Modeling gross primary production of maize and soybean croplands using light quality, temperature, water stress, and phenology

Anthony L. Nguy-Robertson  
*University of Nebraska-Lincoln*

Andrew E. Suyker  
*University of Nebraska-Lincoln*, asuyker1@unl.edu

Xiangming Xiao  
*University of Oklahoma*, xiangming.xiao@ou.edu

Follow this and additional works at: [http://digitalcommons.unl.edu/natrespapers](http://digitalcommons.unl.edu/natrespapers)

Part of the [Atmospheric Sciences Commons](http://digitalcommons.unl.edu/natrespapers), [Meteorology Commons](http://digitalcommons.unl.edu/natrespapers), [Natural Resources and Conservation Commons](http://digitalcommons.unl.edu/natrespapers), [Natural Resources Management and Policy Commons](http://digitalcommons.unl.edu/natrespapers), and the [Other Environmental Sciences Commons](http://digitalcommons.unl.edu/natrespapers)

[http://digitalcommons.unl.edu/natrespapers/526](http://digitalcommons.unl.edu/natrespapers/526)

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Modeling gross primary production of maize and soybean croplands using light quality, temperature, water stress, and phenology

Anthony Nguy-Robertson,¹ Andrew Suyker,¹ and Xiangming Xiao²,3

¹ School of Natural Resources, University of Nebraska–Lincoln, Lincoln, NE
² Center for Spatial Analysis, College of Atmospheric and Geographic Sciences, University of Oklahoma, Norman, OK
³ Department of Microbiology and Plant Biology, College of Arts and Sciences, University of Oklahoma, Norman, OK

Corresponding author — A. Suyker, asuyker1@unl.edu

Abstract
Vegetation productivity metrics, such as gross primary production (GPP) may be determined from the efficiency with which light is converted into photosynthates, or light use efficiency (ε). Therefore, accurate measurements and modeling of ε is important for estimating GPP in each ecosystem. Previous studies have quantified the impacts of biophysical parameters on light use efficiency based GPP models. Here we enhance previous models utilizing four scalars for light quality (i.e., cloudiness), temperature, water stress, and phenology for data collected from both maize and soybean crops at three Nebraska AmeriFlux sites between 2001 and 2012 (maize: 26 field-years; soybean: 10 field-years). The cloudiness scalar was based on the ratio of incident photosynthetically active radiation (PAR_in) to potential (i.e., clear sky) PAR_pot. The water stress and phenology scalars were based on vapor pressure deficit and green leaf area index, respectively. Our analysis determined that each parameter significantly improved the estimation of GPP (AIC range: 2503–2740; likelihood ratio test: p-value < 0.0003, df = 5–8). Daily GPP data from 2001 to 2008 calibrated the coefficients for the model with reasonable amount of error and bias (RMSE = 2.2 gCm⁻²d⁻¹; MNB= 4.7%). Daily GPP data from 2009 to 2012 tested the model with similar accuracy (RMSE = 2.6 gCm⁻²d⁻¹; MNB= 1.7%). Modeled GPP was generally within 10% of measured growing season totals in each year from 2009 to 2012. Cumulatively, over the same four years, the sum of error and the sum of absolute error between the measured and modeled GPP, which provide measures of long-term bias, was ±5% and 2–9%, respectively, among the three sites.

Keywords: Gross primary production, Light use efficiency, Maize, Soybean, Modeling

1. Introduction
The efficiency of light converted into photosynthates, or light use efficiency (ε), is a useful measure of crop productivity (Monteith, 1972). Light use efficiency can be measured at the leaf (Garbulsky et al., 2013), plant (Onoda et al., 2014), or ecosystem/landscape level (Binkley et al., 2013). It is at the landscape level where light use efficiency is used as an important component of many ecosystem production models (e.g., Gilmanov et al., 2013; John et al., 2013) determining net and gross primary production (NPP and GPP, respectively). Therefore, accurate measurements and modeling of ε is important for estimating vegetation productivity in a variety of ecosystems. Many factors impact ε such as water content (e.g., Inoue and Peñuelas, 2006), nitrogen content (e.g., Peltoniemi et al., 2012), temperature (e.g., Hall et al., 2012), and CO₂ concentration (e.g., Haxeltine and Prentice, 1996). Because of the impacts of these factors, a maximum light use efficiency (εₒ) is typically used in ecosystem productivity models (e.g., Li et al., 2012) and down-regulated as environmental conditions change. However, there are known assumptions and errors associated with using εₒ (Xiao, 2006) and improvements in estimating light use efficiency is necessary to improve these ecosystem production models.

Incorporating light quality, a major factor impacting ε (Gu et al., 2003), has been shown to improve ecosystem productivity models (Knobl and Baldocchi, 2008; Suyker and Verma, 2012). This is due to the sensitivity of ε to the light climate in the canopy (He et al., 2013; Zhang et al., 2011). The light quality impact suggests ε should not be defined as a down-regulated maximum value, but as a clear sky value that decreases due to environmental stress and increases due to cloud cover. The light use efficiency has been shown to increase under diffuse light conditions (Gu et al., 2002) in relation to the ratio of diffuse (PAR_d) to incident photosynthetically active radiation (PAR_in) (Schwalm et al., 2006). As diffuse light is not frequently measured, it would be advantageous to have an
alternative to PAR/PAR. Turner et al. (2003) defined a cloudiness coefficient (CC) based on PAR, and the clear-sky potential of photosynthetically active radiation (PAR). The CC was used as a proxy for the quality of light affecting e but not incorporated into their light use efficiency model.

The Vegetation Photosynthesis Model (VPM) is a light use efficiency model that utilizes remote sensing imagery to estimate GPP based on the impacts of temperature, water stress, and phenology (Xiao et al., 2004). These particular factors impact e because (1) plants are affected but can recover quickly (i.e., short-term) from unfavorable temperatures (Crafts-Brandner and Law, 2000), (2) plants take longer to recover (i.e., long-term) from prolonged water stress (Miyashita et al., 2005; Souza et al., 2004), and (3) leaf age impacts photosynthesis rates (Reich et al., 1991). Richardson et al. (2012) indicated that accurate estimates of phenology were necessary for modeling productivity because errors can lead to large biases in cumulative estimates of GPP. In using satellite imagery, the VPM in most situations cannot be applied daily due to limited frequency of clear sky imagery and thus, would not include the impact of light quality on GPP estimates.

