Improving ecosystem productivity modeling through spatially explicit estimation of optimal light use efficiency

Nima Madani

John S. Kimball
*University of Montana - Missoula*

David L.R. Affleck
*University of Montana - Missoula*

Jens Kattge

Jon Graham

*See next page for additional authors*

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/ntsg_pubs

**Recommended Citation**

Improving ecosystem productivity modeling through spatially explicit estimation of optimal light use efficiency

Nima Madani1,2, John S. Kimball1,2, David L. R. Affleck3, Jens Kattge4, Jon Graham5, Peter M. van Bodegom6, Peter B. Reich7,8, and Steven W. Running2

1 Flathead Lake Biological Station, University of Montana, Polson, Montana, USA, 2 Numerical Terradynamic Simulation Group, University of Montana, Missoula, Montana, USA, 3 College of Forestry and Conservation, University of Montana, Missoula, Montana, USA, 4 Max Planck Institute for Biogeochemistry, Jena, Germany, 5 Department of Mathematical Sciences, University of Montana, Missoula, Montana, USA, 6 Department of Systems Ecology, VU University Amsterdam, Amsterdam, Netherlands, 7 Department of Forest Resources, University of Minnesota, St. Paul, Minnesota, USA, 8 Hawkesbury Institute for the Environment, University of Western Sydney, Sydney, New South Wales, Australia

Abstract A common assumption of remote sensing-based light use efficiency (LUE) models for estimating vegetation gross primary productivity (GPP) is that plants in a biome matrix operate at their photosynthetic capacity under optimal climatic conditions. A prescribed constant biome maximum light use efficiency parameter (LUEmax) defines the maximum photosynthetic carbon conversion rate under these conditions and is a large source of model uncertainty. Here we used tower eddy covariance measurement-based carbon (CO2) fluxes for spatial estimation of optimal LUE (LUEopt) across North America. LUEopt was estimated at 62 Flux Network sites using tower daily carbon fluxes and meteorology, and satellite observed fractional photosynthetically active radiation from the Moderate Resolution Imaging Spectroradiometer. A geostatistical model was fitted to 45 flux tower-derived LUEopt data points using independent geospatial environmental variables, including global plant traits, soil moisture, terrain aspect, land cover type, and percent tree cover, and validated at 17 independent tower sites. Estimated LUEopt shows large spatial variability within and among different land cover classes indicated from the sparse tower network. Leaf nitrogen content and soil moisture regime are major factors explaining LUEopt patterns. GPP derived from estimated LUEopt shows significant correlation improvement against tower GPP records (R2 = 76.9%; mean root-mean-square error (RMSE) = 257 g C m⁻² yr⁻¹), relative to alternative GPP estimates derived using biome-specific LUEmax constants (R2 = 34.0%; RMSE = 439 g C m⁻² yr⁻¹). GPP determined from the LUEopt map also explains a 49.4% greater proportion of tower GPP variability at the independent validation sites and shows promise for improving understanding of LUE patterns and environmental controls and enhancing regional GPP monitoring from satellite remote sensing.

1. Introduction

Satellite remote sensing, as the only means of monitoring vegetation changes at global scales, has been widely applied to determine plant productivity and ecosystem dynamics [e.g., Running et al., 2004; Xiao et al., 2005; Kimball et al., 2006; Zhao and Running, 2010]. However, estimation of plant carbon uptake at large scales using remote sensing products is bound with uncertainties [Heinsch et al., 2006; Hilker et al., 2008]. While upscaled site level vegetation gross primary production (GPP) estimates show a mean global terrestrial carbon uptake of 123 ± 8 Pg C yr⁻¹ [Beer et al., 2010]; remote sensing productivity estimates show significantly lower levels (~109.3 Pg C yr⁻¹) [Zhao et al., 2005].

Optical remote sensing data-driven methods for estimating GPP generally rely on spectral vegetation indices of photosynthetic canopy cover derived from visible and near-infrared reflectances and other ancillary biophysical inputs including general land cover and plant functional type characteristics, incident solar radiation, and surface meteorology [Kimball et al., 2009]. However, limitations with respect to available ground truth data for model development, calibration and validation, and poor model assumptions are among the main sources of uncertainties in remote sensing-based ecosystem productivity models [Ahl et al., 2004; Yuan et al., 2007]. Additionally, having a good estimation of ecosystem productivity requires detailed knowledge of vegetation phenology [Jin et al., 2013] and canopy photosynthetic response to variations in environmental conditions [Li et al., 2008] within and between plant functional types.
Optical remote sensing-based productivity models such as the Carnegie–Arms–Stanford Approach [Potter et al., 1993], Terrestrial Uptake and Release of Carbon [Ruimy et al., 1996], C-Fix [Veroustraete et al., 2002], and MOD17 [Running et al., 2004] are based on the light use efficiency (LUE) concept [Monteith, 1972; Kumar and Monteith, 1981]. In this logic, plant production is linearly related to photosynthetically active radiation (PAR) absorbed by the vegetation canopy (absorbed photosynthetically active radiation (APAR)) and the efficiency with which this solar radiant energy is transformed into vegetation biomass through net photosynthesis (light use efficiency).

