~950 site-years). Ecosystem respiration is modeled by a temperature response function (Reichstein et al., 2005) and then subtracted from measured NEE. The temperature response function is locally calibrated using only nighttime data and it is assumed that the calibrated relationship applies during the daytime. Best methods for modeling respiration are a source of ongoing debate among the FLUXNET community. At present, however, the "La Thuile" dataset provides the best available global record for a study like ours. To reduce the probability of using poor quality data, following Richardson et al. (2010), we identified a subset of 157 sites with a total 466 site-years of data (Fig. 1; Table S1) where for each site-year two conditions were met: more than 95% of the days had daily GPP data, and the mean daily quality flag was greater than 0.75 (on a scale from 0 to 1). For each of these site-years, we extracted 30-minute and daily GPP, air temperature, VPD, PAR and their quality-flags. Of the total 157 sites, 21 are croplands (CRO), 10 are shrublands (CSH), 25 are deciduous broadleaf forests, 16 are evergreen broadleaf forests (EBF), 49 are evergreen needleleaf forests (ENF), 29 are grasslands (GRA), and 7 are savannas (SAV). CSH included both open and closed shrublands and SAV included woody savannas and savannas.

2.2. MODIS data

MODIS land products are available at 250, 500, and 1000-m spatial resolution (Justice et al., 2002). For this work we used the fraction of absorbed photosynthetically active radiation (FPAR; Myneni et al., 2002), and day and night land surface temperature products (LST; Wan et al., 2002), which are available at 8-day time steps and 1000-m spatial resolution. We also used nadir bidirectional reflectance distribution function (BRDF) adjusted surface reflectances (and associated quality flags) at 8-day temporal and 500-m spatial resolution (Schaaf et al., 2002) and computed the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and land surface water index (LSWI) from these data. In addition, we obtained land cover and phenology at 500-m spatial resolution from the MODIS Land Cover Type (Friedl et al., 2010) and Land Cover Dynamics Products (Zhang et al., 2003; Ganguly et al., 2010), and the 8-day MODIS GPP product (MOD17) at 1000-m spatial resolution (Running et al., 2004). We obtained all the MODIS data from Land Processes Distributed Active Archive Center (LP DAAC) of the Oak Ridge National Laboratory (ORNL).

To screen for low quality data, we used quality flags and removed the cases identified as having low reliability. For example, in BRDF corrected reflectance, we used quality assurance information from the MCD43A2 product and removed reflectance values where the BRDF inversion was performed with fewer than 7 total observations. Similarly, for the MODIS FPAR data, we removed values produced using the backup algorithm. The gaps created due to the removal of poor quality data were filled using locally weighted regression.

2.3. Remote sensing models of GPP

We evaluated 11 different remote sensing-based models in this study (Table 1). The simplest one was the *EVI-Linear* model, which uses 8-day EVI as the only predictor of GPP in a two-parameter linear model (Schubert et al., 2012; Hashimoto et al., 2012). This model functions as the "null model" in our study. The remaining models become progressively more complex by including additional components or nonlinear relationships, or both (Table 1). The *MOD17* (Running et al., 2004) and *VPRM* (Mahadevan et al., 2008) models follow the LUE approach. To address the issue of errors in upstream meteorological data we calibrated the MOD17 algorithm using meteorological data and GPP from towers (hereafter referred to as *MOD17-Tower*). For the neural network model we used the feed-forward approach with linear transfer function since it is the most effective model to approximate continuous function in static situations. We experimented with the parameters, such as the number of layers, neurons, and transfer function, of the neural network architecture and finally used a single hidden layer with 10 neurons, as it provided an optimum balance between the accuracy of prediction and the complexity of the model. The neural network model uses the same biotic and abiotic inputs as the *MOD17-Tower* model, but combines them in a flexible, nonlinear form.

For the first nine models listed in Table 1, we followed the approach used in previous studies (Heinsch et al., 2006; Sims et al., 2008; Xiao et al., 2010) and averaged 500- and 1000-m MODIS data over 7-by-7 and 3-by-3 pixel windows centered on tower locations, respectively. Studies comparing MODIS-based estimates with field data prefer using a window centered over field locations because averaging over many pixels reduces the effects of geo-location error, random noise, and error due to gridding (Wolfe et al., 2002; Tan et al., 2006). However, for many sites the area covered by the 7-by-7 pixels of 500-m and 3-by-3 of 1000-m is likely to be more than the typical footprint of flux towers, causing a potential mismatch between modeled and tower-derived GPP, especially at heterogeneous sites. To address this possibility we took two steps. First, we calibrated MOD17 with 500-m EVI and 1000-m FPAR using only the pixel co-located with flux towers and included these two models in our analyses. We refer to these two models as MOD17-$Tower_{500m}$ and MOD17-$Tower_{1km}$ (Table 1). Second, for all the remaining 9 models that used MODIS data in 7-by-7 (500-m) or 3-by-3 (1000-m) windows, we calculated biome-level performance by weighing each site with a site heterogeneity index, giving more weight to homogeneous sites (see Section 2.4).

2.4. Model calibration

MOD17 GPP data were downloaded from the Oak Ridge National Laboratory Distributed Active Archive Center, and we followed the approach described by Sims et al. (2008) to estimate GPP from the TG model. These two models were, therefore, not calibrated to tower-derived GPP data. The remaining 9 models were calibrated at daily time scale to tower-derived GPP using tower meteorology and MODIS data. Table 1 lists parameters in each model that were calibrated to tower-derived GPP. Biome specific parameters for each of the 11 models were estimated by minimizing a standard cost function. For example, in the case of the VPRM model, we calibrated two parameters, maximum light use efficiency and half-saturation point (Mahadevan et al., 2008), in each biome. We identified the realistic range of each of the two parameters, randomly sampled the parameters space 1000 times to prescribe the initial value of the parameter vector, and minimized a standard cost function using trust-region method for nonlinear optimization (MATLAB, 2012a, Optimization Toolbox, Mathworks.com). From the 1000 runs, we finally chose the parameter values that resulted in minimum discrepancy between predicted and tower-derived GPP. The same procedure was followed to calibrate parameters in each model.

We used the leave-one-site-out cross-validation method to evaluate the models. This method allows efficient and objective use of available data and enables independent calibration and

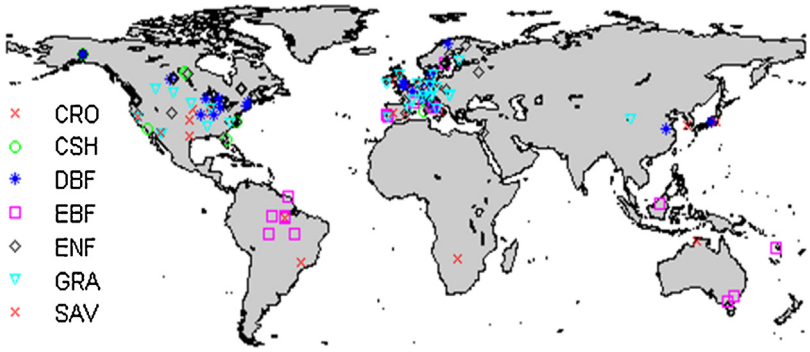

Fig. 1. Location of FLUXNET sites used in the study. CRO is croplands, CSH is shrublands, DBF is deciduous broadleaf forests, EBF is evergreen broadleaf forests, ENF is evergreen needleleaf forests, GRA is grasslands, and SAV is savannas.

evaluation of a model. Thus, in a biome with a total of n sites, we successively used $n-1$ sites to calibrate model parameters and employed these parameters to predict GPP at the left-out nth site.

Note that the official MOD17 is the only model that estimates GPP using coarse scale (0.5° or more) gridded data of PAR, air temperature, and VPD. Other models that used one or more meteorological variables were calibrated using fine resolution PAR, temperature, and VPD data from the "La Thuile" dataset. The MOD17 model provides GPP estimates as 8-day average values. Following Sims et al. (2008) we estimated the TG model at an 8-day time step. Apart from these two models, GPP from the rest of the models was estimated at a daily time step. We interpolated smoothly varying 8-day variables such as EVI and FPAR following the methods suggested by the original developer of each model. For example, to interpolate EVI to daily time scale in the VPRM model we followed the procedure from Mahadevan et al. (2008). Where this information was not provided, we first used a median filter to remove unrealistic spikes and then used locally weighted regression to smooth the data. To maintain consistency, we averaged daily GPP estimates to 8-day average before conducting all the analyses described in Section 2.5.1.

2.5. Statistical analyses

2.5.1. Inter-comparison of models

2.5.1.1. Intra-annual variations in daily GPP. To examine how well each of the 11 models captured intra-annual variation for each site-year, we assessed the agreement between daily modeled and tower-derived GPP. Model performance should be judged against the uncertainty in reference data (Keenan et al., 2012). Uncertainty in instantaneous GPP can be of the same order as the measurements. However, when instantaneous GPP is summed over a coarser temporal resolution the random noise in instantaneous measurements cancels and uncertainty decreases. At a daily time scale uncertainty (one standard deviation) in tower-derived GPP is typically assumed to be 15–20% of the measurements (Falge et al., 2002; Hagen et al., 2006). In this study, we conducted all the analyses at an 8-day time scale and assumed that uncertainty ($\pm\sigma$, where

Table 1
Remote sensing-based models of GPP examined in the study.

