Article

A Production Efficiency Model-Based Method for Satellite Estimates of Corn and Soybean Yields in the Midwestern US

Qinchuan Xin [1,*], Peng Gong [1,2,*], Chaoqing Yu [1], Le Yu [1], Mark Broich [3], Andrew E. Suyker [4] and Ranga B. Myneni [5]

[1] Ministry of Education Key Laboratory for Earth System Modeling, and Center for Earth System Science, Tsinghua University, Beijing 100084, China; E-Mails: chaoqingyu@gmail.com (C.Y.); leyu@tsinghua.edu.cn (L.Y.)

[2] Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720, USA

[3] Plant Functional Biology and Climate Change Cluster Department, University of Technology Sydney, Broadway, NSW 2007, Australia; E-Mail: mark.broich@gmail.com

[4] School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA; E-Mail: asuyker1@unl.edu

[5] Department of Earth and Environment, Boston University, Boston, MA 02215, USA; E-Mail: ranga.myneni@gmail.com

* Authors to whom correspondence should be addressed; E-Mails: xqcchina@gmail.com (Q.X.); penggong@berkeley.edu (P.G.); Tel.: +86-188-1025-3088; Fax: +86-10-6279-7284.

Abstract: Remote sensing techniques that provide synoptic and repetitive observations over large geographic areas have become increasingly important in studying the role of agriculture in global carbon cycles. However, it is still challenging to model crop yields based on remotely sensed data due to the variation in radiation use efficiency (RUE) across crop types and the effects of spatial heterogeneity. In this paper, we propose a production efficiency model-based method to estimate corn and soybean yields with MODerate Resolution Imaging Spectroradiometer (MODIS) data by explicitly handling the following two issues: (1) field-measured RUE values for corn and soybean are applied to relatively pure pixels instead of the biome-wide RUE value prescribed in the MODIS vegetation productivity product (MOD17); and (2) contributions to productivity from vegetation other than crops in mixed pixels are deducted at the level of MODIS resolution. Our estimated yields statistically correlate with the national survey data for rainfed counties in the

Midwestern US with low errors for both corn (R^2 = 0.77; RMSE = 0.89 MT/ha) and soybeans (R^2 = 0.66; RMSE = 0.38 MT/ha). Because the proposed algorithm does not require any retrospective analysis that constructs empirical relationships between the reported yields and remotely sensed data, it could monitor crop yields over large areas.

1. Introduction

Because agricultural productivity directly influences food prices, trade and greenhouse-gas emission policies, and human livelihood [1,2], accurate and timely yield estimation is of primary interest to the scientific community and governments [3]. Compared with traditional ground-based methods, such as visual examination and survey sampling, remote sensing that provides synoptic and repetitive observations of the land surface is well suited for agricultural mapping [4–6] and monitoring [7,8] large geographic areas. In particular, satellite-derived vegetation indices, as measures of plant chlorophyll abundance and vegetation radiation absorption [9], have proven to be closely related to crop growth in field studies and theoretical models [10–12].

Despite tremendous efforts, obtaining accurate yield estimations based on remotely sensed data remains difficult [13,14]. One approach develops regression models that relate satellite-derived vegetation indices directly to historical yield data [15–17]. These models are essentially retrospective and are based empirically on indirect inferences, whereby changes in vegetation indices can determine variations in plant productions [14,18]. Because the regression relationship varies largely on a year-to-year basis due to inter-annual variations in climate, water availability, and management practices, the application of these models is limited to the studied regions and periods and is difficult under extreme conditions (e.g., flooding and drought) beyond historical records. Another approach applies satellite data to calibrate physiology-based crop models [19–21] that simulate the physical process of crop growth, where energy, water, carbon dioxide, and nutrients are converted into biomass [22,23]. These models have proven useful at the field scale but possess one limitation for large-scale application: they often require numerous inputs related to soil characteristics, management practices, and local weather conditions [24–26].

Given the difficulties of the two above-mentioned approaches, improving yield estimation with satellite data based on the Production Efficiency Models (PEMs) is a highly desirable goal. The theory proposed by Monteith [27,28] proposes a basis for estimating the primary productivity of vegetation from remotely sensed data. It suggests that crop yields under non-stressed conditions linearly correlate with the amount of absorbed photosynthetically active radiation. Satellite data at different resolutions have been synthesized in production efficiency models to monitor vegetation productivities [29–32] and estimate crop yields [33–35] on regional or global scales. Based on production efficiency models, a routine product (MOD17) has been derived from high temporal and moderate spatial resolution MODIS data, which has proven useful to studies on global carbon cycles [36,37].

However, validations against flux tower data indicate that MOD17 estimates exhibit large uncertainties in croplands [38]. Crop yield estimates based on the MOD17 products were found to have weak correlation with national agricultural data at the county level for the states of Montana and North Dakota in the US [18]. The MOD17 algorithm prescribes constant values of model parameters such as radiation use efficiency (RUE) for the agricultural biome regardless of crop types [36,37]. However, crops are known to possess distinctive carbon sequestration abilities and thereby different radiation use efficiencies across species, notably the C^3 and C^4 pathways. Moreover, yield estimation based on remotely sensed data requires acknowledging spatial heterogeneity effects because the pixels are typically mixtures of diverse land-cover types at the MODIS spatial resolution [39,40]. Remote sensing models that consider the conceptual partition of mixed pixels are likely to improve the yield estimates of specific crop types.

This study aims to improve yield estimation with MODIS data based on production efficiency models. Our approach is to apply crop-type specific values of radiation use efficiency (RUE) to MODIS pixels based on spatially disaggregated fine-resolution land use maps and to deduct the subpixel productivity contribution from vegetation other than crops. We demonstrate our production efficiency model-based method by quantifying the corn and soybean yields in the Midwestern US. We validate our estimates against national survey data at both the county and state levels.