Maximum light use efficiency (LUE_{max}) defines the canopy photosynthetic capacity or maximum rate of conversion of APAR to vegetation biomass (g C MJ^{-1}) under optimal (nonlimiting) environmental conditions [Monteith, 1972]. The LUE_{max} parameter is reduced under suboptimal temperature and water deficit conditions and varies according to vegetation type and environment [Bartlett et al., 1989; Frapani et al., 1992]. This has been the basis of the NASA MODIS (Moderate Resolution Imaging Spectroradiometer) GPP product (MOD17) [Running et al., 2004; Zhao et al., 2005]. The MOD17 GPP product is currently the only remote sensing global operational ecosystem productivity data record and has been in production from MODIS on board the NASA EOS Terra and Aqua satellites since 2000 and 2002, respectively. The MOD17 GPP product uses MODIS-derived fractional photosynthetically active radiation (FPAR) [Myneni et al., 1999] to estimate terrestrial ecosystem GPP globally at 8 day intervals with 1 km spatial resolution. The MOD17 algorithm is based on the assumption that LUE_{max} variability is conservative within individual biomes [Monteith and Moss, 1977] and defines general biophysical response characteristics to estimate GPP using a Biome Property Look-Up Table (BPLUT) [Zhao et al., 2005] and global land cover classification [Friedli et al., 2010] that defines 11 general plant functional types. However, LUE_{max} can show large variability even within the same plant functional type [Goetz and Prince, 1996; Gower et al., 1999; Turner et al., 2002], and the fixed parameter for LUE_{max} is a major source of uncertainty for ecosystem productivity modeling [Ruimy et al., 1994; Way et al., 2005; Pan et al., 2006; Wang et al., 2010]. The biome background matrix defined using fixed LUE_{max} values increases model GPP uncertainty, because spatial heterogeneity in vegetation light use efficiency is underrepresented [Turner et al., 2002].

Within a given biome type, and independent of direct environmental forcings, LUE and hence ecosystem productivity are affected by stand age and soil nutrition [Huston and Wolverton, 2009; Malhi, 2012], leaf nitrogen concentrations [Kergoat et al., 2008; Ollinger et al., 2008; Reich, 2012], and canopy structure and leaf traits [Wright et al., 2004; Jones et al., 2012; Rogers, 2013]. These factors are not directly represented by the relatively simple LUE model logic due to limitations of available biophysical data required for model development and regional simulations.

Factors constraining plant productivity can be divided into three general groups: constraints governing potential carbon uptake (stressor factors, e.g., temperature and vapor pressure deficit (VPD)), inherent plant physiological characteristics (plant functional types and traits, e.g., leaf nitrogen content), and landscape features (e.g., terrain and microclimate regime). Environmental stressor factors such as minimum temperature and VPD directly affect canopy stomatal conductance and photosynthetic carbon uptake. However, the other two factors are ecosystem properties that vary spatially and can influence LUE and ecosystem productivity within individual biomes. While stressor factors can affect photosynthesis at daily and finer time scales, the other two factors are assumed to be temporally conservative at coarse spatial scales and over limited operational satellite records.

When environmental stressor factors (such as temperature and water deficit) are not constraining to photosynthetic carbon gain, then ecosystem optimal light use efficiency (LUE_{opt}) can be estimated from tower eddy covariance measurements of land-atmosphere carbon (CO_{2}) exchange [Kergoat et al., 2008]. Tower eddy covariance measurement networks, including Flux Network (FLUXNET) [Baldocchi et al., 2001], record CO_{2} fluxes and site-specific climate data, including incoming short wave radiation, and provide useful information for validating ecosystem models and understanding terrestrial carbon budgets and underlying environmental controls for different ecosystems [e.g., Running et al., 1999; Yi et al., 2013]. LUE_{opt} is expected to be spatially heterogeneous and lower than the theoretical maximum rate (LUE_{max}) due to other limiting ecosystem morphological and landscape constraints.

The objective of this study is to improve the accuracy of satellite-based LUE model GPP predictions by implementing a spatially explicit estimation of LUE_{opt}. We apply a regression modeling approach using spatially contiguous landscape attributes and in situ tower eddy covariance-based GPP values and supporting
biophysical measurements to estimate LUEopt over a North American domain. A selection of tower sites representing the major North American biomes is used for estimating LUEopt from multiyear observations at each tower site. A geostatistical regression model is then developed to explain the across-site variability in LUEopt using a set of spatially contiguous predictor variables, including general plant traits and landscape features. The LUEopt predictions are evaluated against other observation-based LUEopt values determined from a set of independent tower validation sites. The LUEopt predictions are also used as primary inputs for satellite (MODIS) LUE model-based GPP predictions over the domain. The model GPP results are then evaluated against independent tower GPP data and baseline LUE model simulations derived using biome-specific constant LUEmax inputs.

2. Data and Methods

2.1. In Situ LUEopt Estimation

Sixty-two tower sites from the FLUXNET La Thuile database [Baldocchi, 2008] were selected for this study representing major biomes of North America (see Table S1 in the supporting information). Most of the tower sites selected had multiyear (≥2) daily values for GPP and surface meteorology, while only two sites had a single year of measurements. The tower eddy covariance method intrinsically measures net ecosystem CO2 exchange (NEE), while GPP is estimated by applying a model to the eddy covariance measurements to partition NEE into GPP and respiration components [Stoy et al., 2006; Lasslop et al., 2010]. Only the best quality daily GPP data determined from tower eddy covariance measured NEE were selected. These data denoted as having the best quality control flag (quality control = 1) are either original data or filled with high confidence [Agarwal et al., 2010]. Daily global shortwave solar radiation measured at the tower sites was used to estimate PAR, which represents approximately half of the total incoming solar radiation [Zhao et al., 2005].

