Ecological Indicators 60 (2016) 702–709

Contents lists available at ScienceDirect

Ecological Indicators

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

Estimating crop yield using a satellite-based light use efficiency model

Wenping Yuan[a], Yang Chen[a], Jiangzhou Xia[a], Wenjie Dong[a,*], Vincenzo Magliulo[b], Eddy Moors[c], Jørgen Eivind Olesen[d], Haicheng Zhang[a]

[a] State Key Laboratory of Earth Surface Processes and Resource Ecology, Future Earth Research Institute, Beijing Normal University, Beijing 100875/Zhuhai 519087, China
[b] National Research Council, Institute for Agricultural and Forest Systems, 80040 Ercolano, Napoli, Italy
[c] Wageningen UR, Alterra, Earth System Science and Climate Change Group, PO Box 47, 6700 AA Wageningen, The Netherlands
[d] Department of Agroecology, Aarhus University, DK-8830 Tjele, Denmark

ARTICLE INFO

Article history:
Received 9 March 2015
Received in revised form 1 July 2015
Accepted 6 August 2015

Keywords:
Light use efficiency
EC-LUE
Gross primary production
Crop yield

ABSTRACT

Satellite-based techniques that provide temporally and spatially continuous information over vegetated surfaces have become increasingly important in monitoring the global agriculture yield. In this study, we examine the performance of a light use efficiency model (EC-LUE) for simulating the gross primary production (GPP) and yield of crops. The EC-LUE model can explain on average approximately 90% of the variability in GPP for 36 FLUXNET sites globally. The results indicate that a universal set of parameters, independent of crop species (except for C4 crops), can be adopted in the EC-LUE model for simulating crops' GPP. At both irrigated and rainfed sites, the EC-LUE model exhibits a similar level of performance. However, large errors are found when simulating yield based on crop harvest index. This analysis highlights the need to improve the representation of the harvest index and carbon allocation for improving crop yield estimations from satellite-based methods.

1. Introduction

Approximately 12% of the Earth's land surface is presently represented by cultivated cropland, which supplies the great part of human food production. Sustained world population growth, rising meat and dairy consumption and expanding biofuel use are exerting increasing pressure on global agriculture (Ray et al., 2012). Global food production, however, is and will be significantly affected by climate change (Parry et al., 2004; Schmidhuber and Tubiello, 2007). In this perspective, crop production monitoring and forecasting is of fundamental importance for agricultural management, food security threats, food trade policy and carbon cycle research (Tilman et al., 2011).

Remotely sensed data provide temporally and spatially continuous information over vegetated surfaces and is useful for accurately monitoring cropland yield and spatial patterns. Generally, there are two approaches for yield estimation using remote sensing data. The first method includes biophysical crop-simulation models that retrieve crop growth parameters from remotely sensed data, which are used as inputs to calibrate and drive the models (Brisson et al., 1998). The main drawback of such models is that they typically require numerous crop-specific inputs such as soil characteristics, management practices, agro-meteorological data and planting dates to simulate crop growth and development through the crop cycle (Moriondo et al., 2007). Such crop-simulation models include CERES (Ritchie and Otter, 1985), WOFOST (Van Depen et al., 1989) and CROPSYST (Van Evert and Campbell, 1994). The second scheme includes statistical regression-based methods, which are the most commonly used remote sensing-based approaches (Wall et al., 2008). These are based on empirical relationships between historic yields and reflectance-based vegetation indices. They are typically straightforward to implement and do not require numerous inputs. A main drawback of empirically based approaches is that the relationships between yield and reflectance are typically localized and are not easily extendable to other areas (Doraiswamy et al., 2003; Moriondo et al., 2007).