However, models incorporating satellite data (e.g., VPM) are critical in developing regional/global estimates of GPP (Yuan et al., 2010). In this study, we adapt a remote sensing-based light use efficiency model to in-situ meteorological (e.g., temperature, VPD) and biophysical data (e.g., green LAI) to estimate the impacts of temperature, water stress, and phenology on e in order to estimate daily GPP. We note that with the development of gridded meteorological data sets (e.g., Maurer et al., 2002) and remotely sensed biophysical parameters (e.g., Nguy-Robertson et al., 2014), this approach could potentially be applicable on a daily basis at regional/global scales. In this study, our objectives are to (1) enhance the light use efficiency model estimation of GPP on a daily and seasonal basis utilizing four scalars for light quality, temperature, water stress, and phenology for in-situ data collected from both maize and soybean at three Nebraskan sites between 2001 and 2008 and (2) evaluate these models from crop data collected at these sites between 2009 and 2012 on a daily, seasonal, and multi-year basis.

2. Materials and methods

2.1. Study site summary

The study area included three fields located at the University of Nebraska-Lincoln (UNL) Agricultural Research and Development Center (ARDC) near Mead, Nebraska, U.S.A. The three sites belong to the AmeriFlux Network, which is sponsored by the U.S. Department of Energy, monitoring carbon fluxes across the North and South American continents. US-Ne1 (41.165°N, 96.476°W, 361 m; 49 ha) and US-Ne2 (41.164°N, 96.471°W, 362 m; 52 ha) were equipped with a center pivot irrigation system while US-Ne3 (41.179°N, 96.439°W, 363 m; 65 ha) was rainfed. In 2001, US-Ne1 and US-Ne2 were equipped with a center pivot irrigation system while US-Ne3 was rainfed. In 2001, a biomass removal study was initiated in the fall of each year starting in 2005. In 2010, a biomass removal study was initiated where the management of US-Ne2 was changed to match US-Ne1 (continuous maize with tillage operations in the fall) except for one factor. Stover was baled and removed from US-Ne2 prior to tillage in order to study the impact of residue removal on carbon and water fluxes. All fields have been fertilized and treated with herbicide and pesticides following best management practices for Eastern Nebraska. For maize, in the irrigated fields, approximately 180 kg N ha$^{-1}$ was applied each year. This was conducted in three applications using the center pivot. Approximately two-thirds (120 kg N ha$^{-1}$) was applied pre-planting and the remaining (60 kg N ha$^{-1}$) was applied in two fertigations. Only a single pre-plant N fertilizer application of 120 kg N ha$^{-1}$ was made on the rainfed site during maize years. Table 1 summarizes the three study sites from 2001 to 2012 (e.g., yield, planting, emergence, and harvest dates).

2.2. Flux measurements

The eddy covariance flux measurements of CO$_2$ (F$_C$), latent heat (LE), sensible heat (H), and momentum fluxes were collected using a Gill Sonic anemometer (Model R3; Gill Instruments Ltd., Lymington, UK), a closed- and open-path CO$_2$/H$_2$O water vapor sensor (LI-6626 and LI-7500, respectively; LI-Cor Lincoln, NE). Storage of CO$_2$ below the eddy covariance sensors was determined from profile measurements of CO$_2$ concentration (LI-6626) and combined with F$_C$ to determine net ecosystem productivity (NEP). Processing methods for correcting flux data due to coordinate rotation (e.g., Baldocchi et al., 1988), inadequate sensor frequency response (e.g., Massman, 1991), and variation in air density (Webb et al., 1980) were applied to all data sets. Key supporting meteorological variables measured included soil heat flux, humidity, incident solar radiation, in situ air and soil temperature, wind speed, and incident photosynthetically active radiation (PAR$_{an}$). Missing data due to sensor malfunction, unfavorable weather, power outages, etc., were gap-filled using a method that combined measurements, interpolation, and empirical data (Baldocchi et al., 1997; Kim et al., 1992; Suyker et al., 2003; Wofsy et al., 1993). Problems associated with insufficient turbulent mixing during nighttime hours was also corrected (Barford et al., 2001; Suyker and Verma, 2012). When mean wind speed (U) was below the threshold value (U = 2.5 ms$^{-1}$ corresponding to a friction velocity of approximately 0.25 ms$^{-1}$), data were filled in using night CO$_2$ exchange-temperature relationships from windier conditions. The daytime estimates of ecosystem respiration (Re) were determined from the temperature-adjusted nighttime CO$_2$ exchange (Xu and Baldocchi, 2004). The GPP was obtained from the difference between NEP and Re (sign convention: GPP and NEP are positive during C uptake by the vegetation and Re is negative).

Energy budget closure is a known issue with regards to eddy covariance measurements and is due, in part, to errors associated with the angle of attack (Frank et al., 2013; Nakai et al., 2006) and phase shifts when estimating energy storage terms (Leuning et al., 2012). For this study, the energy budget closure was determined by comparing the sum of latent and sensible heat fluxes (LE + H) measured by eddy covariance methods with the sum of net radiation and energy storage (R$_{net}$ + G). The growing season energy budget closures for all three sites from 2001 to 2012 (0.78–0.97) were reasonable considering the errors inherent in the measurements of these terms.