The tower LUEopt estimation was based on the assumption that GPP attains a maximum daily rate (defined as ≥98% of the long-term record) at some point over the multiyear measurement record, where LUE and canopy photosynthesis are not limited by one or more environmental stress factors, including light, temperature, or moisture limitations [Kergoat et al., 2008]. In order to avoid the effect of outliers, the higher 0.5% bin of measurements was ignored. Thus, the upper 98–99.5% bin of daily GPP values throughout the measurement years were sampled, representing the maximum estimated daily GPP level (GPPmax) from each site record. For all days with such criteria, LUE was defined as

\[ \text{LUE} = \frac{\text{GPPmax}}{(\text{PAR} \times \text{FPAR})} \]

(1)

using PAR derived from tower incoming short wave solar radiation measurements and colocated MODIS MOD15 FPAR (C5) retrievals [Myhenni et al., 1999]. The MODIS 1 km resolution, 8 day FPAR data were obtained from the Oak Ridge National Laboratory Distributed Active Archive Center (DAAC), where MODIS land products are available in a 7 x 7 km grid centered over individual tower sites. The FPAR data were sampled within 3 x 3 km pixels overlying the individual tower sites, and only the highest-quality values (main algorithm used with no saturation occurrence) were used. The FPAR data were spatially resampled using the MODIS 1 km resolution global land cover product [Friedl et al., 2010] to ensure that all 3 x 3 km windows represented the same land cover type as the local tower footprint. In order to capture the tower footprint, the 3 x 3 km FPAR data were spatially averaged for each 8 day time step [Rahman, 2005], and temporal data gaps were filled using the long-term MODIS FPAR 8 day climatology [Kandasamy et al., 2013]. In order to produce daily FPAR data consistent with daily flux tower GPP values, the continuous 8 day FPAR record was interpolated to a daily time step using smoothing splines [Wahba, 1975]. The FPAR data for each of the 62 tower sites were temporally matched with the local tower GPP and PAR observations.

The resulting mean daily estimated LUE in the defined maximum threshold range from equation (1) was used as the LUEopt value for each tower site and was analyzed within and among the major representative North American biome types represented by the MODIS land cover classification and selected FLUXNET tower sites.

2.2. Modeling LUEopt Patterns

2.2.1. Explanatory Variables Influencing LUEopt Patterns

In order to explain the spatial variability of LUEopt, a set of geospatial environmental data characterizing general canopy traits, climate, and landscape terrain characteristics assumed to be important for ecosystem productivity were considered (see Table 1). Plant traits were characterized by leaf nitrogen content, specific leaf area (SLA; leaf
The long-term average precipitation and temperature, mean annual frozen season, and surface soil moisture dimensionless east and north facing units ranging from with 1 km spatial resolution database and similar methodology but was determined from 154 sampling locations in North America out of available from a global plant trait database 

Variables | Abbreviation | Geophysical Data | Source |
---|---|---|---|
**Landscape characteristics** | MODIS-MOD12 | Land cover types | Friedl et al. [2010] |
| MODIS-MOD13 | Maximum EVI | Huete et al. [2002] |
| MODIS-MOD44 | Percent tree cover | Townshend et al. [2011] |
| Terrain DEM | Elevation (m) | Farr et al. [2007] |
| | Aspect eastness | | |
| | Aspect northness | | |
| **Climate** | Bio1 | Annual temperature (°C) | Hijmans et al. [2005] |
| | Bio12 | Annual precipitation (mm) | | |
| | NOAH GLDAS | Soil moisture (kg m⁻²) | Rodell et al. [2004] |
| | Frozen days | Average number of frozen days | Kim and Kimball [2011] and Kim et al. [2012] |
| **Plant traits** | SLA | Specific leaf area | Katge et al. [2011] |
| | Leaf nitrogen content | Leaf nitrogen per dry mass (%) | Simard et al. [2011] |
| | Height | Global tree height (m) | | |
| **Others** | Ecoregions | Global Ecoregions | Olson et al. [2001] |
| Stand Age | North America stand age map | Pan et al. [2011] |

Denotes the variables used in the final model. 
Land cover types based on University of Maryland scheme includes: Water, Evergreen Needleleaf Forest, Evergreen Broadleaf Forest, Deciduous Needleleaf Forest, Deciduous Broadleaf Forest, Mixed Forests, Closed Shrublands, Open Shrublands, Woody Savannas, Grasslands, Croplands, Urban and Built-Up, Barren and Sparsey Vegetated, and Unclassified.

The SLA map (Figure S2 in the supporting information) was derived from the same MODIS-MOD44 percent tree cover [Townshend et al., 2011] and MOD13A3 16 day Enhanced Vegetation Index (EVI) products [Huete et al., 2002] were mosaicked to make a single map for each product over the North American domain. The maximum EVI was produced using the maximum recorded values for a pixel over the period of record. Forest stand age was obtained from Pan et al. [2011], which provides forest age map products at 1 km resolution for Canada and the United States.

Landscape features and terrain were characterized by elevation, aspect, and canopy cover geospatial data. Elevation was obtained from a global Shuttle Radar Topography Mission-based digital elevation map (DEM) with 1 km spatial resolution [Farr et al., 2007]. Terrain aspect was derived from the DEM and converted to dimensionless east and north facing units ranging from -1 to 1, where flat terrain is 0, and 1 denotes maximum eastward and northward aspects [Zar, 1999].

Selected MODIS land products spanning the satellite record from 2000 to 2006 including the MOD12Q1 IGBP static (2004) land cover classification [Friedl et al., 2010], MOD44 percent tree cover [Townshend et al., 2011], and MOD13A3 16 day Enhanced Vegetation Index (EVI) products [Huete et al., 2002] were used to characterize potential climate characteristics influencing LUETOP. Global temperature and precipitation averages were obtained at 1 km spatial resolution from the WorldClim database; these data are interpolated from 47,554 and 24,542 global weather stations for precipitation and temperature, respectively [Hijmans et al., 2005]. A global ecoregion map was obtained from the world wildlife fund covering 867 land units of distinct biotas [Olson et al., 2001] and was used to define climate zones by aggregating temperature and precipitation within each ecoregion.
The mean annual frozen season (days) from 2000 to 2006 was derived from a consistent global classification of daily landscape freeze-thaw status derived from satellite passive microwave remote sensing [Kim and Kimball, 2011; Kim et al., 2012] and distributed through the National Snow and Ice Data Center DAAC. Global monthly average layer one soil moisture data were obtained from the Global Land Data Assimilation System (GLDAS), representing the outputs of four land surface models and extending from 1948 to 2010 [Rodell et al., 2004]. The aggregated monthly soil moisture data from GLDAS at 0.25° (~27 km) resolution were averaged and resampled to the baseline 1 km spatial resolution geographic projection of this investigation using bilinear interpolation.