Satellite-based light use efficiency (LUE) models are an alternative approach that makes it possible to accurately estimate crop yield, because they can successfully estimate the vegetation's gross primary production (GPP), which is at the basis of the ecosystem's carbon biogeochemical cycle and the main variable determining crop yield. The LUE models build upon the assumption that ecosystem GPP is directly related to the absorbed Photosynthetically Active Radiation (APAR) through LUE (Monteith, 1972, 1977). Actual LUE may be reduced below its theoretical potential value by

* Corresponding author. Tel.: +86 13522824428.
E-mail addresses: yuanwpcn@126.com (W. Yuan), dongwj@bnu.edu.cn (W. Dong).

http://dx.doi.org/10.1016/j.ecolind.2015.08.013
1470-160X/

environmental stresses such as low temperatures or water shortages (Landsberg, 1986; Goerner et al., 2009). Some studies have evaluated LUE models at regional and global scales in major ecosystem types (Potter et al., 1993; Turner et al., 2006; Huntzinger et al., 2012; Yuan et al., 2012; Raczka et al., 2013; Cai et al., 2014).

In recent years, several studies have used the LUE models to address the spatial and temporal patterns of cropland vegetation production (Chen et al., 2014) and to simulate crop yield over large areas (Xin et al., 2013). On the other hand, some authors examined the capability of LUE models to estimate crop yield and demonstrated large uncertainties. For example, based on the MODIS-GPP products (MOD17), Reeves et al. (2005) found a weak correlation between GPP estimates and the national agricultural yield data at the county level for the states of Montana and North Dakota in the United States, and highlighted the need to use crop-specific model parameters. Moreover, the model performance for simulating crop yield has never been evaluated at the site scale. Previous studies often examined model performance over large regions by comparing outputs with regional crop inventory data (Xin et al., 2013; Li et al., 2014). The uncertainties from model inputs, which are needed for regional estimates, will hinder the judgment of model performance. Li et al. (2014) found the misclassification of cropland from a land cover product is one of the major causes of bias in the NPP estimates of cropland in the midwestern United States.

This study aims to examine the performance of a LUE model (i.e., EC-LUE, Yuan et al., 2007) for the prediction of crop biomass and yield production, over multiple crop species globally. The overarching goals of this study are to (1) examine the model performance for vegetative primary production across various crop types and management, (2) assess the possibility for simulating crop yield from crop vegetation production estimates, and (3) investigate the most relevant processes for crop yield simulations based on the LUE models.

2. Model and data

2.1. EC-LUE model

Yuan et al. (2007, 2010) developed the Eddy Covariance-Light Use Efficiency (EC-LUE) model to simulate daily vegetation GPP. The EC-LUE model is driven by only four variables: the Normalized Difference Vegetation Index (NDVI), Photosynthetically Active Radiation (PAR), air temperature (T), and the Bowen ratio of sensible to latent heat flux. The EC-LUE model has the great advantage to map daily GPP over large areas because the potential LUE is invariant across various land cover types (Yuan et al., 2014b). Previous study indicated that a universal of set of parameters, which is independent of vegetation cover type and characteristics can be adopted in EC-LUE model (Yuan et al., 2014b). Availability of this well tested and universal set of parameters would help to improve the accuracy and applicability of LUE models in various biomes and geographic regions.

The model equations are as follows:

$$\mathrm{GPP} = \mathrm{PAR} \times \mathrm{fPAR} \times \varepsilon_{\max} \times Min(T_s, WS_{EF}) \quad (1)$$

$$\mathrm{fPAR} = 1.24 \times \mathrm{NDVI} - 0.168 \quad (2)$$

$$T_s = \frac{(T - T_{min}) \times (T - T_{max})}{(T - T_{\min}) \times (T - T_{max}) - (T - T_{opt})^2} \quad (3)$$