2. Study Area and Materials

The Midwestern US study area (Figure 1), one of the major agricultural regions in the world, includes 12 US states: North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Michigan, Wisconsin, Illinois, Indiana, and Ohio. In this region, diversified crops are cultivated but corn (*Zea mays*) and soybeans (*Glycine max*) are the two dominant crops. In general, crop planting is completed by mid-May, when the daily soil temperatures are warm enough (above 13 °C) to initiate germination. Corn is often planted approximately one to two weeks earlier than soybeans [41]. The crops usually mature in late September and are harvested from early October until late October.

The datasets used in this study (Table 1) were collected over a three-year span between 2009 and 2011. The satellite data include two standard Terra/MODIS products: the eight-day 1000 m Leaf Area Index (LAI) & Fraction of Photosynthetically Active Radiation (FPAR) product (MOD15A2) [42] and the eight-day 1,000 m vegetation productivity product (MOD17A2) [37]. The MODIS tiling scheme separates the earth into 10° by 10° sections referenced with horizontal (H) and vertical (V) notations. Six MODIS tiles (H09V05, H10V04, H10V05, H11V04, H11V05, and H12V04) that cover the Midwestern US were downloaded from the U.S. Geological Survey (USGS) EROS data center [43]. All products were composited every eight days, starting from Day-of-Year (DOY) 1 in each calendar year. The product details are available on the MODIS data website [44].

The Quick Stats database from the National Agricultural Statistics Service (NASS) of the US Department of Agriculture (USDA) provides reliable crop survey data at the county, state, and national levels [45]. We obtained statistical data describing the yields and harvested areas of corn and soybean in each county. Based on the Crop Production 2011 Summary released by the USDA, the Midwestern US produced 277.1 Tg of corn and 70.2 Tg of soybeans in 2011, which were approximately 88.3% and 84.4%, respectively, of the total US corn and soybean productions. Planted areas for corn and soybean in

Midwestern US were 5.6×10^7 ha in 2011, which were approximately 70.9% of the total planted areas (7.9×10^7 ha) for all principle crops such as corn, sorghum, oats, barley, rye, wheat, rice, soybeans, peanuts, sunflower, cotton, and potatoes.

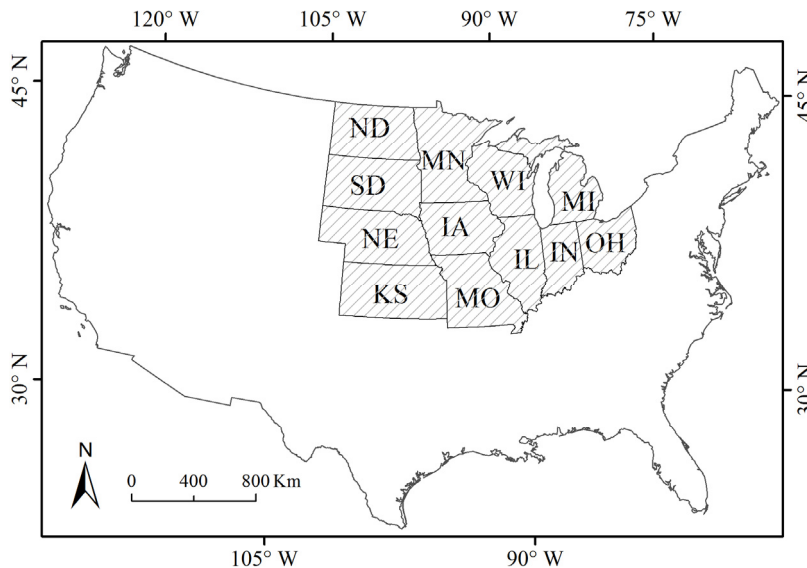**Figure 1.** The study area of the Midwestern US comprises 12 states.

Table 1. The datasets used in this study.

Data Type	Data Source	Description
satellite data	USGS EROS center	MODIS LAI/FPAR data (MOD15A2)
	USGS EROS center	MODIS vegetation productivity data (MOD17A2)
classification maps	NASS Cropland Data Layer	crop-type specific classification maps at 30 or 56 m
	USGS early warning program	a classification map of irrigated areas at 250 m
national survey	NASS Quick Stats	statistics of crop yields and harvested areas

The NASS Cropland Data Layer (CDL) program produces crop-type specific classification maps based on multi-sensor satellite imagery using training data from extensive ground surveys [46]. The classification map is available for each state at a resolution of 30 m. In the present study, the CDL maps were reprojected to a sinusoidal projection, mosaicked, and scaled up to 500 m as percent maps (Figure 2) to match the spatial resolution of MODIS. To avoid errors in regions with sparse agriculture, our analysis was confined to counties that were planted with corn or soybean on at least 5% of their total areas.

Although irrigation helps boost crop yields, most PEMs to date have not compensated for the impacts of irrigation. To investigate the yield estimation in irrigated areas, we sourced the irrigation map (Figure 3) from the USGS early warning program [47]. The irrigated areas in the Midwestern US were mapped for 2007 and were primarily located in Nebraska and Kansas. Because irrigation is unlikely to vary on a year-to-year basis, no change in irrigated area was assumed during the three years (from 2009 to 2011). The spatial pattern of irrigated areas was consistent with that of the agriculture areas (Figure 2), especially for corn (Figure 2A), which consumes large volumes of water before, during and for several weeks after pollination. To study the effects of irrigation, counties were considered as irrigated if more than 1% of the crop areas were under some form of irrigation.