2.3. Other supporting measurements

Destructive leaf area measurements were collected from six small (20×20 cm) plots (i.e., intensive measurement zones or IMZs). The IMZs represent all major soil types of each site, including Tomek, Yutan, Filbert, and Fillmore soil series (Suyker et al., 2004). The green LAI, or photosynthetically active leaf area index, was calculated from a 1m sampling length from one or two rows (6×2 plants) within each IMZ. Samples were collected from each field every 10–14 days starting at the initial growth stages (Abendroth et al., 2011), and ending at crop maturity. To minimize edge effects, collection rows were alternated between sampling dates.
The plants collected were transported on ice to the laboratory where they were visually divided into green leaves, dead leaves, stems, and reproductive organs. The leaf area was measured using an area meter (Model LI-3100, LI-Cor Lincoln, NE). The leaf area index for maize and soybean was determined from use of an extinction coefficient \((\epsilon)\) for each crop. To minimize noise and errors, the average value of \((\epsilon)\) was scaled using all four scalars, light quality or amount of diffuse light \((k^D)\) for each crop was determined different ways: using differences in sunlight vs. shaded leaves (He et al., 2013), temperature and light (McCallum et al., 2013), remote sensing models (Pei et al., 2013), etc. The Vegetation Photosynthesis Model (VPM; Xiao et al., 2004), which was originally developed for satellite imagery, scales \(\epsilon\) using temperature \(T^\text{scalar}\), water stress \(W^\text{scalar}\), and phenology \(P^\text{scalar}\):

\[
\epsilon = \epsilon_0 \times T^\text{scalar} \times W^\text{scalar} \times P^\text{scalar}
\]

where \(\epsilon_0\) is maximum light use efficiency. Suyker and Verma (2012) scaled light use efficiency based on a light quality or amount of diffuse light \(C^\text{scalar}\):

\[
\epsilon = \epsilon_0 \times C^\text{scalar}
\]

where \(\epsilon_0\) is now defined as “clear sky” maximum light use efficiency. In this study, \(\epsilon\) was scaled using all four scalars, light quality, temperature, water stress, and phenology:

\[
\epsilon = \epsilon_0 \times C^\text{scalar} \times T^\text{scalar} \times W^\text{scalar} \times P^\text{scalar}
\]
Thus, daily GPP can be estimated using a cloud-adjusted light use efficiency model (LUEn):

\[ \text{GPP} = \varepsilon_0 \times C_{\text{scalar}} \times T_{\text{scalar}} \times W_{\text{scalar}} \times P_{\text{scalar}} \times \text{APAR} \]  
(5b)

The \( C_{\text{scalar}} \) takes into account improved efficiency of canopy photosynthesis in diffuse compared to direct light. Therefore, \( C_{\text{scalar}} \) scales above 1 using the following equation (Suyker and Verma, 2012):

\[ C_{\text{scalar}} = 1 + \beta \times \left( \frac{\text{PAR}_{d}}{\text{PAR}_{n}} - 0.17 \right) \]  
(6)

where \( \beta \) is the sensitivity of \( \varepsilon \) to diffuse light and \( \text{PAR}_{d}/\text{PAR}_{n} = 0.17 \) on a clear day. However, at many research sites, \( \text{PAR}_{d} \) data are not collected. To incorporate the effect of diffuse light in this \( \varepsilon \) model, \( \text{PAR}_{d}/\text{PAR}_{n} \) was related to the cloudiness coefficient (CC):

\[ CC = 1 - \frac{\text{PAR}_{n}}{\text{PAR}_{\text{pot}}} \]  
(7a)

where \( \text{PAR}_{\text{pot}} \) is the estimated total amount of daily incident PAR assuming cloud-free conditions based on factors, such as latitude, elevation, atmospheric pressure, etc. (Weiss and Norman, 1985). We note corrected equations (A. Weiss, personal communication) for hourly \( \text{PAR}_{\text{pot}} \) as the sum of direct and diffuse \( \text{PAR} \) (\( \text{R}_{\text{DV}} \) and \( \text{R}_{\text{DV}} \) respectively):

\[ \text{PAR}_{\text{pot}} = \text{R}_{\text{DV}} + \text{R}_{\text{DV}} \]  
(7b)

\[ \text{R}_{\text{DV}} = 2428 \times \cos \theta \times \exp\left( -0.185 \times P \right) \]  
(7c)

\[ \text{R}_{\text{DV}} = 0.4 \times (2428 \times \cos \theta - \text{R}_{\text{DV}}) \]  
(7d)

where \( \theta \) is solar zenith angle (midpoint of each hour), \( P \) is site atmospheric pressure (kPa), and \( \text{PAR} \) incident at the top of the atmosphere is 2428 μmol m\(^{-2}\) s\(^{-1}\) (a value of 2760 was used in the original paper). Hourly values of \( \text{PAR}_{\text{pot}} \) were calculated and integrated over each day.

The \( T_{\text{scalar}} \) has been modeled based on the Terrestrial Ecosystem Model (Raich et al., 1991):

\[ T_{\text{scalar}} = \frac{(T - T_{\text{min}}) \times (T - T_{\text{max}})}{(T - T_{\text{min}}) \times (T - T_{\text{max}}) - (T - T_{\text{opt}})^2} \]  
(8)

where \( T \) is daytime average air temperature (when \( \text{PAR}>1 \) μmol m\(^{-2}\) s\(^{-1}\)) and the parameters for \( T_{\text{min}}, T_{\text{max}} \) and \( T_{\text{opt}} \) were 10, 48, and 28 °C, respectively, based on Kalfas et al. (2011). While these temperature parameters could be more narrowly adapted to crop species (i.e., maize or soybean) or regions (i.e., eastern Nebraska), this broad temperature range should reduce the risk of the model becoming specific to a particular plant functional type (\( C_{\text{v}} \) vs. \( C_{\text{c}} \)), growth stage, and/or region.