Model	Input data	Method to estimate GPP	Calibrated parameters	Reference
EVI-Linear	EVI	$GPP = a + b * EVI$	Two parameters at biome level	Schubert et al. (2012)
EVI-NonLinear	Same as above	$GPP = a * EVI^b$	Same as above	Model introduced in this study
PAR-EVI	EVI and PAR	$GPP = a + b * (PAR * EVI)$	Same as above	Wu et al. (2011)
PAR-FPAR	FPAR and PAR	$GPP = a + b * (PAR * FPAR)$	Same as above	Model introduced in this study
TG	EVI, day and night land surface temperature	$GPP = m * scaledLST * scaledEVI$ ('m' in the equation above is linearly related with annual night time LST)	Two parameters that relate 'm' to nighttime LST. Parameters values are different for deciduous and evergreen biomes	Sims et al. (2008)
VPRM	EVI, LSWI, PAR, air temperature	$GPP_t = \varepsilon * Tscale * Pscale * Wscale * EVI * \frac{PAR}{\left(1+\frac{PAR}{PAR_0}\right)}$	Two parameters ε and PAR_0 at biome level	Xiao et al. (2004); Mahadevan et al. (2008)
MOD17	FPAR, PAR, VPD and air temperature	$GPP = \varepsilon * Tscalar * VPDscalar * PAR * FPAR$	Five parameters including ε at biome level	Running et al. (2004)
MOD17-Tower	Same as MOD17, but high quality PAR, VPD, and air temperature from tower measurements	Same as MOD17	Same as above	Heinsch et al. (2006)
MOD17-Tower$_{500m}$	Same as MOD17-Tower, but EVI only from the pixel collocated with tower	Same as MOD17	Same as above	Variation of MOD17 introduced in the model-mix to examine the effect of land surface heterogeneity
MOD17-Tower$_{1000m}$	Same as MOD17-Tower, but FPAR only from the pixel collocated with tower	Same as MOD17	Same as above	Same as above
Neural Network	FPAR, PAR, VPD and air temperature	Feed forward neural network architecture	Weights of the network at biome level	Similar to Xiao et al. (2010)

σ denotes standard deviation) in mean 8-day tower-derived GPP is ±10% of the measurements. Then, using the approach described by Harmel and Smith (2007) we calculated the deviation between paired modeled and tower-derived GPP following Eq. (1) below.

$$e_i = \frac{\mathrm{cf}_i}{0.5}(O_i - P_i) \quad (1)$$

where e_i, cf_i, O_i, and P_i are the deviation between tower-derived and modeled GPP, a correction factor, tower-derived GPP, and modeled GPP, respectively, for the ith pair. For the two boundary cases when the modeled GPP value is more than $3.9^*\sigma$ away or is equal to the tower-derived GPP, the correction factor (cf_i) is 0.5 and zero respectively. When predicted GPP lies between tower-derived GPP and $\pm 3.9\sigma$, the correction factor is $\mathrm{cf}_i = p(|O_i| < Z < |P_i|)$ and is calculated as the area under a normal curve bounded by O_i and P_i, where O_i is the mean of the distribution (see Fig. 2 in Harmel and Smith (2007)).

Using Eq. (1) we calculated Willmott's index of agreement ($\mathrm{IoA}_{m,s}$; Willmott, 1981) and the root mean square error ($\mathrm{RMSE}_{m,s}$) between daily modeled and tower-derived GPP for each model m and site s (see Figure S1 and Eq. (A.1) and (A.2) in Appendix A).

Next, taking the $\mathrm{IoA}_{m,s}$, and $\mathrm{RMSE}_{m,s}$ as variables distributed over space, we calculated the biome specific across-site weighted mean and standard error for each of the two metrics for every model 'm' and biome b.

$$K^i_{m,b} = \sum_{j=1}^{n} \frac{w_j k^i_{m,j}}{w_j} \quad (2)$$

Here, i is one of the two performance metrics (IoA or RMSE), n is the number of sites in biome b, and w_j is the weight for site heterogeneity and k is the value of the metric for site j.

To account for site heterogeneity, we estimated a simple site homogeneity index (w^1_s) at each site s using 500-m MODIS land cover data (Friedl et al., 2010).

$$w^1_s = \left[\frac{\text{no. of pixels with land cover labels similar to tower pixel at site } s}{\text{total number of pixels in window}(=49)}\right]^2 \quad (3)$$

Note that by squaring the fraction we provide non-linearly higher weight to more homogeneous sites and lower weight to heterogeneous sites.

2.5.1.2. Inter-annual variations in daily GPP. To analyze inter-annual variability in daily GPP we first excluded all sites that had only one year of GPP data. For each of the remaining sites we then calculated anomalies in modeled and tower-derived daily GPP for every year. Thus, for a site with y years of data, we first calculated mean GPP by averaging over all the years then calculated daily anomalies by subtracting this average from the GPP for each site year (Figure S1). We then compared anomalies in daily GPP from models with corresponding anomalies from tower measurements for each site-year (Figure S1).

As described in Section 2.5.1.1, we first calculated $\mathrm{IoA}_{m,y}$ and $\mathrm{RMSE}_{m,y}$ between modeled and tower anomalies in GPP for each site year y for every model m. Because anomalies were obtained via a simple linear operation from daily GPP, we assumed that the relative magnitude of uncertainty in anomalies is the same (±10%) as in daily GPP. Next, we calculated the biome specific across-site weighted mean and standard error for each of the two metrics for every model m in a biome b.

We again weighted each site for site heterogeneity, but also included weights for our confidence in anomalies of tower-derived GPP. The weight for site heterogeneity was calculated as described in the previous section. For estimating our confidence in the accuracy of anomalies in tower-derived GPP, we calculated an index as follows.

$$w^1_{s,t,b} = \left[\frac{\text{mean anomaly in tower GPP at sites in year } t \text{ in biome } b}{\text{maximum mean anomaly in biome } b}\right]^2 \quad (4)$$

Thus, site-years with large anomalies were given more weight. Note that we square the fraction and thus assign higher weight to site-years with large anomalies.

2.5.2. Analyses of missing mechanisms

After comparing the eleven remote sensing models of GPP, we examined the potential of five different mechanisms articulated in Introduction for improving model performance in intra- and inter-annual analysis. To do this, we chose the best performing model from the analyses described above, Section 2.5.1, included each of the mechanisms postulated in Introduction in the model one-by-one, and analyzed the improvement in model performance. We specifically focused on DBF and grasslands where the seasonal reflectance signal is likely to be most tightly coupled to photosynthesis, and where remote sensing signals are expected to be more reliable. This allowed us to reduce the possibility of errors due to inaccuracies in remotely sensed phenology and uncoupling of leaf area and photosynthesis.

3. Results

We divide our results into two main sections. First, (Section 3.1) we present results from the inter-comparison of 11 models in intra- and inter-annual analysis. Next (Section 3.2), we share results about the effect of including the five mechanisms, described in the Introduction, on model performance in intra- and inter-annual analysis.

3.1. Agreement between predicted and tower-derived GPP

3.1.1. Intra-annual variation

Fig. 2 shows the values of IoA and RMSE for the models in the 7 biomes (see Figure S2 for pooled data from all the biomes). Collectively, the models had the weakest agreement with tower-derived GPP in EBF (IoA ≈ 0.4) and the strongest agreement in DBF (IoA ≈ 0.8). In terms of RMSE, the models had the lowest value in CSH and the highest in CRO (Fig. 2). Part of the difference between CSH and other biomes in terms of RMSE was because of the low mean GPP in CSH. However, for all most all the models, the RMSE in CRO was nearly twice that of RMSE in biomes such as DBF and GRA where the mean magnitude of daily GPP is similar to that in CRO.

In EBF, mean IoA varied from 0.35 (EVI-Linear) to 0.42 (neural network). The MOD17 model had the highest RMSE, but overall the difference amongst models was small or insignificant in terms of both IoA and RMSE. In CRO too, there was little inter-model difference in both the metrics (IoA ≈ 0.6; RMSE ≈ 3 $\mathrm{gC\,m^{-2}\,day^{-1}}$), although the MOD17 model displayed anomalously high RMSE here too. Similarly, in SAV all the models had nearly the same mean IoA (≈0.6) and RMSE (≈1.7 $\mathrm{gC^{-2}\,day^{-1}}$). Thus, in CRO and SAV, collectively the models performed better than in EBF, but the difference amongst the models was insignificant.

In the remaining four biomes, the difference amongst models was noticeable. In CSH, the neural network model had the highest IoA (≈0.72) and the lowest RMSE (≈0.3 $\mathrm{gC\,m^{-2}\,day^{-1}}$). The mean performance of the neural network model across sites was tightly constrained and showed relatively small variation (see uncertainty bars around mean IoA and RMSE in Fig. 2). In GRA, ENF, and DBF the more complex models (e.g. TG, VPRM, neural network, and MOD17-Tower) performed better than the simple models. These models showed strong agreement with tower-derived GPP with high IoA (≈0.8) and relatively lower RMSE, and showed little differences amongst themselves (also see Figures S2 and S3).