Figure 2. The percentage maps of corn and soybean in 2011 from the NASS CDL program. (**A**) 2011 Corn map from NASS CDL; (**B**) 2011 Soybean map from NASS CDL.
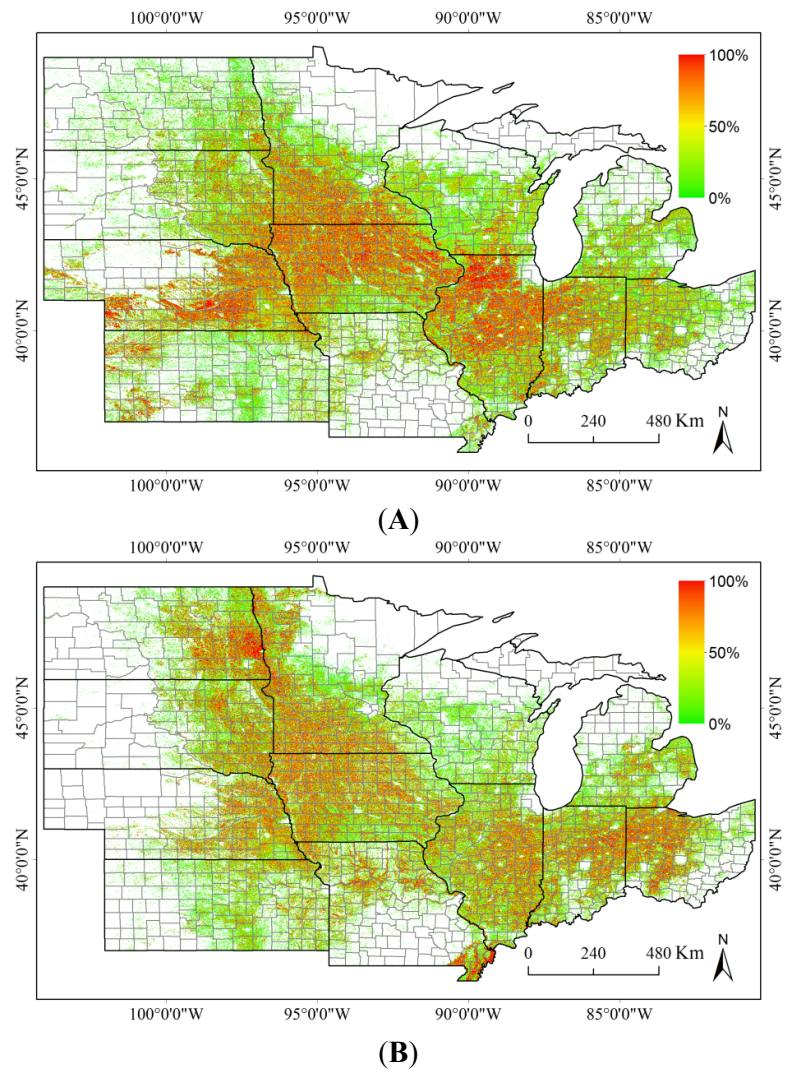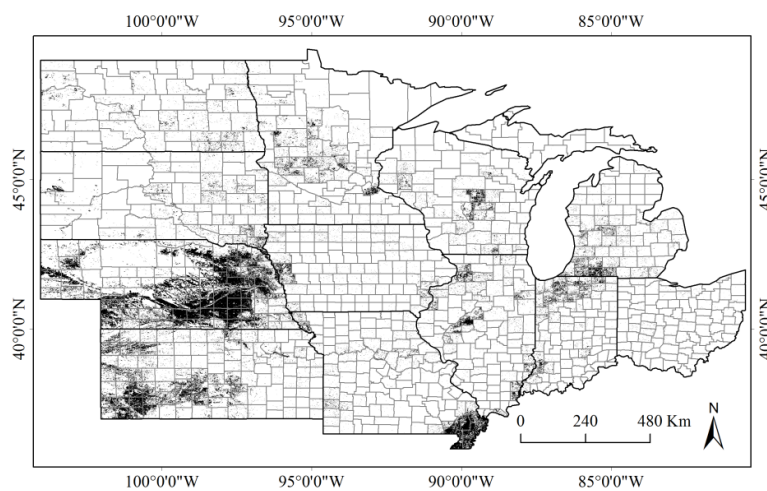

(A)

(B)

Figure 3. The map of irrigated areas from the early warning program of USGS in 2007.

3. Theoretical Background and Improvements

3.1. A Brief Description of the MOD17 Algorithm

Although the MOD17 algorithm has been described in detail in the MOD17 User's Guide [36], we briefly describe it here because the algorithm is fundamental to understanding our proposed method. The MOD17 product [37,48] estimates the gross primary production (GPP) and net photosynthesis (PSN) of each pixel every 8 days and estimates the GPP and NPP for each pixel annually. These estimates are produced based on three upstream datasets: the MODIS land-cover classification product (MOD12Q1) [4], MODIS LAI/FPAR product (MOD15) [42], and large-scale meteorological data from the NASA Data Assimilation Office (DAO). The coefficients for each biome type (decided based on MOD12Q1) are defined in a Biome Parameter Look-Up Table (BPLUT) from a general ecosystem model [49].

The daily MODIS GPP is estimated as:

$$GPP = PAR \times FPAR \times \varepsilon_g \times f(\varepsilon) \tag{1}$$

where PAR (MJ·day^{-1}) is the daily incident photosynthetically active radiation, as estimated using the DAO data; FPAR (dimensionless) is the fraction of absorbed photosynthetically active radiation, as obtained from the upstream MOD15 product; ε_g (g·C·MJ^{-1} PAR) is the radiation use efficiency (RUE) for GPP calculations when the environment is not limiting plant carbon uptake, as defined in the BPLUT for the agricultural biome; and $f(\varepsilon)$ accounts for the influences of environmental stress, such as temperature and vapor pressure deficit, as calculated based on the DAO data and BPLUT.