The \( W_{\text{scalar}} \) takes into account the complex impact of water stress on photosynthesis (i.e., changes in stomatal conductance, leaf water potential, etc.) caused by soil moisture and/or atmospheric water deficits. The \( W_{\text{scalar}} \) is determined using one of multiple techniques from remote sensing data (Wuet al., 2008) or meteorological variables (Maselli et al., 2009; Moreno et al., 2014). Vapor pressure deficit (VPD) is known to affect \( \text{GPP} \) over multiple techniques from remote sensing data (Wuet al., 2008) and \( \text{VPD} \) decreases in the presence of a soil moisture deficit (Hirasawa and Hsiao, 1999). The VPD is already used as a constraint for stomatal conductance in evapotranspiration models. For example, specific biomes are assign values of VPD, along with temperature, for when the stomata are expected to be fully open or closed and these values are applied to the model using look-up tables (Mu et al., 2011, 2007). A similar approach, using one set of VPD values for all crops, was adapted for \( \varepsilon \) models (Yuan et al., 2010). For our study, we modified an approach estimating the plant photosynthetic response to VPD based on varying convexity (Gilmanov et al., 2013). This approach has originally been used in examining changes where the scalar will remain stable (e.g., at 1) until VPD reaches a critical threshold (generally accepted near 1 kPa) that induces a reduction in photosynthesis (El-Sharkawy et al., 1984; Lasslo et al., 2010). However, for this study we seek to determine a scalar useful for daily averages of VPD. Since a daily average of VPD below 1 kPa could contain periods where VPD was greater than 1 kPa, no critical threshold was utilized resulting in the following equation:

\[ W_{\text{scalar}} = \exp\left( -\left[ \frac{\text{VPD}}{\sigma_{W_{\text{scalar}}}} \right]^2 \right) \]  
(9)

where the \( \sigma_{W_{\text{scalar}}} \) is the curvature parameter for water stress.

The \( P_{\text{scalar}} \) determined using remote sensing techniques, accounts for the impact of phenology/leaf age at the canopy level (Kalfas et al., 2011; Wang et al., 2012). Immature leaves do not have the same capacity as mature leaves to photosynthesize (Reich et al., 1991) and mature leaves lose their photosynthetic capacity as they senesce (Dwyer and Stewart, 1986; Field and Mooney, 1983). Green LAI is a good indicator of canopy-level phenological changes in maize and soybean increasing during leaf expansion (vegetative growth stages) and decreasing as canopy chlorophyll is degraded (reproductive growth stages/senescence; Nguy-Robertson et al., 2012). For our study, the equation was adjusted such that the \( P_{\text{scalar}} \) was one at peak green LAI:

\[ P_{\text{scalar}} = \exp\left( -\left[ \frac{\text{green LAI}_{\text{max}} - \text{green LAI}}{\sigma_{P_{\text{scalar}}}} \right]^2 \right) \]  
(10)

where the \( \sigma_{P_{\text{scalar}}} \) is the curvature parameter for phenology and green LAI\(_{\text{max}}\) is the maximal green LAI for each rainfed and irrigated crop. Green LAI\(_{\text{max}}\) is a potential maximum leaf area for a particular crop management (e.g., irrigation, planting density). Other factors (e.g., extreme weather, plant pests/disease) can affect leaf area distribution and peak values in a particular year. These impacts on \( P_{\text{scalar}} \) are discussed in Section 3.1.

2.5 Statistical methods

The four LUEn parameters \( \varepsilon_0, \beta, \sigma_{W_{\text{scalar}}}, \) and \( \sigma_{P_{\text{scalar}}} \) were determined using a step-wise iterative, or “model tuning” approach (Dall’Olmo et al., 2003; Gitelson et al., 2006). While all four parameters could be determined by simultaneous iteration, it would be computationally intensive. Therefore, predetermined ranges of each parameter were established (maize: \( \varepsilon_0: 0.426–1.0 \), gCmol\(^{-1}\), \( \sigma_{W_{\text{scalar}}}: 3–50 \) kPa, and \( \sigma_{P_{\text{scalar}}}: 6–50 \) m\(^2\)m\(^{-2}\)s\(^{-1}\); soybean: \( \varepsilon_0: 0.298–1.0 \) gCmol\(^{-1}\), \( \sigma_{W_{\text{scalar}}}: 3–50 \) kPa, and \( \sigma_{P_{\text{scalar}}}: 6–50 \) m\(^2\)m\(^{-2}\)s\(^{-1}\)) following a \( k \)-fold cross-validation procedure (Kohavi, 1995) where \( k \) was the number of field-years for each crop between 2001 and 2008: 16 for maize and 8 for soybean.

The step-wise process consisted of eight iterations. The first step was to estimate \( \varepsilon_0 \) using the data when \( C_{\text{scalar}}, W_{\text{scalar}}, \) and \( P_{\text{scalar}} \) are assumed to be close to 1. Thus, \( \varepsilon_0 \) was determined during sunny conditions (CC < 0.2) with low water stress (VPD < 1.0) and a relatively mature canopy (LAI > 2m\(^2\)m\(^{-2}\)). After quantifying \( \varepsilon_0, \beta \) was determined by using an expanded data set disregarding the limitation using the CC. Likewise, \( \sigma_{W_{\text{scalar}}} \) was determined with all VPD values included. The fourth iteration isolated \( \sigma_{P_{\text{scalar}}} \) using the entire data set. To ensure relative stability, the four iterations were repeated using the entire data set and the parameters identified in the first four steps. In order to make an accurate comparison between the approach in this study and the approach presented in Suyker and Verma (2012), the Suyker and Verma (2012) model utilized the original coefficients (i.e. \( k, \varepsilon_0 \), etc.) rather than the updated values (Table 2).
The optimal parameters were selected based on a minimum sum of absolute error (MSAE) regression (André et al., 2003; Narula et al., 1999) using R (V. 3.0.1, 2013, The R Foundation for Statistical Computing). MSAE regression has been found to be advantageous when there are outliers in the data set and the median is a more efficient estimator of the parameter rather than the mean (Narula et al., 1999). Due to differences between fields and various climatic conditions, the annual sum of GPP at a given site can be drastically different from normal years. This difference then impacts the mean value of the annual sum of GPP (maize: median = 1669 gCm⁻², average = 1641 gCm⁻²; soybean: median = 916 gCm⁻², average = 944 gCm⁻²). The sum of absolute error (SAE) by field-year (SAEfield-year) reduces both error and bias because self-correcting errors in the annual (i.e., field-year) sums were penalized. Thus, this approach minimizes the absolute value of the annual difference between modeled and measured GPP for a given site:

\[
\text{SAE}_{\text{field-year}} = \sum_{\text{field-year}} | \text{Daily EstimatedGPP} - \Sigma_{\text{daily}} \text{ModeledGPP} | \quad (11)
\]

The approach minimizing SAEfield-year also accentuates annual over daily performance in the model. ASAE analyses for daily values would over-emphasize accuracy for high GPP values. Basic statistical analyses were performed using Excel (V. 2010, Microsoft) where the coefficients of determination (\(R^2\)) were calculated from the best-fit lines and the mean normalized bias (MNB), and root mean square (RMSE) were calculated from the 1:1 line.