### 2.2.2. Spatial Modeling of \( \text{LUE}_{\text{opt}} \)

A generalized additive model (GAM) [Hastie et al., 1986; Wood, 2006] was used to estimate the spatial variability of \( \text{LUE}_{\text{opt}} \) across 45 FLUXNET tower sites having multiyear daily carbon flux measurements. The GAM describes the relationship between two or more explanatory variables and the response variable (\( \text{LUE}_{\text{opt}} \)) by fitting additive and smoothed functions of those explanatory variables [Wood, 2006]. The GAM generally adds nonparametric smoothers to the parametric part of a generalized linear model (GLM). This model can provide improvement over a GLM through the addition of appropriate smoothing functions [Guisan et al., 2002].

The general model structure followed by Wood [2006] is

\[
g(\mu_i) = X_i^\top \theta + f_1(X_{i1}) + f_2(X_{i2}) + f_3(X_{i3}) + \ldots
\]

where \( \mu_i = E(Y_i) \) and \( Y_i \) follows an exponential family distribution. \( Y_i \) is the response variable, \( X_i^\top \) is the \( i \)th row of the model matrix, \( \theta \) is the corresponding parameter vector so that \( X_i^\top \theta \) is a linear function of \( \theta \), and \( f_i \) is the \( i \)th smoothed function of \( X \) covariates.

The GAM was optimized using stepwise variable selection by means of Akaike information criterion [Burnham and Anderson, 2002]. The spatially correlated residuals of the best model \( c(s) \) (the stochastic part of the model) were explained using a semivariogram model [Pebesma, 2004; Hengl et al., 2009], based on the assumption that flux towers with the same land cover type that are closer to each other are more likely to have similar \( \text{LUE}_{\text{opt}} \) values. The spatial continuity between pairs of flux towers was first examined using the empirical semivariogram:

\[
\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{(i,j) \mid |h_{ij}|=h} (y_i - y_j)^2
\]

where \( n(h) \) is the number of pairs of points separated by vector \( h \), \( h = (h_1, h_2, \ldots) \) is the vector of orientation between sites \( i \) and \( j \), and \( y_i \) is the observed response at the \( i \)th location. After calculation of the semivariogram, an isotropic exponential variogram model was fitted

\[
\hat{\gamma}(h) = \begin{cases} 
0 & h = 0 \\
\sigma^2[a + (a^2 - a) \left(1 - e^{-3|h||r}\right)] & h \neq 0
\end{cases}
\]

where \( a \geq 0, \sigma^2 > 0, \) and \( r \geq 0 \); where \( h \) is the distance, \( a \) is the nugget effect, \( \sigma^2 \) is the sill, and \( r \) is the range.

The full model (regression kriging) representing the stochastic and semivariogram deterministic terms was then used for spatial prediction of \( \text{LUE}_{\text{opt}} \) at 1 km resolution over the North American domain using the selected spatially contiguous explanatory variables (represented in Table 1) and was cross validated using tower-based \( \text{LUE}_{\text{opt}} \) values from a set of 17 independent tower sites representing major North American biomes (For the location of test sites, refer to Figure S4 in the supporting information).

### 2.3. Gross Primary Production Modeling

In order to assess potential gain in GPP accuracy by using the spatially explicit \( \text{LUE}_{\text{opt}} \) estimates, both the predicted \( \text{LUE}_{\text{opt}} \) and the MOD17 \( \text{LUE}_{\text{max}} \) values were used to estimate GPP and compared against all tower flux measurement-based GPP records. The MOD17 algorithm [Running et al., 2004] was used for predicting GPP at a daily time step over a single year per tower site with the lowest fraction of missing values. Two sets of GPP simulations were conducted as follows:

\[
\text{GPP}_{\text{LUEopt}} = (\text{LUE}_{\text{opt}} \times f_{\text{VPD}} \times f_T) \times \text{FPAR} \times \text{PAR}
\]

\[
\text{GPP}_{\text{LUEmax}} = (\text{LUE}_{\text{max}} \times f_{\text{VPD}} \times f_T) \times \text{FPAR} \times \text{PAR}
\]

where \( f_{\text{VPD}} \) and \( f_T \) are vapor pressure deficit and temperature scalars that reduce \( \text{LUE}_{\text{max}} \) and \( \text{LUE}_{\text{opt}} \) under suboptimal conditions that vary by land cover type [Zhao et al., 2005]. The daily meteorological inputs
Figure 1. Location of all FLUXNET tower sites used in this investigation and the relative magnitudes of estimated \(LUE_{opt}\) (g C MJ\(^{-1}\)) for each site, overlain on the MODIS-MOD17 maximum light use efficiency (\(LUE_{max}\)) values based on MODIS (MOD12Q1) IGBP land cover classes, including OSH (Open Shrubland), GRA (Grassland), CRO (Croplands), MF (Mixed Forest), DNF (Deciduous Needleleaf Forest), DBF (Deciduous Broadleaf Forest), SA (Savanna), WSA (Woody Savanna), EBF (Evergreen Broadleaf Forest), and CSH (Closed Shrubland).