The concept of daily PSN considers the maintenance respiration of leaves and fine roots:

$$PSN = GPP - Leaf\_MR - Froot\_MR \tag{2}$$

where the daily maintenance respiration is estimated for the leaves (Leaf_MR) and fine roots (Froot_MR) separately:

$$Leaf\_MR = LAI / SLA \times leaf\_mr\_base \times Q10\_mr^{[(Tavg - 20.0) / 10.0]} \tag{3}$$

$$Froot\_MR = LAI / SLA \times froot\_leaf\_ratio \times froot\_mr\_base \times Q10\_mr^{[(Tavg - 20.0) / 10.0]} \tag{4}$$

where LAI (m^2·leaf·m^{-2} ground area) is the leaf area index and SLA (projected leaf area m^2·kg^{-1}·leaf C) is the specific leaf area for a given pixel; leaf_mr_base and froot_mr_base (kg·C·kg·C^{-1}·day^{-1}·20 C) are the maintenance respiration of leaves and fine roots per unit mass at 20 C, respectively; froot_leaf_ratio (unitless) is the ratio of the fine root to leaf mass; Q10_mr (unitless) is an exponent shape parameter that controls respiration as a function of temperature; and Tavg (°C) is the average daily temperature. LAI is obtained from MOD15, Tavg is estimated from the DAO meteorological data, and other biome-specific coefficients are obtained from BPLUT.
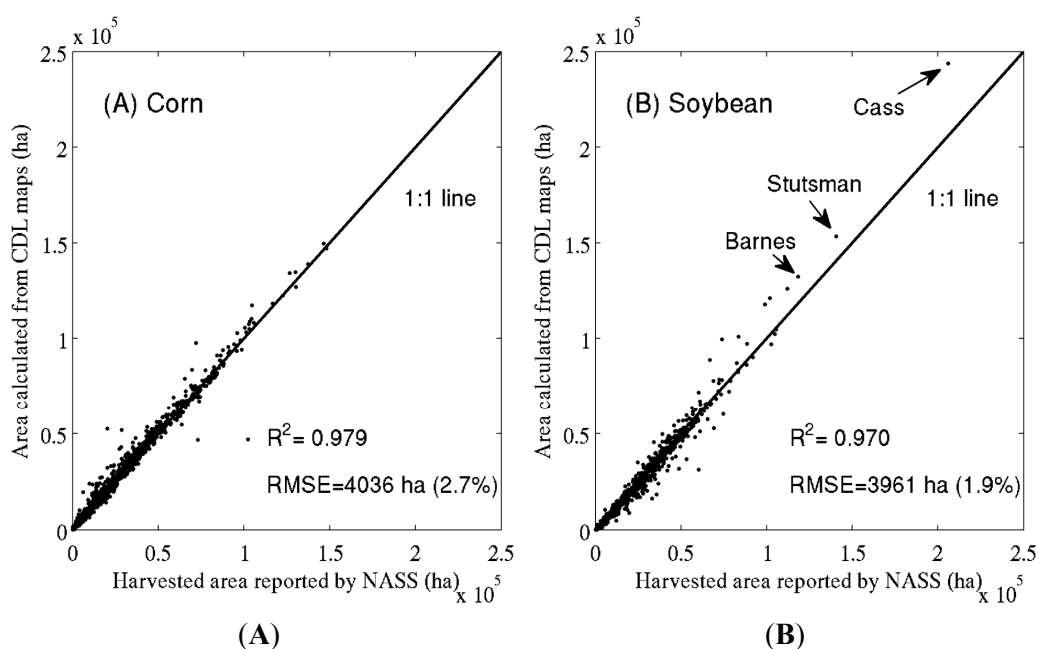
3.2. Improvements to Estimate Crop Productivity

Based on the MOD17 algorithm, we attempted to improve the crop yield estimation in the Midwestern US using the following techniques: (1) crop-type specific masks were derived from fine-resolution NASS CDL maps to identify relatively pure MODIS pixels; (2) a specific RUE was

applied for each crop type; and (3) contributions to primary production from other vegetation types were deducted when possible.

First, the MOD17 algorithm relies on the MODIS land cover product (MOD12), a classification scheme that does not differentiate crop types. Several studies have indicated that the use of the MODIS land cover product is likely to overestimate the agriculture areas relative to those reported by the NASS [18]. For this reason, we analyzed relatively pure corn or soybean pixels based on spatially disaggregated CDL maps (Figure 2) rather than applying a general cropland mask based on the MODIS land cover product. By counting corn pixels in the CDL maps, we observed that the estimated corn areas agreed well with the harvested corn areas reported by the NASS in 2011 (Figure 4A). The soybean areas were slightly overestimated, especially for the top soybean-producing counties (Figure 4B). It is appropriate to assume that pixels with more than 75% corn or soybean cover are relatively pure endmembers [41,50]. The satellite data (resampled to 500 m) were spatially averaged based on those representative pixels to derive the mean values for each county every 8 days.

Figure 4. Comparisons between crop areas calculated from CDL maps and reported by NASS in 2011 for counties in the Midwestern US. (**A**) Corn area estimates; (**B**) Soybean area estimates.

Second, we applied crop-type specific RUE values for corn and soybeans (Table 2) instead of a biome-wide value in the MOD17 algorithm to estimate the yields. The MOD17 algorithm prescribes a universal RUE value for all crops, regardless of their types [36,49]. However, the field measurements that quantify RUE values based on the relationship between biomass accumulation and cumulative radiation interception suggest that the carbon sequestration abilities vary widely across crop types [34,51]. Sinclair and Muchow [52] reviewed several studies and reported that the seasonal corn RUE ranges from 2.6 to 3.4 $g \cdot MJ^{-1}$ and soybean RUE ranges from 1.32 to 2.52 $g \cdot MJ^{-1}$. To account for the RUE differences, we applied 3.35 $g \cdot MJ^{-1}$ for corn and 1.44 $g \cdot MJ^{-1}$ for soybeans as obtained from recent field

measurements in Iowa [53]. Although some studies observed that the RUE values varied across time [38,52], the RUE values applied here were derived as seasonal mean values in the experiments. Because the respiration costs of annual crop plants were deducted from the field measurements of accumulated biomass, the use of these RUE values would result in crop productivities that can be linked to crop yields directly.