When incorporating a new factor into the VPM (\(C_{\text{scalar}}\)) and modifying other scalars (\(T_{\text{scalar}}, W_{\text{scalar}}\) and \(P_{\text{scalar}}\)), their statistical significance must be evaluated in explaining the variability in daily GPP. Since LUE, non-linear, the model was transformed logistically to perform two separate model selection analyses, Akaike information criterion (AIC) and likelihood ratio test, in R (V. 3.0.1, 2013, The R Foundation for Statistical Computing). To determine if each scalar statistically improves the model we used the following process. From the base model (GPP = \(\epsilon_o \times \text{APAR}\), the AIC was used to determine which singular scalar improved the model the most. The model with the lowest AIC values among the tested models will have the optimal number of parameters for explaining the data while minimizing complexity (Akaike, 1974; Held and Sabanès Bové, 2014). The likelihood ratio test identified if the model was significantly improved. The likelihood ratio test compares a simple model with a nested and more complex model to provide a measure of statistical significance to any improvement of the model by adding a parameter (Fischer, 1921; Held and Sabanès Bové, 2014). The optimal parameter at each level of complexity (i.e., number of scalars), determined from AIC, was used as the simpler model in the likelihood ratio test for the increasingly complex model up to the proposed cloud-adjusted light use efficiency model (LUE).
studies we determined $\beta$ to be $0.347 \pm 0.051$ and $0.411 \pm 0.056$ for maize and soybean, respectively. This discrepancy was likely due to differences in model calibration. The original determination of $\beta$ was from a single site in a single year for each crop. This study determined $\beta$ using the entire calibration data set (24 field-years). The $\sigma_W$ scalar was determined to be $6 \pm 0$ and $4 \pm 0$ kPa for maize and soybean, respectively. The $\sigma_P$ scalar was determined to be $18 \pm 5$ and $18 \pm 7$ m$^2$ m$^{-2}$ for maize and soybean, respectively. The wide range in the variation using the $k$-fold cross-validation technique may be due to fitting the same $\sigma_P$ scalar for both irrigated and rainfed crops despite the different maximal green LAI values. However, other factors not incorporated into the model can also impact green LAI (e.g., disease, damage by pests) and increase the uncertainty in the $\sigma_P$ scalar. The resulting range of values for the scalars and other parameters are shown in Table 4. While the average for each scalar was close to one (0.9–1.1), on particular days the impact of some individual scalars was substantial. The temperature severely reduced $\epsilon$ on some days for both maize and soybean ($T_{\text{scalar}} = 0.02–0.05$) which occurred towards the end of the season when daily daytime temperature averages reached the minimum of 10 °C necessary for physiological activity. The lowest values for the $W_{\text{scalar}}$ was in the rainfed soybean (0.46 when VPD was high (>3 kPa)). However, this was relatively infrequent for all three sites ($n = 36$ days). The relatively small range of $P_{\text{scalar}}$ (~0.7–1.0) was expected as young leaves and canopies can photosynthesize, even if they are inefficient compared to fully mature leaves. This narrow range and the uncertainty in quantifying green LAI during later reproductive stages (Gitelson et al., 2014; Peng et al., 2011) may have contributed to the wider confidence intervals associated with the curvature parameter, $\sigma_P$. Despite multiple factors that reduce maximal green LAI for maize and soybean for their respective management, the $P_{\text{scalar}}$ approached one each field-year (>0.985). The $C$ scalar increased to a maximum of 1.4 in both crops, supporting earlier studies demonstrating that cloudy conditions increase $\epsilon$ (e.g., Knohl and Baldocchi, 2008).

### 3.2. Model selection analysis, calibration, and validation

The LUE was developed using the 2001–2008 data. The likelihood ratio test demonstrated that each successive scalar, while adding complexity to the basic model, significantly improved the

---

**Parameter Utilized for Scalar**

Figure 2. The relationships between the parameters utilized for the scalars; cloudiness coefficient (CC), average daytime temperature ($T$), vapor pressure deficit (VPD), and green leaf area index (green LAI); and the scalars; $C_{\text{scalar}}$, $T_{\text{scalar}}$, $W_{\text{scalar}}$, $P_{\text{scalar}}$. Summary statistics for each parameter and scalar are in Table 4.