(VPD, \(T\), and PAR) were derived from tower site level measurements, while daily FPAR was obtained from the MODIS MOD15 FPAR (C5) product. The MODIS BPLUT for the VPD and \(T\) scalars is summarized in Table S3 in the supporting information.

3. Results

3.1. \(LUE_{opt}\) Estimation for Selected Tower Sites

\(LUE_{opt}\), derived by applying the LUE model in equation (1) range from 0.28 to 2.82 (g C MJ\(^{-1}\)) for tower sites representing the different land cover types. Shrubland sites have the lowest \(LUE_{opt}\) (0.508 ± 0.01), while cropland sites have the highest \(LUE_{opt}\) rates of the nine major North American land cover and plant functional types represented. The estimated tower \(LUE_{opt}\) results show high spatial variability both within and among biome types (Figures 1 and 2), while croplands have the highest range of \(LUE_{opt}\) spatial variability. The \(LUE_{opt}\) results are summarized for individual tower sites in Table S1 in the supporting information.

Comparison of average \(LUE_{opt}\) values for each land cover type derived from this study with the biome-specific constant \(LUE_{max}\) values used in the MODIS MOD17 LUE algorithm [Running et al., 2004] shows relatively large differences in these parameters for similar land cover types (Table 2). The MOD17 algorithm prescribes much lower optimum LUE rates for croplands, deciduous broadleaf forest, grasslands, and mixed forest plant functional types than the estimated mean \(LUE_{opt}\) rates inferred from the tower GPP values, implying that MOD17 underestimates GPP for these land cover types (at least under near optimal conditions); the largest LUE differences occur for croplands, where the \(LUE_{opt}\) results are approximately 2.5 times larger than MOD17. The prescribed MOD17 \(LUE_{max}\) rates are also lower than the lower 95th percentile of tower-estimated \(LUE_{opt}\) rates for cropland (1.04 versus 1.77 g C MJ\(^{-1}\)), grassland (0.86 versus 0.91 g C MJ\(^{-1}\)), and deciduous broadleaf forest (1.16 versus 1.29 g C MJ\(^{-1}\)), which are three important biomes that cover ~20% of the total area (and ~1/3 of the vegetated area) of the study domain. In contrast, MOD17 \(LUE_{max}\) is approximately 2.5 times higher than the upper 95th percentile of \(LUE_{opt}\) for closed shrublands (1.28 versus 0.53 g C MJ\(^{-1}\)).
3.2. LUE\textsubscript{opt} Prediction Over the North American Domain

The best GAM used to derive LUE\textsubscript{opt}, based on its predictive power (root-mean-square error (RMSE)) in estimating LUE\textsubscript{opt} from the tower test sites, used land cover type (grasslands and croplands), leaf nitrogen content, soil moisture, percent tree cover, and terrain aspect eastness coefficients. The model showed an adjusted $R^2$ of 68.2% with an associated RMSE of 0.22 g C MJ\textsuperscript{-1} for the tower training site-derived LUE\textsubscript{opt} values. The GAM used a linear relationship between all covariates except soil moisture, which showed a nonlinear relationship (Figure S6 in the supporting information) with tower observation-derived LUE\textsubscript{opt}. The full model (combination of GAM and kriging of residuals) increased the GAM accuracy in explaining LUE\textsubscript{opt} spatial variability among the tower training sites ($R^2 = 88.6\%$, RMSE = 0.15 g C MJ\textsuperscript{-1}) and validation sites ($R^2 = 91.1\%$, RMSE of 0.22 g C MJ\textsuperscript{-1}), so that 88.6% of the variance across all sites (training and validation) was explained (RMSE = 0.17 g C MJ\textsuperscript{-1}) (Figure 3). As indicated by Kergoat et al. [2008], leaf nitrogen content was the most important factor explaining spatial variability in LUE\textsubscript{opt}, accounting for 43% of the observed spatial variability across all tower sites ($p < 0.0001$) (Figure S5 in the supporting information). Surface soil moisture was the next most important factor, accounting for 19% ($p < 0.0001$) of LUE\textsubscript{opt} variability. Among land cover types, only grassland and cropland identifiers were significant in explaining LUE\textsubscript{opt} variability between sites ($p < 0.01$ and $p < 0.0001$, respectively) and were used in the fitted GAM (GAM coefficients of the covariates are summarized in Table S2 in the supporting information). The resulting model predictions showed that LUE\textsubscript{opt} for the three training cropland sites were underestimated by 11% compared to the tower-estimated LUE\textsubscript{opt} for these sites.

Table 2. Comparison of Biome Average and Spatial Standard Deviation of Estimated LUE\textsubscript{opt} (g C MJ\textsuperscript{-1}) From This Study and the Prescribed Biome-Specific LUE\textsubscript{max} Values Used in the MODIS MOD17 Operational Algorithm and GPP Product for North American Land Cover Types With the Greatest Proportional Coverage Over the Study Domain