There were insignificant differences between models calibrated with data from single pixel collocated with towers ($\mathrm{MOD17\text{-}Tower_{500m}}$ and $\mathrm{MOD17\text{-}Tower_{1km}}$) and the models (e.g.

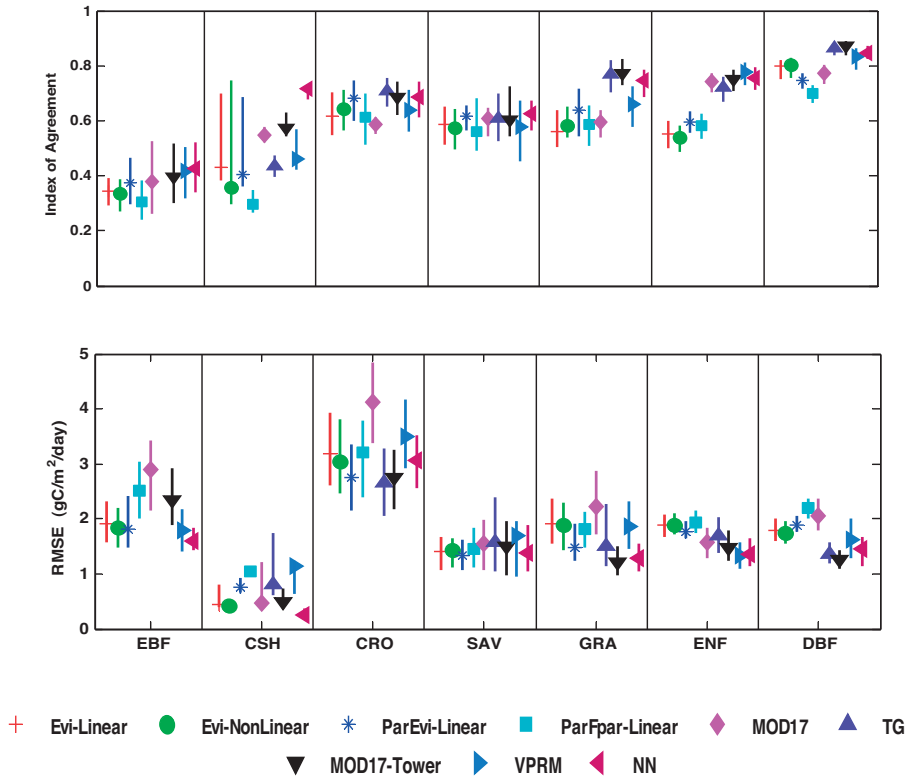

Fig. 2. Index of Agreement and root mean square error (RMSE) between daily GPP estimated from the nine models and derived from FLUXNET measurements. CSH, CRO, DBF, EBF, ENF, GRA, and SAV are shrublands, croplands, deciduous broadleaf, evergreen broadleaf, evergreen needle leaf forest, grasslands, and savanna, respectively.

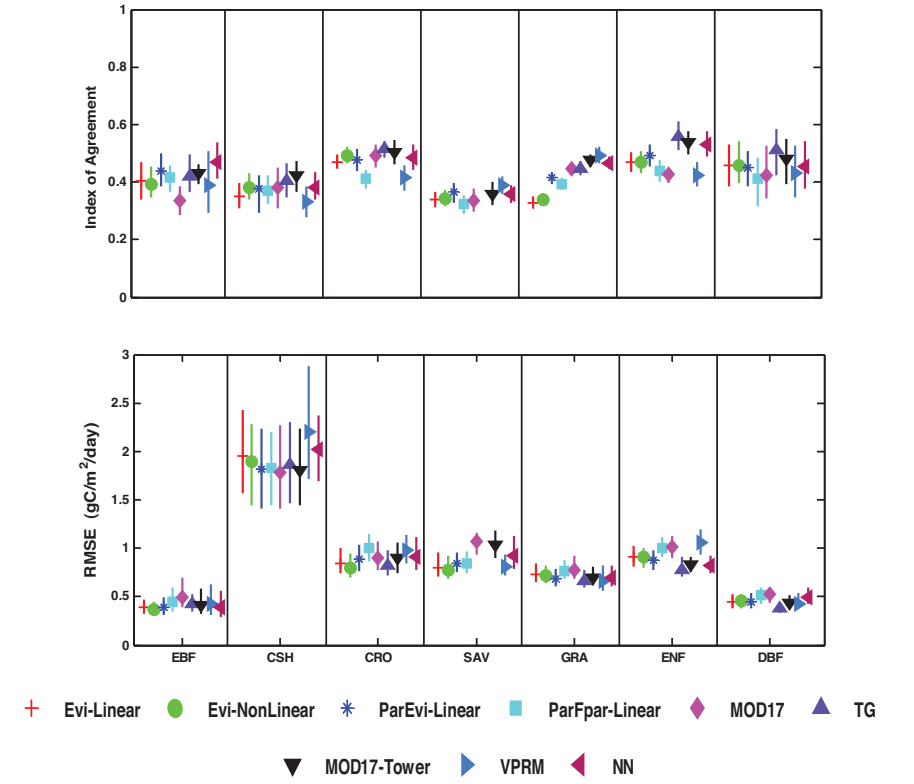

Fig. 3. Index of Agreement and root mean square error (RMSE) between anomalies in daily GPP estimated from the nine models and derived from FLUXNET measurements. CSH, CRO, DBF, EBF, ENF, GRA, and SAV are shrublands, croplands, deciduous broadleaf, evergreen broadleaf, evergreen needleleaf forest, grasslands, and savanna, respectively.

MOD17-Tower and VPRM) calibrated with data from 3-by-3 or 7-by-7 window centered on towers (results not shown).

3.1.2. Inter-annual variation

In contrast to their strong performance in capturing intra-annual variations in daily GPP, the models showed little skill (IoA < 0.5) in tracking inter-annual anomalies in daily GPP in all biomes (Fig. 3; Figure S3). There were a few site years in different biomes (e.g. in DBF) where model anomalies correlated well (IoA > 0.65) with tower anomalies, but overall, mean IoA values at the biome level were low. RMSE's in all biomes, (0.4–2.5 gC m^{-2} day^{-1}) were high and were comparable in magnitude with those observed in the intra-annual analysis, despite the fact that inter-annual variations in GPP are an order of magnitude smaller than intra-annual variations. We also found an insignificant difference in performance between the simplest model (EVI-Linear) and the more complex models except in GRA where more complex models had higher IoA.

There were insignificant differences between the two models calibrated using data from single pixels and the ones that used remote sensing data averaged over 3-by-3 or 7-by-7 window centered on tower locations (results now shown).

3.2. Examination of different hypotheses

The results above (in Section 3.1) showed that the TG, VPRM, MOD17-Tower and neural network model performed better than the other models. Of these four models, the structure of the VPRM, MOD17-Tower, and neural network can be easily modified to examine the additional predictive power of the five mechanisms identified in Introduction. Although, the value of the IoA for the VPRM model was a bit less than the other two models in GRA, based on goodness of fit and parsimony (Johnson and Omland, 2004) and the overall performance (Figures S2 and S3) we chose the VPRM model (with only 2 tunable biome specific parameters relative to 5 in the MOD17-Tower and an order of magnitude more in the neural network model; Table 1) to examine each of the hypotheses postulated in the Introduction for both intra- and inter-annual analysis. Our approach was to introduce each of the five modifications in the model one by one and assess if it resulted in better agreement with tower-derived GPP relative to the original VPRM model used in Section 3.1 (referred to as VPRM-daily). Since the number of observations used to assess model performance across the five different modifications was different, we used R^2 in this part of the analyses to assess improvement in model performance.

3.2.1. Model Calibration at 30-minute time scale

Model calibration at 30-minute time steps did not lead to better agreement between daily predicted and tower-derived GPP along intra-annual variation in DBF or GRA (Fig. 4). In DBF, R^2 increased by more than 0.05 for only 2 out of 79 site-years and for 11 out of 85 site-years in GRA. In fact, 30-minute calibration resulted in a decrease of more than 0.05 points in R^2 for 37 site-years out of 79, and in 34 out of 85 site-years, in DBF and GRA, respectively.

At inter-annual time scales there was a noticeable improvement in the correlation between model and tower anomalies when the model was calibrated at 30-minute time steps. The agreement between predicted and tower anomalies improved by more than 0.05 points for 24 (35%) and 20 site-years (25%) in DBF in GRA, respectively (Fig. 4, bottom row). However, R^2 also decreased by more than 0.05 for 11 (15%) and 10 (12%) site-years.

3.2.2. Separation of direct and diffuse radiation

We had data for the direct and diffuse component for 11 and 26 site-years in DBF and GRA, respectively. We re-calibrated the VPRM model at 30-minute resolution accounting for direct and diffuse PAR separately as below.

$$GPP_t = tscale_t * pscale_t * wscale_t * evi_t * \left(\varepsilon_{direct} \frac{1}{1 + \frac{directPAR_t}{h_{direct}}} + \varepsilon_{diffuse} \frac{1}{1 + \frac{diffusePAR_t}{h_{diffuse}}} \right) \quad (5)$$

Here, ε_{direct}, and h_{direct} are the maximum light use efficiency and half saturation point, respectively, for direct PAR, and $\varepsilon_{diffuse}$, and $h_{diffuse}$ are the corresponding parameters for diffuse PAR (see Figure 1 in Mahadevan et al. (2008) for the meaning of the remaining terms). Thus, here, there were four tunable parameters. As described earlier, we first identified the realistic range of each of the four parameters, randomly sampled the parameter space, and used the trust region algorithm to find the optimum value of the parameter vector. We also used alternative optimization methods and confirmed that there was no significant difference in final parameters selected by different methods.