### Table 3. Maximal light use efficiency ($\epsilon_o$) values in units of g Cmol$^{-1}$ determined by various studies. For Prince and Goward (1995), the $\epsilon_o$ is adjusted by a temperature factor (α).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Maize</th>
<th>Soybean</th>
<th>Developed specifically for maize or soybean?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running et al.</td>
<td>2004</td>
<td>0.148</td>
<td>0.148</td>
<td>No</td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>2014</td>
<td>0.915</td>
<td>0.567</td>
<td>Yes</td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>2014</td>
<td>1.207</td>
<td>0.612</td>
<td>Yes</td>
</tr>
<tr>
<td>He et al.</td>
<td>2013</td>
<td>0.631</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Kalfas et al.</td>
<td>2011</td>
<td>1.500</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Lobell et al.</td>
<td>2002</td>
<td>0.4–0.8</td>
<td>0.4–0.8</td>
<td>No</td>
</tr>
<tr>
<td>Mahadevan et al.</td>
<td>2008</td>
<td>0.900</td>
<td>0.768</td>
<td>Yes</td>
</tr>
<tr>
<td>Norman and Arkebauer</td>
<td>1991</td>
<td>0.457–0.486</td>
<td>0.356–0.379</td>
<td>Yes</td>
</tr>
<tr>
<td>Prince and Goward</td>
<td>1995</td>
<td>0.600</td>
<td>12α</td>
<td>No</td>
</tr>
<tr>
<td>Suyker and Verma</td>
<td>2012</td>
<td>0.426</td>
<td>0.298</td>
<td>Yes</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2010</td>
<td>0.560</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2012</td>
<td>0.578</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Yan et al.</td>
<td>2009</td>
<td>0.920</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>This study</td>
<td>2015</td>
<td>0.526</td>
<td>0.374</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Maize</th>
<th>Soybean</th>
<th>Developed specifically for maize or soybean?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running et al.</td>
<td>2004</td>
<td>0.148</td>
<td>0.148</td>
<td>No</td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>2014</td>
<td>0.915</td>
<td>0.567</td>
<td>Yes</td>
</tr>
<tr>
<td>Cheng et al.</td>
<td>2014</td>
<td>1.207</td>
<td>0.612</td>
<td>Yes</td>
</tr>
<tr>
<td>He et al.</td>
<td>2013</td>
<td>0.631</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Kalfas et al.</td>
<td>2011</td>
<td>1.500</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Lobell et al.</td>
<td>2002</td>
<td>0.4–0.8</td>
<td>0.4–0.8</td>
<td>No</td>
</tr>
<tr>
<td>Mahadevan et al.</td>
<td>2008</td>
<td>0.900</td>
<td>0.768</td>
<td>Yes</td>
</tr>
<tr>
<td>Norman and Arkebauer</td>
<td>1991</td>
<td>0.457–0.486</td>
<td>0.356–0.379</td>
<td>Yes</td>
</tr>
<tr>
<td>Prince and Goward</td>
<td>1995</td>
<td>0.600</td>
<td>12α</td>
<td>No</td>
</tr>
<tr>
<td>Suyker and Verma</td>
<td>2012</td>
<td>0.426</td>
<td>0.298</td>
<td>Yes</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2010</td>
<td>0.560</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Wang et al.</td>
<td>2012</td>
<td>0.578</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Yan et al.</td>
<td>2009</td>
<td>0.920</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>This study</td>
<td>2015</td>
<td>0.526</td>
<td>0.374</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 4. Summary of the parameters and corresponding equation number (Eq.) utilized in this study. The minimum (min), maximum (max), and average (avg) of each parameter was presented for each crop. Numbers in square brackets indicate values for the rainfed site (US-Ne3) while those to the left were for the two irrigated sites (US-Ne1 and US-Ne2).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Symbol</th>
<th>Eqs.</th>
<th>Units</th>
<th>Maize</th>
<th>Soybean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross primary production</td>
<td>GPP</td>
<td>(1)</td>
<td>gC m⁻² d⁻¹</td>
<td>0.0</td>
<td>33.5[29.5] 13.5[12.0]</td>
</tr>
<tr>
<td>Green leaf area index</td>
<td>green LAI</td>
<td>(2), (10)</td>
<td>m² m⁻²</td>
<td>0.0</td>
<td>6.78[4.93] 3.26[2.35]</td>
</tr>
<tr>
<td>Absorbed PAR by green</td>
<td>APAR</td>
<td>(2)</td>
<td>Mol photons m⁻² d⁻¹</td>
<td>0.0</td>
<td>60.5[54.4] 28.4[24.7]</td>
</tr>
<tr>
<td>components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incident PAR</td>
<td>PARin</td>
<td>(2)</td>
<td>Mol photons m⁻² d⁻¹</td>
<td>1.0(1.4)</td>
<td>65.1[64.9] 30.9[31.0]</td>
</tr>
<tr>
<td>Ratio of diffuse PAR and PAR</td>
<td>PARD/PARin</td>
<td>(6), (12)</td>
<td>Unitless</td>
<td>0.0</td>
<td>0.90[0.89] 0.25</td>
</tr>
<tr>
<td>Cloudiness coefficient</td>
<td>CC</td>
<td>(7), (12), (13)</td>
<td>Unitless</td>
<td>0.0</td>
<td>1.14[1.08] 0.48[0.49]</td>
</tr>
<tr>
<td>Potential PARin</td>
<td>PARpot</td>
<td>(7)</td>
<td>Mol photons m⁻² d⁻¹</td>
<td>27.6</td>
<td>65.5      54.2</td>
</tr>
<tr>
<td>Vapor pressure deficit</td>
<td>VPD</td>
<td>(9)</td>
<td>kPa</td>
<td>0.0[0.03]</td>
<td>3.52[3.70] 1.22[1.32]</td>
</tr>
<tr>
<td>Cloudiness scalar</td>
<td>Cscalar</td>
<td>(6), (13)</td>
<td>Unitless</td>
<td>1.0(1.02)</td>
<td>1.35      1.11</td>
</tr>
<tr>
<td>Temperature scalar</td>
<td>Tscalar</td>
<td>(8)</td>
<td>Unitless</td>
<td>0.04</td>
<td>1.0       0.92</td>
</tr>
<tr>
<td>Water stress scalar</td>
<td>Wscalar</td>
<td>(9)</td>
<td>Unitless</td>
<td>0.71[0.68]</td>
<td>1.0       0.95</td>
</tr>
<tr>
<td>Phenology scalar</td>
<td>Pscalar</td>
<td>(10)</td>
<td>Unitless</td>
<td>0.67[0.93]</td>
<td>1.0       0.95[0.97]</td>
</tr>
</tbody>
</table>

Table 5. Summary of model selection results for the Akaike Information Criterion (AIC) and likelihood ratio test. The difference between the AIC and minimum Akaike Information Criterion (AICmin) was shown to make it easier to identify optimal models at each level of complexity. The optimal parameter at each level of complexity (in bold) was used as the simpler model in the likelihood ratio test for the increasingly complex model up to the proposed cloud-adjusted light use efficiency model (LUE). These results indicate that the addition of each remaining parameter was statistically significant (p-value < 0.001).