$$W_{SEF} = \frac{\mathrm{LE}}{R_n} \quad (4)$$

where fPAR is the fraction of intercepted incident PAR. ε_{max} is the potential light use efficiency without environmental stress (2.14 g C m^{-2} MJ^{-1} APAR). *Min* denotes the minimum values of T_s and W_{SEF} (it is assumed that the impacts of temperature and moisture follow Liebig's Law, so that LUE is only affected by the most limiting factor at any given time). T_{min}, T_{max} and T_{opt} are the minimum, maximum and optimum air temperatures (°C) for photosynthetic activity, respectively. If the air temperature falls below T_{min} or increases beyond T_{max}, T_s is set to zero. In this study, T_{min} and T_{max} were set to 0 and 40 °C (Yuan et al., 2007), respectively, while T_{opt} was determined using nonlinear optimization to be 21 °C (Yuan et al., 2007). LE is the daily latent heat flux (MJ m^{-2}), which is estimated using the revised RS-PM (Remote Sensing – Penman Monteith) model (Yuan et al., 2010). R_n is the daily net radiation (MJ m^{-2}).

2.2. Data

Data collected at 36 eddy covariance (EC) sites (78 site-year) were used in this study to examine the performance of the EC-LUE model (Table 1). These sites covered several dominant cropland ecosystem types (Table 1). EC data were obtained from the websites: FLUXNET (http://www.fluxdata.org), HiWATER (http://westdc.westgis.ac.cn/hiwater) (Li et al., 2013), and AsiaFlux (http://www.asiaflux.net). Supplementary information on the vegetation, climate and soil at each site was also available at the above web sites. Half-hourly or hourly averaged PAR, T and friction velocity (u^*) were used along with Net ecosystem CO_2 exchange (NEE) in this study. FLUXNET datasets that were gap-filled by site investigators were used directly for this study (i.e., the LaThuile database) (Agarwal et al., 2010).

For the sites that were not in the LaThuile FLUXNET database, the following established procedures were used to fill data gaps (Yuan et al., 2014a): nonlinear regression relationships were fitted between the measured fluxes and controlling environmental variables (air temperature, PAR) and subsequently used to fill the missing values using a 15-day moving window. The van't Hoff equation was used to estimate missing night time NEE ($F_{c,night}$) (Lloyd and Taylor, 1994):

$$F_{c,night} = A \times e^{(B \times T)} \quad (5)$$

where A and B are estimated model coefficients and T is the air temperature. A Michaelis–Menten light response equation was used for daytime NEE ($F_{c,day}$) (Falge et al., 2001):

$$F_{c,day} = \frac{\alpha \times \mathrm{PAR} \times F_{GPP,sat}}{F_{GPP,sat} + \alpha \times \mathrm{PAR}} - F_{RE,day} \quad (6)$$

where $F_{GPP,sat}$ (the GPP at saturating light) and α (the initial slope of the light response function) are empirically estimated coefficients, and $F_{RE,day}$ (ecosystem respiration) was estimated by the extrapolation of Eq. (5) using the daytime air temperature. The daily NEE, ecosystem respiration (*Re*), and meteorological variables were averaged based on half-hourly or hourly values, and the daily values were flagged as missing when more than 20% of the data for a given day was lacking; otherwise, the daily values were calculated by multiplying the averaged half-hourly or hourly rate by 24 h. The GPP was calculated as the sum of the NEE and Re. Based on the daily dataset, the 8-day GPP mean value could be calculated. If more than 2 days of daily data were missing within a given 8-day period, the 8-day value was indicated as missing.

We adapted the harvest index methods (Prince et al., 2001) to estimate crop yield using the following equations:

$$Yield = \mathrm{GPP} \times \mathrm{AR} \times \mathrm{HI} \times \mathrm{RS} \quad (7)$$

where *Yield* is the crop yield (g C m^{-2} yr^{-1}). AR is the autotrophic respiration proportion accounting for GPP (0.53) (Waring et al., 1998), and GPP × AR indicates the net primary production (NPP) (g C m^{-2} yr^{-1}). HI refers to the harvest index, a standard measure of the proportion of total crop aboveground biomass allocated to the economic yield of the plant (Donald and Hamblin, 1976; Hay,