Accounting for direct and diffuse PAR did not have an overall positive impact in both intra- and inter-annual analysis in DBF (Fig. 5). In GRA the agreement between modeled and tower-derived data showed both improvement and deterioration. In intra-annual analysis, correlation improved for nearly 40% of the sites but also decreased for nearly an equal number of sites. In the inter-annual analysis agreement improved for 18 site-years, but also decreased for 8 site-years.

3.2.3. Lagged effects and model calibration at 15-day time scale

To capture lagged effects within the model framework, we calibrated the model at a 15-day timescale (~2 weeks) and found that calibration at this coarser time scale resulted in improved performance at 17 out of 79 (21%) in DBF and in 33 out of 85 (39%) site-years in GRA (Fig. 6). At the remaining sites, R^2 values were the same for both daily and 15-day calibration. Thus, correlation either improved or remained same in both DBF and GRA. In inter-annual analysis correlation between tower and model anomalies increased for 22 out of 70 (31%) in DBF and 28 out of 81 (34%) site-years in GRA (Fig. 6). However, R^2 values also decreased for 14 site-years in DBF and 5 in GRA in inter-annual analysis.

3.2.4. Better representation of soil moisture dynamics

We used tower evaporative fraction (EF) instead of LSWI to capture soil moisture controls on stomatal functioning (Yuan et al., 2007). We calculated EF as LE/(LE + H + G), where LE, H, and G are the latent heat, sensible heat, and ground heat flux, respectively (Brutsaert and Chen, 1996). Note that tower data includes the independent measurement of surface net radiation, but because of the energy balance closure issue (e.g., Barr et al., 2006) we chose to estimate total surface net radiation as a sum of the three fluxes. To minimize the effect of seasonal variability in leaf area, we normalized evaporative fraction with EVI, and used the resulting fraction to calculate the moisture control scalar (w_{scale} in Eq. (5)) in the VPRM model. The moisture control scalar ranges between 0 and 1 and was calculated as (1 + nEF)/(1 + nEF_{max}), where nEF is the normalized EF, and nEF_{max} is the seasonal maximum of normalized EF for each site year.

Inclusion of EF had mixed effects on model performance (Fig. 7). In intra-annual analysis correlation between daily model and tower-derived GPP increased for 13 (16%), remained the same for 57 (72%), and decreased for 9 (12%) site-years in DBF. In GRA, the correlation increased for 39 (47%), remained same in 28 (34%), and decreased in 15 (19%) site-years. Thus, the positive effect of including EF was stronger in GRA than DBF.

In inter-annual analysis correlation between modeled and tower anomalies increased for 16 (22%), remained same for 27 (39%), and decreased for 27 (39%) site years in DBF (Fig. 6). In GRA R^2

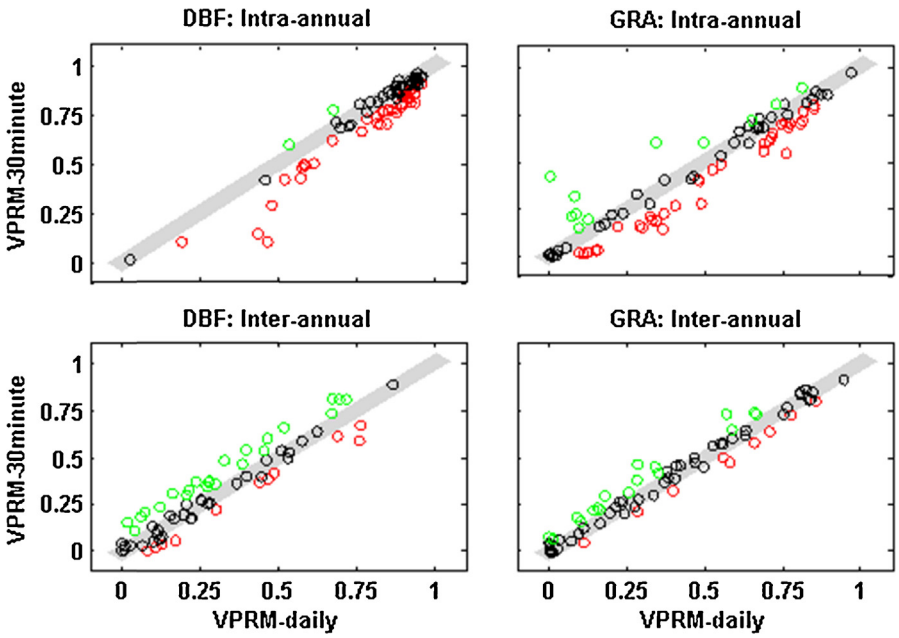

Fig. 4. R^2 between daily modeled and tower-derived GPP for the VPRM model calibrated at daily (VPRM-daily) and 30-minute (VPRM-30minute) time scale in intra- and inter-annual analysis. DBF and GRA are deciduous broadleaf forest and grasslands, respectively. Shaded bar shows 1:1 line with ±0.05 units. Green circles show site-years with better R^2, black circles depict site-years with no improvement, and red circles mark site-years that registered a decline in performance relative to the original VPRM model calibrated at daily time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increased for 26 (33%), remained same for 19 (24%), and decreased for 33 (43%) site years.

3.2.5. Spatial variation in parameters

To assess whether geographic variation in model parameters influences model results, we allowed model parameters to vary from site-to-site (but not from year-to-year for the same site), and calibrated the VPRM model at every site. For a site with n site-years of data, we successively used $n-1$ years of data for calibration and employed the calibrated parameters to predict GPP at the left-out nth site year. Results from this analysis showed that allowing model parameters to vary over space (site-to-site) did not improve R^2 in GRA or DBF at either intra- or interannual time scales (Fig. 8).

4. Discussion

4.1. Seasonal variation in daily GPP

With the exception of EBF, mean seasonal GPP predicted by the simplest model (EVI-Linear) showed good agreement with daily

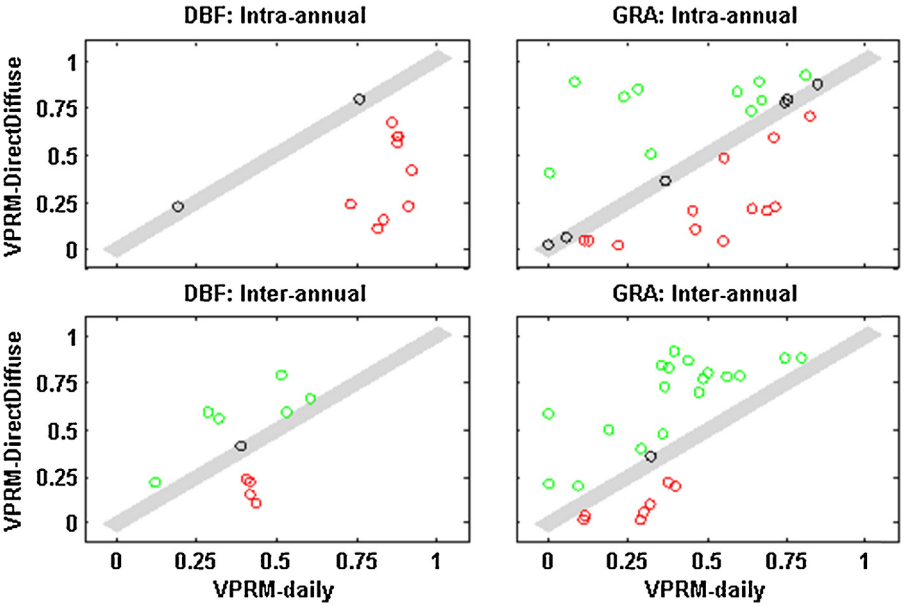

Fig. 5. R^2 between daily modeled and tower-derived GPP for the VPRM model calibrated at daily (VPRM-daily) and 30-minute time scale accounting for direct and diffuse radiation separately (VPRM-DirectDiffuse) in intra- and inter-annual analysis. DBF and GRA are deciduous broadleaf forest and grasslands, respectively. Shaded bar shows 1:1 line with ±0.05 units. Green circles show site-years with better R^2, black circles depict site-years with no improvement, and red circles mark site-years that registered a decline in performance relative to the original VPRM model calibrated at daily time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