<table>
<thead>
<tr>
<th>Model</th>
<th>Akaike information criterion</th>
<th>Likelihood ratio test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>AIC-AICmin</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>df</td>
</tr>
<tr>
<td>APAR × εₒ</td>
<td>7065</td>
<td>4563</td>
</tr>
<tr>
<td>APAR × εₒ × Cscalar</td>
<td>2694</td>
<td>191</td>
</tr>
<tr>
<td>APAR × εₒ × Tscalar</td>
<td>2735</td>
<td>233</td>
</tr>
<tr>
<td>APAR × εₒ × Wscalar</td>
<td>2740</td>
<td>238</td>
</tr>
<tr>
<td>APAR × εₒ × Pscalar × Cscalar</td>
<td>2676</td>
<td>174</td>
</tr>
<tr>
<td>APAR × εₒ × Pscalar × Tscalar × Cscalar</td>
<td>2598</td>
<td>96</td>
</tr>
<tr>
<td>APAR × εₒ × Pscalar × Wscalar</td>
<td>2652</td>
<td>150</td>
</tr>
<tr>
<td>APAR × εₒ × Pscalar × Tscalar × Wscalar</td>
<td>2528</td>
<td>25</td>
</tr>
<tr>
<td>LUEC</td>
<td>2503</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. The estimated and measured gross primary production (GPP) relationships from the 2001–2008 calibration data for the two light use efficiency models: (A) cloud-adjusted (LUEC) and (B) Suyker and Verma (2012) model. The coefficient of determination (R²) was determined from the best-fit line for maize and soybean data combined. The mean normalized bias (MNBE) and root mean square error (RMSE) was determined from the 1:1 line.

estimation of daily GPP (p-value = 0.0002, df = 8; Table 5). The largest decrease in AIC occurred when adding any one of the scalars and the Pscalar contributed the most to the variability in GPP for these maize and soybean crops. The model estimated GPP with reasonable accuracy and low bias (RMSE: 2.2 gC m⁻² d⁻¹; MNBE: 4.7%; Figure 3A). Minimizing bias has two benefits. Firstly, error due to bias will compound over time and thus reduce the accuracy in monitoring long-term trends in GPP. Secondly, lower bias indicates that over-and/or under-estimation of GPP was minimized for specific periods of the growing season (i.e., early, peak, etc.). The daily trends of the measured and modeled GPP between 2001 and 2008 roughly matched for US-Ne1 (Figure 4), US-Ne2 (Figure 5), and US-Ne3 (Figure 6). This indicates that the model was reasonably estimating both low and high values of GPP.

The model was tested using the 2009–2012 data by evaluating daily and yearly RMSE and bias. While there was slightly
Modeling gross primary production of maize and soybean croplands

Figure 4. Growing season distributions of the measured daily gross primary production (GPP) and the estimated GPP from the cloud-adjusted light use efficiency model (LUEc) at the AmeriFlux site US-Ne1 located near Mead, NE, USA from 2001 to 2012. The site was managed as irrigated continuous maize during the entire study.

Figure 5. Growing season distributions of the measured daily gross primary production (GPP) and the estimated GPP from the cloud-adjusted light use efficiency model (LUEc) at the AmeriFlux site US-Ne2 located near Mead, NE, USA from 2001 to 2012. The site was irrigated and managed as a maize (odd years) and soybean (even years) rotation from 2001 to 2009. From 2010 to 2012 the site was managed as continuous maize.
increased scatter in the daily modeled vs. measured GPP relationships (RMSE = 2.6 gCm$^{-2}$ d$^{-1}$), this error was still reasonable (Figure 7A). The temporal behavior of the modeled and measured GPP for 2009–2012 was similar to those in 2001–2008 (Figs. 4–6). Yearly estimates of GPP (RMSE = 27.4 gCm$^{-2}$ y$^{-1}$) were also reasonable (Figure 7C). Desai et al. (2008) found the errors associated with the method of measuring GPP and gap-filling to be less than 10% across several methods in various biomes. For LUE$^c$, all the data points in the validation data set fell within this 10% error threshold from measured GPP except for US-Ne3 in 2010 (13.5%) and 2012 (−13.5%).

The accuracy of the LUE$^c$ over the period of validation (2009–2012) was strikingly good even with a change in management for US-Ne2 (from maize/soybean rotation to continuous maize) to accommodate a biomass study and several unforeseen events that influenced crop growth and the carbon flux. For example, at the end of the 2010 growing season there was a hail storm that damaged all three sites, but impacted US-Ne1 the most with an estimated loss of grain carbon of over 400 gCm$^{-2}$ (stalks were lodged by large hail). This grain was incorporated in the field following fall conservation tillage to decompose the following growing seasons, yet this additional respiration did not impact GPP estimates for LUE$^c$ (US-Ne1 2011: RMSE= 2.4 gCm$^{-2}$ d$^{-1}$). Another unexpected event was the drought in 2012. While the LUE$^c$ performed worse in 2012 compared to other years in several metrics (2012: RMSE= 3.4 gCm$^{-2}$ d$^{-1}$; MNB= 13.5%), the model still had less error and bias than the Suyker and Verma (2012) model (2012: RMSE=3.9 gCm$^{-2}$ d$^{-1}$; MNB= 30.0%). This indicates that the LUE$^c$ was fairly robust even during extreme events, likely due to using VPD as a metric for estimating the $W_{\text{scalar}}$.

In addition to evaluating the LUE$^c$ and the significance of each parameter scaling $\epsilon$, we also wanted to quantify the improvement in this model compared to Suyker and Verma (2012). The Suyker and Verma (2012) modeled values underestimated daily GPP compared to measured values for the developmental period (slope = 0.885 from 2001 to 2008; Figure 3B) and the test period (slope = 0.839 from 2009 to 2012; Figure 7B). Growing season totals show larger RMSE, too (Figure 7D). Generally for all metrics utilized in this study (i.e., error, bias), the approach incorporating four scalars outperformed the single scalar based model. This suggests multiple factors are significantly impacting light use efficiency that ultimately affects daily and seasonal estimates of GPP.