Table 2. The parameters used in our study for estimating crop yields.

Parameter	Description	Corn	Soybean	Units
ε	the radiation use efficiency	3.35	1.44	$g \cdot MJ^{-1}$ PAR
RS	the root: shoot ratio	0.18	0.15	dimensionless
HI	the harvest index	0.53	0.42	dimensionless
MC	the moisture content of the grain	0.11	0.10	dimensionless

Finally, our analysis attempted to address the productivity contributed by vegetation other than crops because pixels are commonly mixed at the subpixel level of MODIS. The conceptual partition of mixed pixels could avoid the biased estimation of crop yields. In other words, the GPP for each pixel was partitioned into two components in our algorithm:

$$GPP_{total, DOY\ n} = GPP_{crop, DOY\ n} + GPP_{other, DOY\ n} \tag{5}$$

where $GPP_{total,\ DOY\ n}$ is the GPP value for a specific Day Of Year (DOY) n; $GPP_{crop,\ DOY\ n}$ is the contribution from crops (*i.e.*, corn or soybeans) for the same time period; and $GPP_{other,\ DOY\ n}$ is the contribution from other vegetation types.

Separating each component in mixed pixels is challenging in remote sensing models because dynamic information of subpixel land-cover composition is often unavailable. Here, we propose an algorithm to approximate the $GPP_{other,\ DOY\ n}$ component based on plant LAI. Given that corn and soybeans in the Midwestern US have not emerged or in very early growth stages by mid-May, when the natural vegetation has already accumulated significant leaf area [41], it is appropriate to approximate the LAI of vegetation other than crops during the growing season by the maximum LAI before mid-May:

$$LAI_{other} \approx max(LAI_{total, DOY\ n = 89:N_0}) \tag{6}$$

where LAI_{other} is the estimated LAI of vegetation other than crops; and $max(LAI_{total,DOY\ n=89:N_0})$ is the maximum LAI value between April (DOY 89) and before mid-May (DOY N_0), as obtained from MOD15. The DOY N_0 value was assumed to be 129 for corn and 137 for soybeans because corn is typically planted approximately one to two weeks earlier than soybeans.

Similar to Equations (3) and (4), the maintenance respiration of other vegetation is estimated to be:

$$Leaf\_MR_{other} = LAI_{other} / SLA \times leaf\_mr\_base \times Q10\_mr^{[(Tavg - 20.0) / 10.0]} \tag{7}$$

$$Froot\_MR_{other} = LAI_{other} / SLA \times froot\_leaf\_ratio \times froot\_mr\_base \times Q10\_mr^{[(Tavg - 20.0) / 10.0]} \tag{8}$$

$$MR_{other} = Leaf\_MR_{other} + Froot\_MR_{other} \tag{9}$$

where MR_{other} is the maintenance respiration of other vegetation; $Leaf\_MR_{other}$ and $Froot\_MR_{other}$ are the maintenance respiration of the leaves and fine roots of other vegetation, respectively; Tavg (°C) is the

average daily temperature, as estimated from the DAO datasets; and the values for the other parameters are the same as the agricultural biome in the MOD17 algorithm.
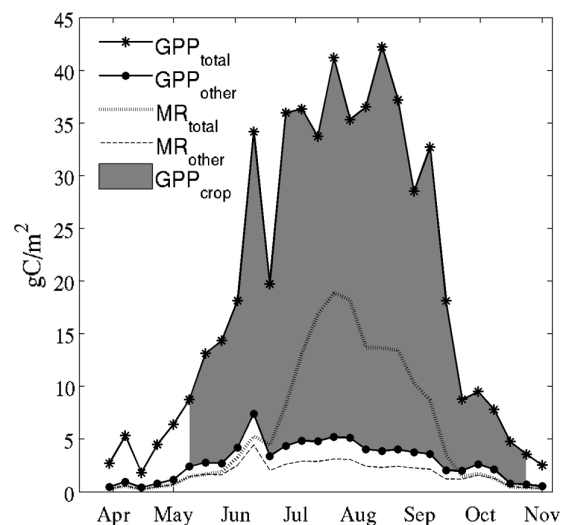
Because the ratio of NPP to GPP, termed as the carbon use efficiency (CUE), is modeled as a constant in many ecosystem models [29,31], the GPP contribution from other vegetation can then be estimated to be:

$$\text{GPP}_{\text{other, DOY n}} = \text{MR}_{\text{other, DOY n}} / (1 - \text{CUE}) \tag{10}$$

where $\text{GPP}_{\text{other, DOY n}}$ is the GPP of other vegetation types for a specific DOY n, $\text{MR}_{\text{other, DOY n}}$ is the maintenance respiration, and CUE is the carbon use efficiency. Published studies indicate a CUE value from 0.3 to 0.5 for natural vegetation such as forests [54]. We approximated a carbon use efficiency of 0.4 in this study.

Our estimation of each GPP and MR component throughout the growing season is further illustrated in Figure 5, which presents the example of Olmsted County, Minnesota. The model was applied at the pixel level, and estimates were aggregated to the county level. The proposed algorithm captured the subtle variation of the GPP and MR caused by climatic factors in the time series. The sudden drop of MODIS GPP in the middle of the growing season (DOY 169) was most likely due to cloud contamination rather than harvest or other cultural practices. Clouds and aerosols can largely influence crop GPP in a short period of time by reducing the incident PAR [55].

Figure 5. A simple representation of MODIS GPP (Gross Primary Production) and MR (Maintenance Respiration) components of corn pixels in Olmsted County, Minnesota.