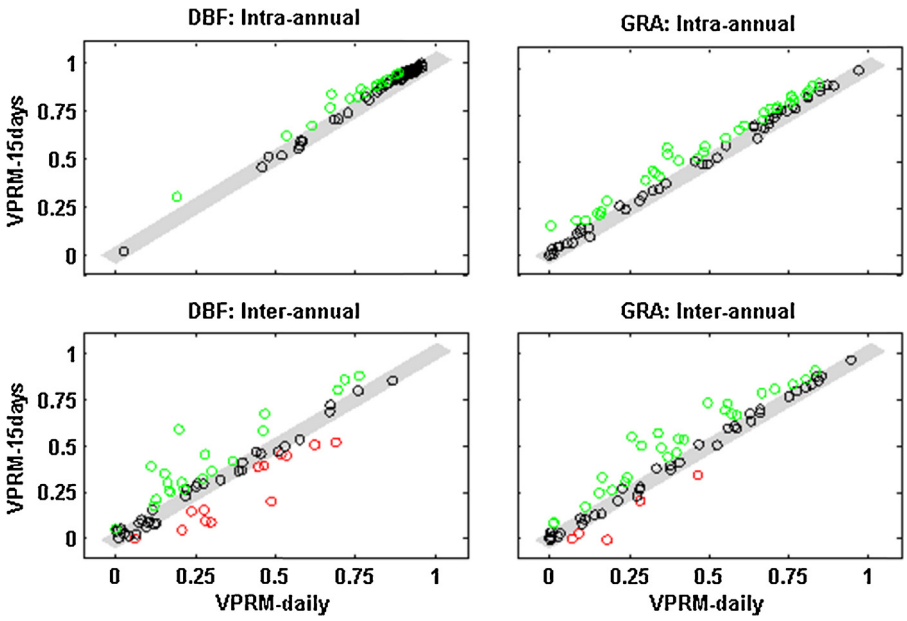

Fig. 6. R^2 between daily modeled and tower-derived GPP for the VPRM model calibrated at daily (VPRM-daily) and 15-day (VPRM-15days) time scale in intra- and inter-annual analysis. DBF and GRA are deciduous broadleaf forest and grasslands, respectively. Shaded bar shows 1:1 line with ±0.05 units. Green circles show site-years with better R^2, black circles depict site-years with no improvement, and red circles mark site-years that registered a decline in performance relative to the original VPRM model calibrated at daily time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tower-derived GPP. Because the EVI-Linear model only includes green leaf phenology (via 8-day EVI), these results confirm that leaf area is the main driver of seasonal variation in GPP across deciduous biomes (Xiao et al., 2004; Xu and Baldocchi, 2004; Wohlfahrt et al., 2008; Table 1).

The performance of the two absorbed PAR-based models (PAR*FPAR and PAR*EVI) were also indistinguishable from the simplest model in most biomes. Seasonal and spatial variations in leaf area in natural ecosystems can be limited by several factors including leaf-area, light, water, and nutrients. However, the fact that EVI provided as much information about the variability in GPP as absorbed PAR suggests that seasonal variation in leaf area is optimized to capture variability in PAR. In EBF however, EVI and PAR did not show large seasonal variations, were not synchronous, and did not drive seasonal variation in GPP. Therefore, the assumption of a linear relationship between EVI and GPP is significantly weaker (Asner and Alencar, 2010) in EBF and the performance of the EVI-Linear model is anomalously poor.

The MOD17 model did not perform better than the simplest model, likely because of errors and uncertainty in the coarse resolution meteorological forcing data used by the algorithm (Heinsch et al., 2006; also see discussion below).

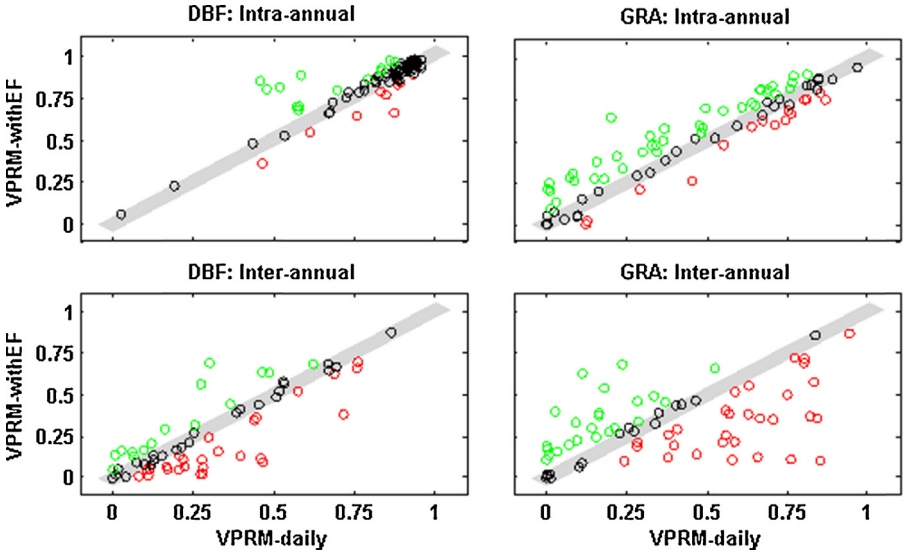

Fig. 7. R^2 between daily modeled and tower-derived GPP for the VPRM model calibrated at daily time scale with MODIS derived land surface water index (LSWI) and evaporative fraction (VPRM-EF) to capture the effect of soil moisture availability in intra- and inter-annual analysis. DBF and GRA are deciduous broadleaf forest and grasslands, respectively. Shaded bar shows 1:1 line with ±0.05 units. Green circles show site-years with better R^2, black circles depict site-years with no improvement, and red circles mark site-years that registered a decline in performance relative to the original VPRM model calibrated at daily time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

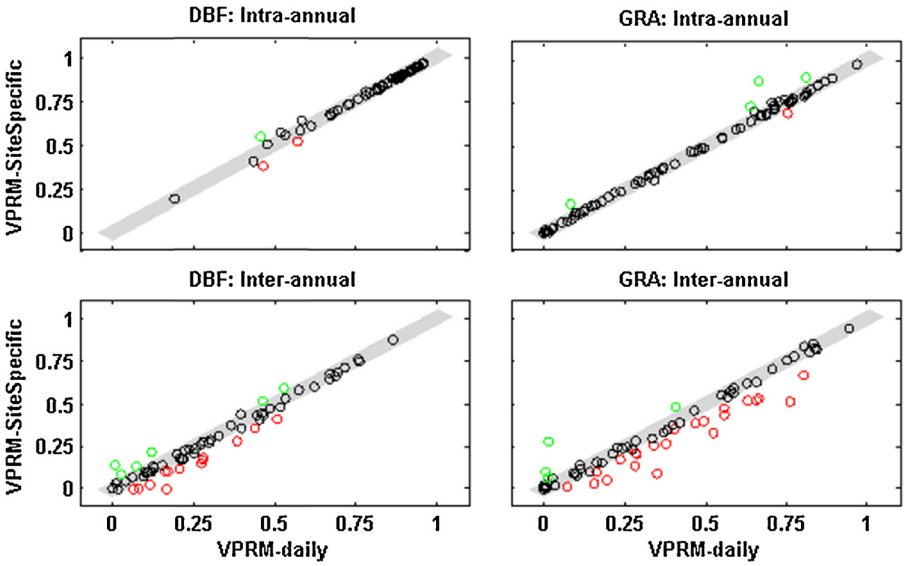

Fig. 8. R^2 between daily modeled and tower-derived GPP for the VPRM model calibrated at daily time scale with biome specific (VPRM-daily) and site-specific parameters (VPRM-SiteSpecific) in intra- and inter-annual analysis. DBF and GRA are deciduous broadleaf forest and grasslands, respectively. Shaded bar shows 1:1 line with ±0.05 units. Green circles show site-years with better R^2, black circles depict site-years with no improvement, and red circles mark site-years that registered a decline in performance relative to the original VPRM model calibrated at daily time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In EBF and CRO, the more complex models did not perform better than the null model. In deep-rooted EBF, VPD can be a poor indicator of moisture availability and does not provide accurate information about stomatal behavior. On the other hand, in CRO, management practices, which are not included in the models here, play a more important role than VPD or temperature fluctuations in influencing GPP.

The TG, MOD17-Tower, VPRM, and neural network models, showed significantly better performance than the EVI-Linear model in the remaining four biomes. The TG model, despite its relatively simple formulation, displayed comparable performance as the other three models (The MOD17-Tower, VPRM, and neural network model). Seasonal GPP predicted by the four models agreed better with tower-derived GPP than the "null" model in four of the seven biomes in at least one of the two performance criteria. The TG model's additional power relative to the "null" model was derived from including daytime land surface temperature in the model. Land surface temperature not only captures temperature driven dynamics in GPP at higher latitudes, but may also capture the effect of VPD on GPP because surface temperatures are partly controlled by moisture availability (Sims et al., 2008). Some studies have suggested that meteorological information is redundant for estimating GPP at the daily time scale (Sims et al., 2008; Jung et al., 2008). However, our results do not support this. Specifically, the fact that the more complex models perform better than the models that did not include meteorological forcing suggests that temperature and VPD provide useful information related to seasonal variation in GPP. Differences between the performance of the neural network and the other three models were small. Neural network models have been previously used in modeling and upscaling GPP and evapotranspiration (Xiao et al., 2010). However, our results suggest that the empirical, flexible, and nonlinear structure of the neural network model did not provide significant advantage over the relatively simple models. The ability of the neural network model to predict GPP was not significantly different from the simple but physiologically based models such as the VPRM and MOD17-Tower.