3.3. Long-term error accumulation and bias associated with the models

While the daily accuracy of the model is important, small biases in modeled GPP can accumulate over multiple years. There are two types of cumulative error that reflect the quality of the model: (1) error that is self-correcting where over-estimations in some years can be offset by under-estimations in subsequent years which reduces bias (sum of error; SOE) and (2) error that accumulates the absolute difference between modeled and measured GPP each year (sum of absolute error; SAE). For the LUE$^c$ from 2009 to 2012 for all three sites under differing management practices (e.g., rainfed vs. irrigated, continuous maize vs. maize/soybean rotation), the magnitude of SOE (US-Ne1: −33.7; US-Ne2: 272.7; US-Ne3: −231.4 gCm$^{-2}$) was within ±5% of measured cumulative GPP. The values of SAE ranged from 2 to 9% of GPP (US-Ne1: 157.0; US-Ne2: 398.5; US-Ne3: 441.2 gCm$^{-2}$). The cumulative error and bias of LUE$^c$ were within reason when compared to other light use efficiency models. For example, a direct comparison across the three sites, the SOE and SAE from the Suyker and Verma (2012) model ranged from −2 to 4% and 3 to 13%, respectively. The LUE$^c$ demonstrates that it reduces self-correction compared to the earlier approach by Suyker and Verma (2012). Using the VPM between
Figure 7. The (A–B) daily and (C–D) yearly estimated vs. measured gross primary production (GPP) relationships from the 2009–2012 validation data set for the two light use efficiency models, (A, C) cloud-adjusted (LUE\textsubscript{c}) and (B, D) Suyker and Verma (2012) model. The coefficient of determination ($R^2$) was determined from the best-fit line for both maize and soybean. The mean normalized bias (MNB) and root mean square error (RMSE) was determined from the 1:1 line. Ten percent error bars (dashed lines) are included in the yearly estimated GPP graphs.

Figure 8. Cumulative annual sum of error (SOE) between measured and estimated gross primary production (GPP) from 2001 to 2012 for (A) the cloud adjusted light use efficiency model (LUE\textsubscript{c}) and (B) the Suyker and Verma (2012) model and cumulative annual sum of absolute error (SAE) for (C) LUE\textsubscript{c} and (D) Suyker and Verma (2012) model.
2001 and 2005, Xiao et al., (2014) over-estimated GPP in each year for US-Ne2 for a total of 458 gCm\(^{-2}\) (SOE = SAE = 7%).

While the long-term analysis here is limited to four years, we repeated the analysis with data from 2001 to 2012 (Figure 8A and C). Inclusion of the calibration data into this error analysis may not be ideal; however, it does provide some additional insights to the long-term trends. The SOE was −0.5 to 2% and SAE was 3 to 7% for all three sites where cumulative GPP measured 14,000 to 20,000 gCm\(^{-2}\). The corresponding SOE and SAE for Suyker and Verma (2012) was −1 to 2% and 4 to 10%, respectively (Figure 8B and D). From 2001 to 2005 at US-Ne2, the SOE and SAE were lower (0.7 and 2%, respectively) compared to Xiao et al., (2014). This error analysis suggests incorporating multiple scaling factors (regulated by meteorological and biophysical variables) into light use efficiency models can provide long-term GPP estimates with small bias.

4. Conclusion

The cloud-adjusted light use efficiency model (LUE\(_c\)) was able to model GPP utilizing field-based meteorological and biophysical measurements from three Nebraska AmeriFlux sites growing two different crops, maize and soybean, from 2001 to 2012. This light use efficiency (\(\epsilon\)) model incorporated four scalars for estimating GPP, including light, climate, impacts of temperature, water stress, and phenology. The model coefficients for LUE\(_c\) were calibrated using a k-fold cross-validation procedure using data collected between 2001 and 2008. A computationally efficient iterative procedure ascertained initial parameter estimates from a limited range of environmental conditions and final parameters were determined from the entire data set. The likelihood ratio test demonstrated that all four scalars were statistically significant in improving the model estimation of daily GPP. On a day to day basis, temperature scalar can range from zero to one while the phenology scalar has the smallest range (0.7–1). However, based on the Akaike Information Criterion analysis, phenology explained more GPP variability compared to temperature and the other two scalars. This model was validated on data collected between 2009 and 2012. The LUE\(_c\) had low error and bias estimates for daily and growing season GPP. On a cumulative basis, the sum of error between the measured and modeled GPP which provides a measure of long-term cumulative bias (2001–2012), was less than 350 gCm\(^{-2}\).

This error analysis suggests incorporating multiple scaling factors (regulated by meteorological and biophysical variables) into light use efficiency models can provide long-term GPP estimates with small bias.

Acknowledgments — The US-Ne1, US-Ne2, and US-Ne3 AmeriFlux sites were supported by the DOE Office of Science (BER; Grant Nos. DE-FG03-00ER62996, DE-FG02-03ER63639, and DE-EE0003149), DOE-EPS-CoR (Grant No. DE-FG02-00ER45827), and NASA NACP (Grant No. NNX08AI75G). We are grateful to be supported by the resources, facilities, and equipment by the Carbon Sequestration Program, the School of Natural Resources, and the Nebraska Agricultural Research Division located within the University of Nebraska-Lincoln. We would like to thank Dr. Tim Arkebauer and Dave Scoby for the destructive leaf area measurements. We gratefully acknowledge the technical assistance of Sheldon Sharp, Ed Cunningham, Brent Riehl, Tom Lowman, Todd Schimelfenig, Dan Hatch, Jim Hines, and Mark Schroeder.

References


Fischer, R.A., 1921. On the probable error of a coefficient of correlation deduced from a small sample. Metron 1, 3–32.


Gitelson, A.A., Peng, Y., Arkebauer, T.J., Schepers, J., 2014. Relationships between gross primary production, green LAI, and canopy chlo-
rophyll content in maize: implications for remote sensing of pri-
rse.2014.01.004


Hall, F.G., Hilker, T., Coops, N.C., 2012. Data assimilation of photosyn-
rse.2012.02.007

Haxeltine, A., Prentice, I.C., 1996. A general model for the light-


Hirasawa, T., Hsiao, T.C., 1999. Some characteristics of reduced leaf pho-


lcates LTD. Montreal, Quebec, Canada, pp. 1137–1143.


Li, A., Bian, J., Lei, G., Huang, C., 2012. Estimating the maximal light use efficiency for different vegetation through the CASA model com-
bined with time-series remote sensing data and ground measure-