<table>
<thead>
<tr>
<th>Land Cover Type</th>
<th>Area (%)</th>
<th>LUE\textsubscript{opt} (This Study)</th>
<th>LUE\textsubscript{max} (MOD17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Shrubland (OSH)</td>
<td>27.86</td>
<td>0.631 ± 0.37</td>
<td>0.841</td>
</tr>
<tr>
<td>Evergreen Needleleaf Forest (ENF)</td>
<td>12.16</td>
<td>0.835 ± 0.24</td>
<td>0.962</td>
</tr>
<tr>
<td>Cropland (CRO)</td>
<td>10.86</td>
<td>2.201 ± 0.66</td>
<td>1.044</td>
</tr>
<tr>
<td>Grassland (GRA)</td>
<td>7.81</td>
<td>1.294 ± 0.32</td>
<td>0.962</td>
</tr>
<tr>
<td>Mixed Forest (MF)</td>
<td>6.94</td>
<td>1.171 ± 0.23</td>
<td>1.051</td>
</tr>
<tr>
<td>Woody Savanna (WSA)</td>
<td>5.53</td>
<td>0.983 ± 0.20</td>
<td>1.239</td>
</tr>
<tr>
<td>Deciduous Broadleaf Forest (DBF)</td>
<td>5.12</td>
<td>1.453 ± 0.14</td>
<td>1.165</td>
</tr>
<tr>
<td>Evergreen Broadleaf Forest (EBF)</td>
<td>1.50</td>
<td>0.980 ± 0.21</td>
<td>1.268</td>
</tr>
<tr>
<td>Closed Shrubland (CSH)</td>
<td>0.50</td>
<td>0.508 ± 0.01</td>
<td>1.281</td>
</tr>
</tbody>
</table>

Figure 2. Box plot of estimated tower LUE\textsubscript{opt} (g C MJ\textsuperscript{-1}) representing different land cover (MODIS MOD12Q1) types for 62 FLUXNET sites having multiyear data records that were used in this study. Each symbol (black circles) represents an individual tower site grouped by land cover class; the numbers above each box plot denote the number of tower sites representing the specified land cover class, including OSH (Open Shrubland), ENF (Evergreen Needleleaf Forest), CRO (Croplands), GRA (Grassland), MF (Mixed Forest), WSA (Woody Savanna), DBF (Deciduous Broadleaf Forest), EBF (Evergreen Broadleaf Forest), and CSH (Closed Shrubland); and the asterisk denotes MODI7 maximum light use efficiency (LUE\textsubscript{max}) values (g C MJ\textsuperscript{-1}) for each land cover type.
The predicted LUE_{opt} map for the North American domain is defined at 1 km spatial resolution consistent with the MODIS land cover classification inputs and shows that, as expected, areas of intense agriculture have the highest regional LUE_{opt} (Figure 4a). The model predictions also show that the largest relative prediction error is associated with regions where there were only a limited number of tower sites available for model development (Figure 4b). These areas are generally cropland regions (Figure 1).

### 3.3. GPP Modeling Improvements

Predicted LUE_{opt} at the tower sites was used as an ancillary input to the MOD17 LUE algorithm in place of a prescribed LUE_{max} for each biome type to estimate daily GPP at each site; the GPP simulations were conducted at each site over one specific year having the longest available tower GPP record. The resulting daily GPP estimates derived using LUE_{opt} (GPP_{LUEopt}) estimation were evaluated against independent tower GPP values and daily GPP estimates determined using prescribed LUE_{max} constants (GPP_{LUEmax}) from the MOD17 algorithm. Tower site comparisons between LUE_{opt} and MOD17 LUE_{max} and the resulting annual GPP simulations are summarized in Figures 5a and 5b. These results indicate that the use of a spatially explicit LUE_{opt} input dramatically improves GPP estimation accuracy relative to the tower data and baseline MOD17 calculations derived from biome-prescribed LUE_{max} constants. A list of sites used for validation and associated summary of annual GPP results for

---

**Figure 3.** The relationship between model-predicted and tower-estimated LUE_{opt} (g C MJ^{-1}) for all FLUXNET sites represented in this study. The predicted LUE_{opt} values account for 88.6% ($R^2$) of spatial variability in tower GPP, with associated RMSE differences of 0.17 g C MJ^{-1}.

**Figure 4.** (a) Predicted map of LUE_{opt} (g C MJ^{-1}) for the North American domain, encompassing all vegetated land areas. Croplands have the highest estimated LUE_{opt}, with generally lower LUE_{opt} levels at higher latitudes and elevations. Gray shading denotes mean LUE_{opt} variability by latitude and longitude; latitudinal means range from 0.2 to 1.6 (g C MJ^{-1}), and longitudinal means range from 0.2 to 1.2 (g C MJ^{-1}). (b) Map of the model standard error as a proportion (percentage) of the predicted LUE_{opt}; greater relative model uncertainty occurs over regions with sparse tower data.
Comparison of annual GPP (g C m\(^{-2}\) yr\(^{-1}\)) estimated from daily LUE model simulations using (a) LUE\(_{\text{opt}}\) and (b) baseline-prescribed LUE\(_{\text{max}}\) inputs relative to tower GPP from all the 62 tower sites and 17 independent tower test sites (inset); the symbols denote the dominant land cover type of each tower site, while symbol colors denote the magnitude of LUE\(_{\text{opt}}\) and LUE\(_{\text{max}}\) used to estimate GPP.

Each tower site and model are presented in Table S4 in the supporting information. The RMSE difference between GPP-LUE\(_{\text{opt}}\) and the tower data was 257 g C m\(^{-2}\) yr\(^{-1}\) for all 62 tower sites, which represented a 41% improvement in model performance relative to the baseline MOD17 GPP-LUE\(_{\text{max}}\) calculations (RMSE: 439 g C m\(^{-2}\) yr\(^{-1}\)). The model results for the independent tower validation sites also indicate that the LUE\(_{\text{opt}}\) inputs provide more accurate simulations of observed annual GPP spatial variability compared to the baseline MOD17 simulations, accounting for more than 4 times as much observed GPP variability ($R^2$ of 64.2% versus $R^2$ of 14.8%). For both sets of simulations, we used local tower meteorological data, including incoming shortwave solar radiation, air temperature, and MODIS FPAR inputs, and the only differences were the baseline LUE\(_{\text{max}}\) and model-predicted LUE\(_{\text{opt}}\) inputs to the LUE model GPP calculations.