Between biomes, the proportion of variance in tower-derived GPP explained by modeled GPP was not strongly dependent on the magnitude of variance in tower-derived GPP. For example, in ENF the standard deviation in daily tower-derived GPP was lower than in CRO, but the IoA between modeled and tower-derived GPP was higher in ENF than in CRO for most models.

4.2. Inter-annual variations in daily GPP

Relative to their ability to explain intra-annual variation in GPP, all the models were significantly less successful in capturing inter-annual variability and more complex models did not perform better than the simpler models. For nearly 20–25% of the total site-years, the IoA value for the MOD17-Tower and VPRM models was greater than 0.5 in GRA. However, in most of these site-years the anomalies in GPP correlated equally strongly with the anomalies in the simplest model, suggesting that inter-annual variations in GPP were mainly caused by changes in leaf-area. The ability of the models to capture inter-annual variations was inconsistent across years in GRA. Thus, the models successfully captured anomalies in one year but failed to do so in other years at the same site.

4.3. Improving the performance of "big-leaf" models

We noticed a significant difference in the potential of the five mechanisms in improving model performance in GRA and DBF. Allowing model parameters to vary from site to site did not improve agreement between modeled and tower-derived GPP in either biome in both intra- and inter-annual analyses. We also compared Akaike Information Criterion (AIC) and found that AIC did not exhibit significant difference in model performance between the site- and biome-level calibration. Unlike some studies (Horn and Schulz, 2011; Madani et al., 2014) our results did not support the hypothesis that site-to-site variation in maximum light use efficiency is an important source of intra- or inter-annual variation in daily GPP. Using data from 168 FLUXNET sites Yuan et al.

(2014b) showed that biome specific parameters are not required in LUE models. Although we used biome-specific parameters as suggested by the developers of each model, our results together with the results from Yuan et al. (2014b) might suggest that within the errors and uncertainties in input data, spatial variability in parameters does not significantly affect the accuracy of predicted GPP.

Results from our analysis of the remaining four hypotheses were more nuanced. Calibration at the 30-minute time scale did not lead to overall improvement in intra-annual analysis. However, it did result in better agreement for a number of site-years in both GRA and DBF in inter-annual analysis. Similarly, accounting for direct and diffuse radiation showed improvement at a number of sites in inter-annual analyses, but not in intra-annual analyses. Empirical studies have shown that maximum light use efficiencies for direct and diffuse PAR are different (Gu et al., 2002) and formalizing this reality in model structure is considered important to capture variation in GPP (McCallum et al., 2009). However, it appears that this difference does not have a significant impact on model performance in the context of "big-leaf" models in intra-annual analyses. Seasonal (intra-annual) variation in daily GPP is strongly controlled by variation in daily PAR, EVI, and temperature and sub-daily variation does not provide extra information. However, once the mean seasonal cycle is removed (i.e. for anomalies) the effect of sub-daily fluctuations in meteorological variables on photosynthesis becomes more important. Thirty-minute measurements provide important information, but they also have significantly higher levels of noise relative to daily values. Thus, there is a tradeoff between additional information and larger proportions of noise. Sites where sub-daily fluctuations in meteorological forcing are relatively large and frequent are likely to have higher signal to noise ratio and greater non-linearity in photosynthetic response to environmental conditions. Hence, higher-frequency inputs are likely to be more important in locations where meteorological forcing tends to be more variable at short time scales.

A number of studies have reported poor performance of models in capturing inter-annual variations in GPP (Jung et al., 2008; Keenan et al., 2012; Verma et al., 2014). Similarly, poor correlation between environmental drivers and interannual variation in GPP has been reported in site-level studies (Wohlfahrt et al., 2008). As discussed in Section 4.2 above, we also noted the poor performance of the models in capturing inter-annual variation. However, we noticed a perceptible improvement in model performance in inter-annual analysis for the model calibrated at the 15-day time scale. This result points to both the importance of lagged effect and reduction of noise in input data at longer time scales.

Including EF resulted in significant improvement in predictions for a number of site-years, but it also decreased the quality of model predictions for many other site-years. Thus, overall at biome level the improvement in performance was negligible. AIC also showed that at biome-level the improvement was small or negligible (results not shown). Substituting LSWI with EF improved model performance for site-years where modeled anomalies correlated poorly ($R^2 < 0.4$) with tower anomalies for the VPRM-daily model in GRA. However, the site-years where VPRM-daily performed well ($R^2 > 0.4$), agreement between measured and modeled GPP from the VPRM-with-EF model decreased. Normalized EF is sensitive to errors and noise in both EF and EVI. Hence, even though EF has the potential to provide useful information regarding moisture control on GPP, our results suggest that noise in EF data limits its utility to sites where moisture constraints are significant. Currently EF, as a standard product, is not available from remote sensing data. However, it can be derived at the satellite overpass time following methods suggested in different studies. Instantaneous net-radiation from MODIS can be estimated following Bisht and Bras (2011) and Bisht et al. (2005). Similarly, ET at the satellite overpass time can be estimated from remotely sensed data using simple formulations based on Priestley-Taylor (e.g. Fisher et al., 2008) or Penman-Monteith (such as Mu et al., 2011). Instantaneous EF at the satellite overpass time can then be calculated from the instantaneous ET and net-radiation. Because EF remains nearly constant over the course of a day, it can be safely assumed that daily EF is same as the instantaneous EF (Brutsaert and Chen, 1996).

Among the five additional mechanisms that we considered here, it appears that accurately modeling the soil moisture availability is likely to improve model performance at both intra- and inter-annual time scales. However, our analyses also suggest that the level of improvement in model performance was moderate to small especially at biome level. In intra-annual analyses few sites showed large improvement (>0.10 in R^2). In inter-annual analysis some site-years did show large improvement (>0.10 in R^2, e.g. Figs. 5 and 7), but this was also accompanied by decrease in agreement at other sites. We conducted our analyses using high-quality tower data. Global models often use coarse resolution data with larger errors and uncertainty. Considering the uncertainty and error in tower-derived GPP and estimated parameters, and the moderate or no improvement at a large number of sites the likelihood of getting significant improvement in global models with any of the five modifications discussed here is not likely to be very high.

5. Conclusion

We compared seasonal GPP predicted from 11 remote sensing-based models with tower-derived GPP to assess the biome-specific relative performance of each model. The selected models are based on different approaches, formalize a variety of hypotheses about processes that control seasonal GPP, and cover a spectrum of model complexity. We then analyzed the potential of five different mechanisms that have been identified in the literature, but which are not represented in any of the models, for improving model predictions of GPP at intra- and inter-annual time scales.

Remote-sensing based models were more successful in capturing intra- than inter-annual variation in seasonal GPP in every biome. More complex models performed better than the null model in capturing intra-annual variations in daily GPP, but the difference between the null model and other models was insignificant at inter-annual time scales.

Inclusion of some mechanisms such as better representation of soil moisture improved agreement between modeled and tower-derived data. However, we also noticed negative or no improvement at a large number of sites. We cannot rule out the effects of noise and uncertainty in both the data and model parameters, which can distort correlations between modeled and tower-derived GPP. However, weak correlations in many site-years, especially at inter-annual time scale, also points to the possibility that "big-leaf" representations have incomplete formalization of the biotic and abiotic factors that control GPP and do not appropriately capture scaled-up leaf-level processes or biotic variability. Inter-annual variations arise because of complex interactions between ecosystem and environmental changes, and also have significantly smaller magnitudes than intra-annual variations. Better understanding regarding how leaf level processes scale up to ecosystem level and the relative importance of biotic and abiotic factors on variation in GPP in different ecosystems are required to improve currently available models.

Acknowledgements

This research was partially supported by NASA grant number NNX11AE75G, the National Science Foundation Macrosystem

Biology program (award EF-1065029), and AmeriFlux [the Office of Science (BER), US Department of Energy (DOE; DE-FG02-04ER63917 and DE-FG02-04ER63911)]. MV and MAF gratefully acknowledge the efforts of the FLUXNET community to compile and make available the La Thuile data set. 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)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, 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, 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 and the University of Virginia.

Appendix A.

We calculated Willmott's Index of Agreement (IoA; Eq. (A.1)) and root mean square error (RMSE, Eq. (A.2)) between predicted and tower-derived GPP for each site-year 's' as follows:

$$\mathrm{IoA}_s = \frac{\sum_{i=1}^{n} \frac{cf_i}{0.5}|O_i - P_i|}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)} \quad (A.1)$$

$$\mathrm{RMSE}_s = \sqrt{\frac{\sum_{i=1}^{n}\left[\frac{cf_i}{0.5}(O_i - P_i)\right]^2}{n}} \quad (A.2)$$

Here, P_i, O_i, and $\bar{O}$ are predicted GPP, tower-derived GPP and mean of tower-derived GPP.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agrformet.2015.09.005.

References

Alton, P.B., North, P.R., Los, S.O., 2007. The impact of diffuse sunlight on canopy light-use efficiency, gross photosynthetic product and net ecosystem exchange in three forest biomes. Glob. Change Biol. 13, 776–787, http://dx.doi.org/10.1111/j.1365-2486.2007.01316.x.