3.3. Converting MODIS GPP Estimates to Crop Yields

To compare with the NASS Quick Stats dataset, the GPP estimated by our method and the GPP derived from the MOD17 product (hereafter referred as to the MODIS-GPP algorithm) were converted to crop yields based on a common method used in agricultural studies [18,56]:

$$\text{Yield} = \sum_{n > N_0}^{n = N_1} \text{GPP}_{\text{crop, DOY n}} \times \frac{\text{HI}}{(1 + \text{RS})} \times \frac{1}{1 - \text{MC}} \tag{11}$$

where $GPP_{crop,\ DOYn}$ is the estimated GPP of crops (*i.e.*, corn or soybeans) for a specific DOY n. The yield was integrated for an apparent growing season, which began in mid-May (DOY N_0) and ended in late-October (DOY N_1). The value for N_1, the growing season endpoint, was set as 297 in our study, and N_0 was defined as 129 for corn and 137 for soybeans, as in Equation (6). RS is the root to shoot ratio, HI is the harvest index, and MC is the grain moisture content. We chose MODIS GPP instead of NPP in this formulation because our preliminary tests as similar to other studies [18] revealed that yield estimates derived from MODIS NPP were relatively low.

The crop-type specific parameters involved in our calculations are defined in Table 2. The root: shoot ratio, which depends on the growing stages and fertilization of crops, is considered to be constant for a specific crop type when harvested. Prince [56] reviewed past studies and chose the root: shoot ratio of 0.18 for corn and 0.15 for soybeans. The harvest index (HI) varies little for each crop species, except under extreme stress conditions [56]. The HI of corn is 0.53, as measured by experiments in nine US locations [57]. The soybean HI is 0.42 based on reported regression equations [56,58]. The moisture content is estimated as 11% for corn grain and 10% for soybeans [35]. The values of these parameters vary in different studies. Lobell *et al.* [34] observed that regional yield predictions were insensitive to the variability in the radiation use efficiency and harvest index. Prince *et al.* [56] attributed most of the uncertainty to the variability in the harvest index and root:shoot ratio, which resulted in an overall variability of ~10%. The uncertainty associated with variation in these defined parameters was considered to be of secondary importance in the context of other modeling uncertainties such as the RUE variations and the MODIS data.

4. Results

4.1. Analysis at the County Level

Our estimated corn and soybean yields for 2011 agree well with the NASS reported survey data for counties dominated by rainfed cultivation (Figure 6). The coefficients of determination (R^2), as measures of the correlations between estimated values and reference data, are 0.77 for corn and 0.66 for soybeans. The corn yields were slightly underestimated, as indicated by the mean error (ME) of $-0.18\ MT \cdot ha^{-1}$, which was possibly due to the variation of parameters defined in Table 2. The variation in the parameters, as the linear coefficients in the GPP-Yield relationship (Equation (11)), has little influence on the R^2 values. The high R^2 values obtained for both the corn and soybeans demonstrate that the crop yields estimated from MODIS data strongly correlate with the NASS survey data at the county level.

Table 3 compares our algorithm with the MODIS-GPP algorithm (in parentheses) in the period 2009 to 2011. Our algorithm demonstrates improved crop yield estimation over the MODIS-GPP algorithm, as indicated by higher R^2 values and lower root mean square errors (RMSE). Similar to the findings in other studies [18], the R^2 values based on the MODIS-GPP algorithm between estimated crop yields and NASS survey data were only 0.15~0.46 for corn and 0.35~0.53 for soybeans in rainfed counties. In addition, the yields were considerably underestimated for corn (negative mean error) but overestimated for soybeans (positive mean error). Our estimated biases were considerably smaller, as indicated by the mean errors. Because the soybean RUE (1.44 $g \cdot MJ^{-1}$ PAR) is close to the biome-wide value in MOD17

(1.36 g·MJ^{-1} PAR), the GPP contributed by other vegetation was the major factor that resulted in overestimation in MOD17. In contrast, corn as a C4 plant has a much higher RUE value (3.35 g·MJ^{-1} PAR) based on a photosynthetic pathway different from soybean, a C3 plant. The low RUE value prescribed in MOD17 is the main contributor to the large underestimation of corn yields. Therefore, we need to consider the effects of mixed pixels and the RUE variation across crop types when estimating yields based on remotely sensed data.

Figure 6. Comparisons between crop yields estimated from MODIS data and reported by the NASS for rainfed counties in the Midwestern US. The black line is the 1:1 line. (**A**) Rainfed corn; (**B**) Rainfed soybean.
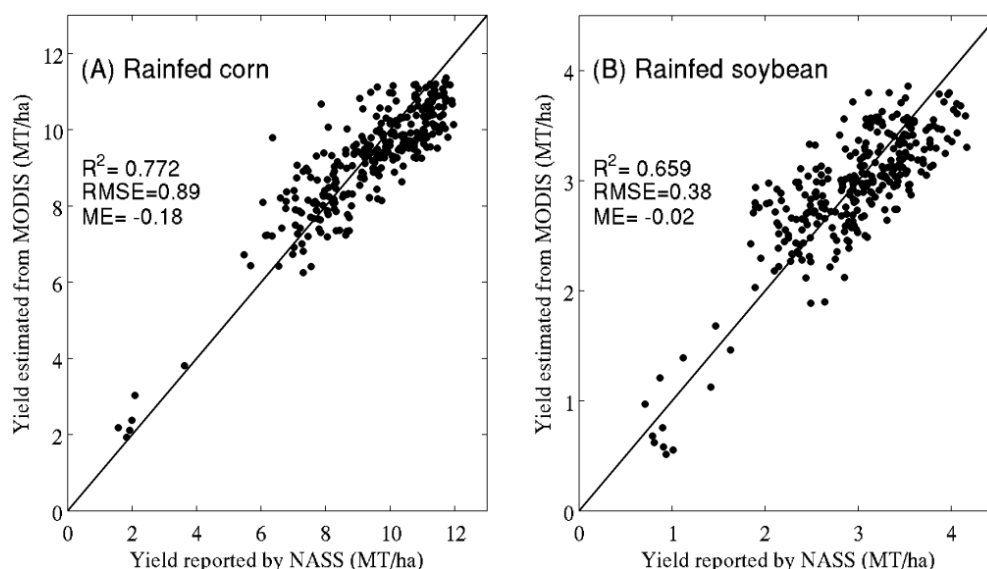