The model GPP results show that for all land cover types represented, the predicted LUE\(_{\text{opt}}\) inputs produce more accurate predictions of ecosystem GPP (Table 3) relative to the use of prescribed biome-specific LUE\(_{\text{max}}\) constants in the baseline MOD17 simulations. The seasonal progression of estimated daily GPP derived from alternative

<table>
<thead>
<tr>
<th>Land cover type</th>
<th>GPP-LUE(_{\text{opt}})</th>
<th>GPP-LUE(_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>RMSE$^c$</td>
</tr>
<tr>
<td>Open Shrubland (OSH)</td>
<td>0.34</td>
<td>177</td>
</tr>
<tr>
<td>Evergreen Needle Leaf Forest (ENF)</td>
<td>0.60</td>
<td>308</td>
</tr>
<tr>
<td>Cropland (CRO)</td>
<td>0.75</td>
<td>233</td>
</tr>
<tr>
<td>Grassland (GRA)</td>
<td>0.93</td>
<td>158</td>
</tr>
<tr>
<td>Mixed Forest (MF)</td>
<td>0.72</td>
<td>238</td>
</tr>
<tr>
<td>Woody Savanna (WSA)</td>
<td>0.20</td>
<td>20</td>
</tr>
<tr>
<td>Deciduous Broadleaf Forest (DBF)</td>
<td>0.66</td>
<td>223</td>
</tr>
<tr>
<td>Evergreen Broadleaf Forest (EBF)</td>
<td>0.44</td>
<td>443</td>
</tr>
<tr>
<td>Closed Shrubland (CSH)</td>
<td>-</td>
<td>245</td>
</tr>
</tbody>
</table>

$^a$For all land cover types, GPP is modeled using LUE\(_{\text{opt}}\) inputs and shows improvements over the baseline MOD17 simulations derived using prescribed LUE\(_{\text{max}}\) inputs.

$^b$MRE (mean residual error).

$^c$RMSE (root-mean-square error).

$^d$The hyphen (-) denotes only two tower sites representing the land cover type.
LUE\textsubscript{opt} and LUE\textsubscript{max} inputs is presented in Figure 6 for four selected tower sites representing the major North American biome types, based on having the largest aerial coverage within the domain and representing shrubland, evergreen needleleaf forest, cropland, and grassland biome types. The prescribed LUE\textsubscript{max} inputs from the MOD17 operational algorithm are generally higher than the average estimated LUE\textsubscript{opt} values for closed shrubland and evergreen needleleaf forest and lower for cropland and grasslands, while LUE\textsubscript{opt} levels from this investigation are generally independent of land cover type and exhibit large spatial variability. Overall, the use of alternative LUE\textsubscript{opt} inputs leads to 58.5% lower RMSE differences and 42.9% higher daily \( R^2 \) correspondence against the tower data and relative to the LUE\textsubscript{max}-based GPP simulations at the tower sites.

4. Discussion

The tower analysis showed that LUE\textsubscript{opt} is spatially heterogeneous within individual land cover types and across the landscape. The largest LUE\textsubscript{opt} variability was within croplands, which also showed the highest LUE\textsubscript{opt} compared to other land cover types. Previous studies have shown that LUE\textsubscript{max} is underestimated for Maize and Soybean in the MODIS MOD17 algorithm [Xin et al., 2013], and a cropland field study showed that LUE\textsubscript{max} of Maize can be as high as 3.84 g C MJ\(^{-1}\) [Lindquist et al., 2005]. Spatial modeling of tower-estimated LUE\textsubscript{opt} facilitated GAM-based estimation of LUE\textsubscript{opt} regional patterns across North America. Among the factors influencing LUE\textsubscript{opt} patterns, percent tree cover as a surrogate for vegetated area was positively associated with LUE\textsubscript{opt}. Terrain aspect as a surrogate for solar illumination and microclimate (air temperature and precipitation) has the lowest impacts as predictor variables in our model. Our results also highlight the importance of soil moisture and leaf nitrogen content in determining LUE\textsubscript{opt} patterns. Leaf
nitrogen content, because of its relationship with rubisco and photosynthetic capacity, is an important factor determining plant productivity. The availability of large databases for plant traits [Kattge et al., 2011] provided a unique opportunity to include leaf nitrogen content data in the \( \text{LUE}_{\text{opt}} \) prediction model. Leaf traits such as leaf nitrogen content and SLA can be temporally dynamic in response to seasonal canopy changes, stand age, and disturbance recovery [Parolin et al., 2002; Nouvellon et al., 2010], affecting canopy photosynthetic capacity, light use efficiency, and GPP; and others have shown that canopy N can explain a substantial amount of variation in productivity from minute to year scales [Kergoat et al., 2008; Ollinger et al., 2008; Reich, 2012]. Our investigation focus was on estimating spatial patterns in \( \text{LUE}_{\text{opt}} \) using limited ground observations of general leaf traits, while the effects of temporal leaf trait variations on \( \text{LUE}_{\text{opt}} \) and associated GPP calculations were not explicitly represented and require further investigation. Some attempts have been made to infer LUE by estimating leaf nitrogen or chlorophyll content using remote sensing [e.g., Grace et al., 2007; Goerner et al., 2009; Frankenberg et al., 2011; Schlemmer et al., 2013], which is expected to improve global ecosystem productivity estimation. However, the general methods and leaf trait data used in this study, including leaf nitrogen, show promising results for spatially continuous estimation of \( \text{LUE}_{\text{opt}} \) and improved estimation of GPP spatial and temporal dynamics.