Asner, G.P., Alencar, A., 2010. Drought impacts on the Amazon forest: the remote sensing perspective. New Phytol. 187, 569–578, http://dx.doi.org/10.1111/j.1469-8137.2010.03310.x.

Barr, A., Morgenstern, K., Black, T., McCaughey, J., Nesic, Z., 2006. Surface energy balance closure by the eddy-covariance method above three boreal forest stands and implications for the measurement of the CO_2 flux. Agric. For. Meteorol. 140, 322–337.

Beer, C., Reichstein, M., Tomelleri, E., et al., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838, http://dx.doi.org/10.1126/science.1184984.

Bisht, G., Bras, R., 2011. Estimation of net radiation from the moderate resolution imaging spectroradiometer over the continental united states. IEEE T. Geosci. Remote 49, 2448–2462.

Bisht, G., Venturini, V., Islam, S., Jiang, L., 2005. Estimation of the net radiation using MODIS (Moderate Resolution Imaging Spectroradiometer) data for clear sky days. Remote Sens. Environ. 97, 52–67.

Bonan, G.B., 1995. Land-atmosphere interactions for climate system models: coupling biophysical, biogeochemical, and ecosystem dynamical processes. Remote Sens. Environ. 51, 57–73, http://dx.doi.org/10.1016/0034-4257(94)00065-U.

Brutsaert, W., Chen, D., 1996. Diurnal variation of surface fluxes during thorough drying (or severe drought) of natural prairie. Water Resour. Res. 32, 2013–2019.

Chasmer, L., Barr, A., Hopkinson, C., McCaughey, H., Treitz, P., Black, A., Shashkov, A., 2009. Scaling and assessment of GPP from MODIS using a combination of airborne lidar and eddy covariance measurements over jack pine forests. Remote Sens. Environ. 113, 82–93, http://dx.doi.org/10.1016/j.rse.2008.08.009.

Cook, B.D., Bolstad, P.V., Martin, J.G., et al., 2008. Using light-use and production efficiency models to predict photosynthesis and net carbon exchange during forest canopy disturbance. Ecosystems 11, 26–44, http://dx.doi.org/10.1007/s10021-007-9105-0.

Coops, N., Black, T., Jassal, R., Trofymow, J., Morgenstern, K., 2007. Comparison of MODIS, eddy covariance determined and physiologically modelled gross primary production (GPP) in a Douglas-fir forest stand. Remote Sens. Environ. 107, 385–401, http://dx.doi.org/10.1016/j.rse.2006.09.010.

Dai, Y., Dickinson, R.E., Wang, Y.P., 2004. A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. J. Clim. 17, 2281–2299, http://dx.doi.org/10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2.

Dietze, M.C., Vargas, R., Richardson, A.D., et al., 2011. Characterizing the performance of ecosystem models across time scales: a spectral analysis of the North American Carbon Program site-level synthesis. J. Geophys. Res. 116, G04029, http://dx.doi.org/10.1029/2011JG001661.

Falge, E., Baldocchi, D., Tenhunen, J., et al., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. For. Meteorol. 113, 53–74, http://dx.doi.org/10.1016/S0168-1923(02)00102-8.

Fisher, J.B., Tu, K.P., Baldocchi, D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X.M., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182, http://dx.doi.org/10.1016/j.rse.2009.08.016.

Ganguly, S., Friedl, M.A., Tan, B., Zhang, X.Y., Verma, M., 2010. Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. Remote Sens. Environ. 114, 1805–1816, http://dx.doi.org/10.1016/j.rse.2010.04.005.

Gitelson, A.A., Viña, A., Verma, S.B., et al., 2006. Relationship between gross primary production and chlorophyll content in crops: implications for the synoptic monitoring of vegetation productivity. J. Geophys. Res. 111, D08S11, http://dx.doi.org/10.1029/2005JD006017.

Groenendijk, M., Dolman, A.J., Molen, M.K., et al., 2011. Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. Agric. For. Meteorol. 151, 1–17, http://dx.doi.org/10.1016/j.agrformet.2010.08.013.

Gu, L., Baldocchi, D., Verma, S.B., Black, T., Vesala, T., Falge, E.M., Dowty, P.R., 2002. Advantages of diffuse radiation for terrestrial ecosystem productivity. J. Geophys. Res. Atmos. 107, http://dx.doi.org/10.1029/2001JD001242, ACL 2-1–ACL 2-23.

Hagen, S.C., Braswell, B.H., Linder, E., Frolking, S., Richardson, A.D., Hollinger, D.Y., 2006. Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at Howland Forest, Maine. J. Geophys. Res. Atmos. 111, 1–12.

Harmel, R.D., Smith, P.K., 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. J. Hydrol. 337, 326–336.

Hashimoto, H., Wang, W., Milesi, C., et al., 2012. Exploring simple algorithms for estimating gross primary production in forested areas from satellite data. Remote Sens. 4, 303–326, http://dx.doi.org/10.3390/rs4010303.

Heinsch, F.A., Zhao, M.S., Running, S.W., et al., 2006. Evaluation of remote sensing based terrestrial productivity from MODIS using regional tower eddy flux network observations. IEEE Trans. Geosci. Remote Sens. 44, 1908–1925, http://dx.doi.org/10.1109/TGRS.2005.853936.

Horn, J., Schulz, K., 2011. Spatial extrapolation of light use efficiency model parameters to predict gross primary production. J. Adv. Model. Earth Syst. 3, M12001, http://dx.doi.org/10.1029/2011MS000070.

Irvine, J., Law, B.E., Kurpius, M.R., Anthoni, P.M., Moore, D., Schwarz, P.A., 2004. Age-related changes in ecosystem structure and function and effects on water and carbon exchange in ponderosa pine. Tree Physiol. 24, 753–763, http://dx.doi.org/10.1093/treephys/24.7.753.

Jahan, N., Gan, T.Y., 2009. Modeling gross primary production of deciduous forest using remotely sensed radiation and ecosystem variables. J. Geophys. Res. 114, G04026, http://dx.doi.org/10.1029/2008JG000919.

Jenkins, J.P., Richardson, A.D., Braswell, B.H., Ollinger, S.V., Hollinger, D.Y., Smith, M.L., 2007. Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements. Agric. For. Meteorol. 143, 64–79, http://dx.doi.org/10.1016/j.agrformet.2006.11.008.

Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108, http://dx.doi.org/10.1016/j.tree.2003.10.013.

Jung, M., Verstraete, M., Gobron, N., Reichstein, M., Papale, D., Bondeau, A., Robustelli, M., Pinty, B., 2008. Diagnostic assessment of European gross primary production. Glob. Change Biol. 14, 2349–2364, http://dx.doi.org/10.1111/j.1365-2486.2008.01647.x.