Table 3. Statistics between crop yields estimated from our approach using MODIS data and reported by the NASS for rainfed counties for each year from 2009 to 2011. For comparison, values in parentheses are statistics using the standard MOD17 products.
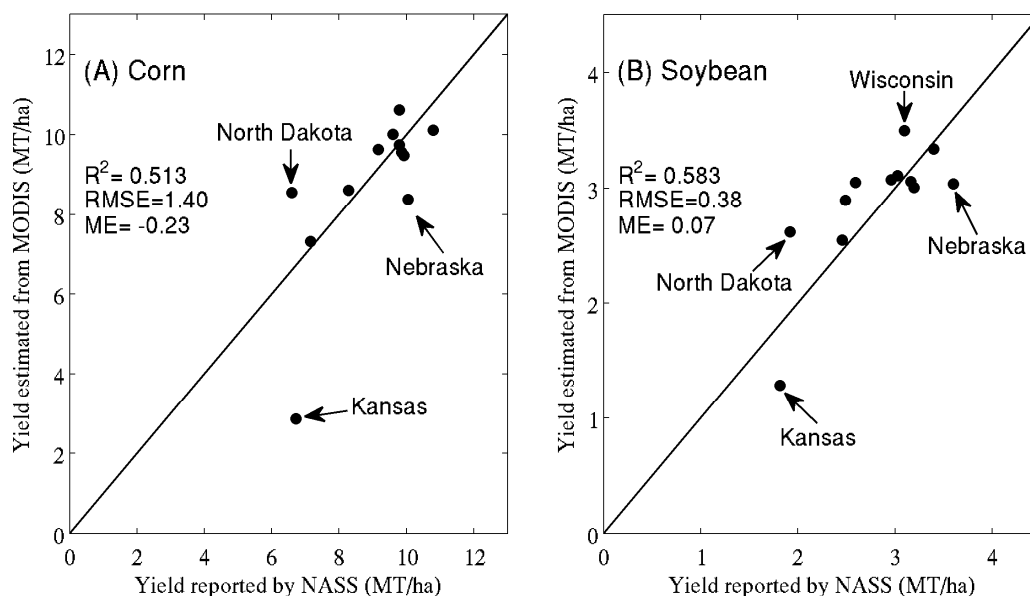
Year	Corn			Soybean		
	R^2	RMSE (MT/ha)	ME (MT/ha)	R^2	RMSE (MT/ha)	ME (MT/ha)
2009	0.55 (0.15)	1.21 (5.52)	−0.60 (−5.39)	0.50 (0.35)	0.38 (0.86)	−0.07 (0.77)
2010	0.54 (0.22)	1.17 (4.65)	−0.14 (−4.38)	0.73 (0.53)	0.30 (0.97)	−0.09 (0.89)
2011	0.77 (0.46)	0.89 (4.56)	−0.18 (−4.28)	0.66 (0.53)	0.38 (1.06)	−0.02 (0.95)

4.2. Analysis at the State Level

The analysis at the state level was performed in the same manner as at the county level, except that we did not try to separate irrigated counties. The lower R^2 values of the yields in 2011 (Figure 7) for corn (0.51) and for soybeans (0.58) were partly caused by a large underestimation in Kansas and Nebraska due to the irrigation effects. Because plants in irrigated areas often have higher transpiration rates that help cool their leaves, the MODIS-GPP algorithm, which is sensitive to the environmental factors such as daily temperature or vapor pressure deficit, would tend to underestimate GPP in irrigated areas. Interestingly, the yields were overestimated for the northern states, such as North Dakota and Wisconsin, which may

have been because crop planting typically starts later in those states than in other regions. It would be more appropriate to use the LAI values from late May rather than mid-May to approximate the LAI of other vegetation during the growing season. Future analysis should incorporate the timing of crop phenology derived from MODIS to improve the yield estimation [59].

Figure 7. Comparisons between crop yields estimated from MODIS data and reported by the NASS for states in the Midwestern US. The black line is the 1:1 line. (**A**) Corn; (**B**) Soybean.

We also estimated the final crop production based on satellite data using the following techniques: (i) the crop yields were estimated from MODIS data; (ii) the harvested areas were derived from the CDL maps by counting pixels within each state, and (iii) the crop production was calculated as the product of crop yields and harvested areas. Compared with the NASS reported data in the Midwestern US, the total crop productions were underestimated (Table 4) due to the underestimated yields in irrigated counties. The errors of estimated total productions were −1.5% for corn and −0.6% for soybeans in 2011 and fall within the acceptable ranges in other years. Note that the crop areas derived from the NASS CDL maps may also contribute errors to our estimation of crop production here.

Table 4. Estimated and reported corn and soybean production for 12 states in the Midwestern US.

Year	Corn Production			Soybean Production		
	Estimated (Tg)	Reported (Tg)	Error	Estimated (Tg)	Reported (Tg)	Error
2009	253.6	291.7	−13.04%	68.2	75.6	−9.74%
2010	268.1	276.3	−2.95%	73.9	77.7	−4.79%
2011	273.0	277.1	−1.49%	69.8	70.2	−0.56%

5. Discussions

5.1. Major Findings

We employed three key methods to estimate crop yields based on satellite data. First, there are clear benefits to retrieving subpixel areas of different crop types and other vegetation based on land-cover maps at a fine resolution. Although some studies concluded that coarse spatial resolution satellite data (0.05 degree or approximately 5.5 km) were advantageous over higher spatial resolution data for estimating the yield because individual pixels will likely shift between crop types on a year-to-year basis [15], we observed that a combination of fine and moderate resolution data was preferable for accurately estimating yield because retrieving information related to crop types is essential to yield modeling.