The GAM results showed that a static land cover classification was only useful when grassland and cropland classes were considered as independent predictors of \( \text{LUE}_{\text{opt}} \) spatial variability. The covariates used in this study were also generally insufficient for explaining \( \text{LUE}_{\text{opt}} \) spatial variability within cropland and grassland areas due in part to sparse tower representation of grassland and cropland heterogeneity. These results also imply the need for testing other covariates, including temporally dynamic land cover (e.g., maize and soybean rotation) and irrigation regime inputs, as potential explanatory variables in these areas. However, despite the fact that our model underestimated \( \text{LUE}_{\text{opt}} \) for these sites (Figure 3), GPP estimation accuracy was significantly improved using estimated \( \text{LUE}_{\text{opt}} \) compared to baseline \( \text{LUE}_{\text{max}} \) inputs to the MOD17 algorithm.

In this study, we used a limited number of predictor variables reported to have an important role in the carbon uptake capacity of ecosystems to predict \( \text{LUE}_{\text{opt}} \) at regional scales. Other factors such as groundwater storage and soil chemistry also influence plant photosynthesis and may provide additional landscape characteristics for estimating \( \text{LUE}_{\text{opt}} \) and productivity. Additionally, uncertainty associated with the covariates used in the spatial regression model likely has a negative impact on \( \text{LUE}_{\text{opt}} \) and GPP accuracy. The model predictions and validation activities from this study were also derived from a relatively sparse North American tower observation network that may not fully capture the range of variability in regional vegetation and climate patterns. Despite these limitations and uncertainties, our model results produce regional patterns of \( \text{LUE}_{\text{opt}} \) with favorable accuracy that enhanced the accuracy of higher-order GPP simulations from a satellite data-driven LUE algorithm. These findings and the global availability of similar plant traits information and geospatial data required for model extrapolation imply the potential for similar global mapping of \( \text{LUE}_{\text{opt}} \) utilizing more extensive plant trait and tower (FLUXNET) measurement records available from global networks and spanning a broader range of global biomes. New global biophysical data from next generation satellite sensors may also lead to better \( \text{LUE}_{\text{opt}} \) and GPP predictions; these new observations include canopy fluorescence, landscape freeze-thaw, and soil moisture dynamics from the NASA Orbiting Carbon Observatory 2 and Soil Moisture Active Passive missions that may provide near direct measures of LUE and underlying environmental controls. \( \text{LUE}_{\text{opt}} \) also likely varies temporarily with changes in vegetation and environmental conditions, whereas this study only provides static map of \( \text{LUE}_{\text{opt}} \) from limited tower observations and spatially coarse geophysical data. Future research and new satellite observations may enable temporal modeling of \( \text{LUE}_{\text{opt}} \) while also considering land cover change and disturbance recovery impacts. This is especially important for cropland regions with annual rotation of C4 and C3 crops.

Here our primary focus was on predicting optimal light use efficiency, while \( \text{LUE}_{\text{opt}} \) predictions were used for enhancing GPP estimation accuracy relative to using prescribed biome \( \text{LUE}_{\text{max}} \) constants in the MOD17 LUE model. The MOD17 algorithm for GPP modeling only accounts for VPD and air temperature as direct daily stressor factors. Even though \( \text{GPP}_{\text{LUEopt}} \) showed improved accuracy over the \( \text{GPP}_{\text{LUEmax}} \) estimates and relative to the tower GPP data, the remaining unexplained variance between the model predictions and tower GPP values imply that additional environmental factors are needed to further improve model accuracy (Figure 6); the application of additional dynamic stressor factors such as freeze-thaw and soil moisture status may improve the seasonality of LUE-modeled GPP [Kimball et al., 2009].
In summary, our results showed that the spatial variability in LUEopt between sites and within individual biome types should not be ignored. The LUEopt predictions from this study led to a 42.9% improvement in LUE model and tower GPP correspondence. The LUEopt retrievals show large spatial variability that is largely independent from the static land cover classification data, and even though the current study domain was limited to North America, the potential exists for extrapolating these methods to a global domain using a larger global tower network [Baldocchi, 2008] and plant traits database [Kattge et al., 2011]. Spatially explicit LUEopt data derived from general landscape characteristics and plant traits information, and associated improvements in GPP estimation accuracy, should promote better understanding of terrestrial carbon sinks and sources and biospheric capacity for mitigation of the human carbon footprint.

Acknowledgments
This study was conducted with funding provided by the National Aeronautics and Space Administration (NASA) Earth Science program (NNX11AD46G, NNNX09AP62G). The study has been supported by the TRY initiative on plant traits (http://www.try-db.org). The TRY initiative and database is hosted, developed, and maintained by J. Kattge and G. Bittern (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is has been supported by DIVERSITAS, IGBP, the Global Land Project, the UK Natural Environment Research Council through its program QUEST (Quantifying and Understanding the Earth System), the French Foundation for Biodiversity Research, and GIS “Climat Environment et Societe,” France. 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)), Fluxnet-Canada (supported by CFCA, NSERC, BICAP, Environment Canada and NRCan). We acknowledge the financial support to the eddy covariance data harmonization provided by CARBOEurope, FAO-GTOS-TCO, ELEPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval, Environment Canada and U.S. 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.

References
Fried, M. A., D. Sull-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang (2010), MGDIS Collection 5 global land cover: Independent from the static land cover classification data, and even though the current study domain was limited to North America, the potential exists for extrapolating these methods to a global domain using a larger global tower network [Baldocchi, 2008] and plant traits database [Kattge et al., 2011]. Spatially explicit LUEopt data derived from general landscape characteristics and plant traits information, and associated improvements in GPP estimation accuracy, should promote better understanding of terrestrial carbon sinks and sources and biospheric capacity for mitigation of the human carbon footprint.