Justice, C.O., Townshend, J.R.G., Vermote, E.F., et al., 2002. An overview of MODIS land data processing and product status. Remote Sens. Environ. 83, 3–15, http://dx.doi.org/10.1016/S0034-4257(02)00084-6.
Kanniah, K.D., Beringer, J., Hutley, L.B., Tapper, N.J., Zhu, X., 2009. Evaluation of collections 4 and 5 of the MODIS gross primary productivity product and algorithm improvement at a tropical savanna site in northern Australia. Remote Sens. Environ. 113, 1808–1822, http://dx.doi.org/10.1016/j.rse.2009.04.013.
Keeling, C., Whorf, T., Wahlen, M., Plicht, J., 1995. Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. Nature 375, 666–670, http://dx.doi.org/10.1038/375666a0.
Keenan, T.F., Baker, I., Barr, A., et al., 2012. Terrestrial biosphere model performance for interannual variability of land atmosphere CO_2 exchange. Glob. Change Biol. 18, 1971–1987, http://dx.doi.org/10.1111/j.1365-2486.2012.02678.x.
Law, B., Waring, R., 1994. Combining remote sensing and climatic data to estimate net primary production across Oregon. Ecol. Appl. 4, 717–728, http://dx.doi.org/10.2307/1942002.
Law, B.E., Goldstein, A.H., Anthoni, P.M., et al., 2001. Carbon dioxide and water vapor exchange by young and old ponderosa pine ecosystems during a dry summer. Tree Physiol. 21, 299–308, http://dx.doi.org/10.1093/treephys/21.5.299.
Leuning, R., Cleugh, H., Zegelin, S., Hughes, D., 2005. Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates. Agric. For. Meteorol. 129, 151–173, http://dx.doi.org/10.1016/j.agrformet.2004.12.004.
Lin, J.C., Pejam, M.R., Chan, E., Wofsy, S.C., Gottlieb, E.W., Margolis, H.A., Mccaughey, J.H., 2011. Attributing uncertainties in simulated biospheric carbon fluxes to different error sources. Glob. Biogeochem. Cycles 25, GB2018, http://dx.doi.org/10.1029/2010GB003884.
Madani, N., Kimball, J.S., Affleck, D.L.R., et al., 2014. Improving ecosystem productivity modeling through spatially explicit estimation of optimal light use efficiency. J. Geophys. Res. Biogeosci. 119, 1755–1769, http://dx.doi.org/10.1002/2014JG002709.
Mahadevan, P., Wofsy, S.C., Matross, D.M., et al., 2008. A satellite-based biosphere parameterization for net ecosystem CO(2) exchange: Vegetation Photosynthesis and Respiration Model (VPRM). Glob. Biogeochem. Cycles 22, GB2005, http://dx.doi.org/10.1029/2006GB002735.
Makela, A., Pulkkinen, M., Kolari, P., et al., 2008. Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Glob. Change Biol. 14, 92–108, http://dx.doi.org/10.1111/j.1365-2486.2007.01463.x.
McCallum, I., Wagner, W., Schmullius, C., Shvidenko, A., Obersteiner, M., Fritz, S., Nilsson, S., 2009. Satellite-based terrestrial production efficiency modeling. Carbon Balance Manage. 4, http://dx.doi.org/10.1186/1750-0680-4-8.
Medvigy, D., Wofsy, S.C., Munger, J.W., Moorcroft, P.R., 2010. Responses of terrestrial ecosystems and carbon budgets to current and future environmental variability. Proc. Nat. Acad. Sci. U. S. A. 107, 8275–8280, http://dx.doi.org/10.1073/pnas.0912032107.
Monteith, J., 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9, 747–766.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800.
Myneni, R.B., Hoffman, S., Knyazikhin, Y., et al., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83, 214–231.
Olofsson, P., Lagergren, F., Lindroth, A., Lindström, J., Klemedtsson, L., Eklundh, L., 2008. Towards operational remote sensing of forest carbon balance across Northern Europe. Biogeosciences 5, 817–832, http://dx.doi.org/10.5194/bg-5-817-2008.
Propastin, P., Ibrom, A., Knohl, A., Erasmi, S., 2012. Effects of canopy photosynthesis saturation on the estimation of gross primary productivity from MODIS data in a tropical forest. Remote Sens. Environ. 121, 252–260, http://dx.doi.org/10.1016/j.rse.2012.02.005.
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. Glob. Change Biol. 11, 1424–1439, http://dx.doi.org/10.1111/j.1365-2486.2005.001002.x.
Richardson, A.D., Black, T.A., Ciais, P., et al., 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. Philos. Trans. R. Soc. Lond. B 365, 3227–3246, http://dx.doi.org/10.1098/rstb.2010.0102.
Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M.S., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. BioScience 54, 547–560, http://dx.doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.
Sakamoto, T., Gitelson, A.A., Wardlow, B.D., Verma, S.B., Suyker, A.E., 2011. Estimating daily gross primary production of maize based only on MODIS WDRVI and shortwave radiation data. Remote Sens. Environ. 115, 3091–3101, http://dx.doi.org/10.1016/j.rse.2011.06.015.
Schaaf, C.B., Gao, F., Strahler, A.H., et al., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sens. Environ. 83, 135–148, http://dx.doi.org/10.1016/S0034-4257(02)00091-3.
Schimel, D., 2007. Carbon cycle conundrums. Proc. Nat. Acad. Sci. U. S. A. 104, 18353–18354.
Schubert, P., Lagergren, F., Aurela, M., et al., 2012. Modeling GPP in the Nordic forest landscape with MODIS time series data-Comparison with the MODIS GPP product. Remote Sens. Environ. 126, 136–147, http://dx.doi.org/10.1016/j.rse.2012.08.005.
Schwalm, C.R., Williams, C.A., Schaefer, K., et al., 2010. A model-data intercomparison of CO_2 exchange across North America: results from the North American Carbon Program site synthesis. J. Geophys. Res. 115, G00H05, http://dx.doi.org/10.1029/2009JG001229.
Sims, D.A., Rahman, A.F., Cordova, V.D., et al., 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote Sens. Environ. 112, 1633–1646, http://dx.doi.org/10.1016/j.rse.2007.08.004.
Sjöström, M., Ardö, J., Arneth, A., et al., 2011. Exploring the potential of MODIS EVI for modeling gross primary production across African ecosystems. Remote Sens. Environ. 115, 1081–1089, http://dx.doi.org/10.1016/j.rse.2010.12.013.
Sjöström, M., Zhao, M., Archibald, S., et al., 2013. Evaluation of MODIS gross primary productivity for Africa using eddy covariance data. Remote Sens. Environ. 131, 275–286, http://dx.doi.org/10.1016/j.rse.2012.12.023.
Solomon, S., Qin, D., Manning, M., et al., 2007. Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change.
Tan, B., Woodcock, C., Hu, J., et al., 2006. The impact of gridding artifacts on the local spatial properties of MODIS data: implications for validation, compositing, and band-to-band registration across resolutions. Remote Sens. Environ. 105, 98–114.
Thomas, C.K., Law, B.E., Irvine, J., Martin, J.G., Pettijohn, J.C., Davis, K.J., 2009. Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semiarid mature ponderosa pine forest in central Oregon. J. Geophys. Res. 114, G04006, http://dx.doi.org/10.1029/2009JG001010.
Turner, D.P., Ritts, W.D., Zhao, M.S., et al., 2006. Assessing interannual variation in MODIS-based estimates of gross primary production. IEEE Trans. Geosci. Remote Sens. 44, 1899–1907.
Ueyama, M., Harazono, Y., Ichii, K., 2010. Satellite-based modeling of the carbon fluxes in mature black spruce forests in Alaska: a synthesis of the eddy covariance data and satellite remote sensing data. Earth Interact. 14, 1–27, http://dx.doi.org/10.1175/2010EI319.1.
Verma, M., Friedl, M.A., Richardson, A.D., et al., 2014. Remote sensing of annual terrestrial gross primary productivity from MODIS: an assessment using the FLUXNET La Thuile data set. Biogeosciences 11, 2185–2200, http://dx.doi.org/10.5194/bgd-10-11627-2013.
Wan, Z.M., Zhang, Y.L., Zhang, Q.C., Li, Z.L., 2002. Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sens. Environ. 83, 163–180, http://dx.doi.org/10.1016/S0034-4257(02)00093-7.
Willmott, C.J., 1981. On the validation of models. Phys. Geogr. 2, 184–194.
Wohlfahrt, G., Hammerle, A., Haslwanter, A., Bahn, M., Tappeiner, U., Cernusca, A., 2008. Seasonal and inter-annual variability of the net ecosystem CO_2 exchange of a temperate mountain grassland: effects of weather and management. J. Geophys. Res.: Atmos. 113 (D8), 1984–2012.
Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., Patt, F.S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sens. Environ. 83, 31–49Wohlfahrt, G., Hammerle, A., Haslwanter, A., Bahn, M., Tappeiner, U., Cernusca, A., 2008. Seasonal and inter–annual variability of the net ecosystem CO2 exchange of a temperate mountain grassland: Effects of weather and management, Journal of Geophysical Research: Atmospheres, 1984–2012, 113(D8)..
Wu, C., Chen, J.M., Huang, N., 2011. Predicting gross primary production from the enhanced vegetation index and photosynthetically active radiation: evaluation and calibration. Remote Sens. Environ. 115, 3424–3435, http://dx.doi.org/10.1016/j.rse.2011.08.006.
Xiao, X.M., Zhang, Q.Y., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Berrien, M., Ojima, D., 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. Remote Sens. Environ. 91, 256–270, http://dx.doi.org/10.1016/j.rse.2004.03.010.
Xiao, J.F., Zhuang, Q.L., Law, B.E., et al., 2010. A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. Remote Sens. Environ. 114, 576–591, http://dx.doi.org/10.1016/j.rse.2009.10.013.
Xu, L.K., Baldocchi, D.D., 2004. Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. Agric. For. Meteorol. 123, 79–96, http://dx.doi.org/10.1016/j.agrformet.2003.10.004.
Yang, F.H., Ichii, K., White, M.A., et al., 2007. Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach. Remote Sens. Environ. 110, 109–122, http://dx.doi.org/10.1016/j.rse.2007.02.016.
Yuan, W.P., Liu, S., Zhou, G.S., et al., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agric. For. Meteorol. 143, 189–207, http://dx.doi.org/10.1016/j.agrformet.2006.12.001.
Yuan, W., Cai, W., Xia, J., et al., 2014a. Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. Agric. For. Meteorol. 192–193, 108–120, http://dx.doi.org/10.1016/j.agrformet.2014.03.007.
Yuan, W., Cai, W., Liu, S., et al., 2014b. Vegetation-specific model parameters are not required for estimating gross primary production. Ecol. Model. 292, 1–10, http://dx.doi.org/10.1016/j.ecolmodel.2014.08.017.

Zhang, X.Y., Friedl, M.A., Schaaf, C.B., et al., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475, http://dx.doi.org/10.1016/S0034-4257(02)00135-9.
Zhao, M., Heinsch, F., Nemani, R.R., Running, S.W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. Remote Sens. Environ. 95, 164–176, http://dx.doi.org/10.1016/j.rse.2004.12.011.
Zhao, M., Running, S.W., Nemani, R.R., 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. J. Geophys. Res. 111, G01002, http://dx.doi.org/10.1029/2004JG000004.
Zielis, S., Etzold, S., Zweifel, R., Eugster, W., Haeni, M., Buchmann, N., 2014. NEP of a Swiss subalpine forest is significantly driven not only by current but also by previous year's weather. Biogeosciences 11, 1627–1635, http://dx.doi.org/10.5194/bg-11-1627-2014.