Second, our analysis of corn and soybeans reveals that applying crop-type specific RUE values in the PEMs is necessary to estimate crop yields properly. Global-scale PEMs mainly focus on quantifying vegetation productivity for a general biome type [38]. For example, the default RUE value in MOD17 is the same for all crops and grasses [36,48]. However, it would be inappropriate to apply flat RUE values to estimate the yields for different crops, especially for productive C4 plants such as corn.

Finally, it is necessary to deduct the GPP contributed by vegetation other than crops when estimating the yield at the spatial resolution of MODIS. As illustrated, crop yields are likely to be overestimated without such considerations when applying RUE values measured in field studies. Our findings are consistent with other studies, which observed that the direct use of field-measured RUE values in global-scale PEMs would result in an overestimated GPP [35].

5.2. Limitations and Future Improvements

Although we were able to adequately estimate crop yields over large geographical regions, some limitations necessitate further improvements. Our algorithm is based on the assumption that the LAI of other vegetation is relatively constant after mid-May. For much of the study area where corn and soybeans are planted, this assumption is reasonable, and our model improved the estimated crop yields as demonstrated. However, such an assumption is not entirely correct and does not hold for crops with different phenological cycles such as winter wheat. It would be of interest to identify appropriate solutions to modeling GPP by integrating MODIS data and fine-resolution land use maps.

There are also uncertainties related to irrigation practice and cloud cover. Although the MODIS GPP/NPP algorithm considers the daily atmospheric water deficit [48], our study implies that it still requires better parameterization for irrigated areas. Given the availability of time-series measurements from the flux towers located in irrigated areas [60], we should be able to calibrate the production efficiency models to improve yield estimation. Cloud cover is another factor that may influence the yield estimation from remotely sensed data. Even though the fitting process can reduce the influence of intermittent cloud-cover, the proposed method may not perform well in areas with persistent cloud covers, which could result in large data gaps in the MODIS time series.

6. Conclusions

This study proposes a production efficiency model-based algorithm for estimating crop yield in the Midwestern US. The yield estimates have shown to agree well with the NASS survey data in rainfed counties for both corn (R^2 = 0.77; RMSE = 0.89 MT·ha^{-1}) and soybeans (R^2 = 0.66; RMSE = 0.38 MT·ha^{-1}). Different from previous studies, we found that the MODIS GPP algorithm was capable of making reasonable yield estimates with certain considerations. Our analysis suggests using field-measured RUE values instead of biome-wide RUE values in the production efficiency models when estimating the yield for a specific crop type. Moreover, it is important to consider the productivity contribution from other vegetation in mixed pixels. The proposed algorithm does not require a retrospective analysis that constructs the empirical relationships between reported yields and remotely sensed data, and it has the potential for monitoring crop yields over large areas. Future work on yield estimation based on production efficiency models should focus on investigations that consider the subpixel spatial heterogeneity and irrigation effects.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Hertel, T.W.; Burke, M.B.; Lobell, D.B. The poverty implications of climate-induced crop yield changes by 2030. *Glob. Environ. Chang.* **2010**, *20*, 577–585.
2. Lobell, D.B.; Schlenker, W.; Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* **2011**, *333*, 616–620.
3. Doraiswamy, P.C.; Moulin, S.; Cook, P.W.; Stern, A. Crop yield assessment from remote sensing. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 665–674.
4. Friedl, M.A.; McIver, D.K.; Hodges, J.C.F.; Zhang, X.Y.; Muchoney, D.; Strahler, A.H.; Woodcock, C.E.; Gopal, S.; Schneider, A.; Cooper, A.; *et al.* Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302.
5. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654.
6. Thenkabail, P.S.; Wu, Z. An automated cropland classification algorithm (ACCA) for Tajikistan by combining Landsat, MODIS, and secondary data. *Remote Sens.* **2012**, *4*, 2890–2918.

7. Pan, Y.; Li, L.; Zhang, J.; Liang, S.; Zhu, X.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-EVI time series data using the crop proportion phenology index. *Remote Sens.Environ.* **2012**, *119*, 232–242.

8. Vancutsem, C.; Marinho, E.; Kayitakire, F.; See, L.; Fritz, S. Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. *Remote Sens.***2012**, *5*, 19–41.

9. Myneni, R.B.; Hall, F.G.; Sellers, P.J.; Marshak, A.L. The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.***1995**, *33*, 481–486.

10. Myneni, R.B.; Nemani, R.R.; Running, S.W. Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.***1997**, *35*, 1380–1393.

11. Veroustraete, F.; Patyn, J.; Myneni, R.B. Estimating net ecosystem exchange of carbon using the normalized difference vegetation index and an ecosystem model. *Remote Sens.Environ.* **1996**, *58*, 115–130.

12. Thenkabail, P.S.; Smith, R.B.; de Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens.Environ.* **2000**, *71*, 158–182.

13. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.***2013**, *5*, 949–981.

14. Moulin, S.; Bondeau, A.; Delecolle, R. Combining agricultural crop models and satellite observations: From field to regional scales. *Int. J. Remote Sens.***1998**, *19*, 1021–1036.

15. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens.Environ.* **2010**, *114*, 1312–1323.

16. Mulianga, B.; Bégué, A.; Simoes, M.; Todoroff, P. Forecasting regional sugarcane yield based on time integral and spatial aggregation of MODIS NDVI. *Remote Sens.***2013**, *5*, 2184–2199.

17. Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sens.***2010**, *2*, 673–696.

18. Reeves, M.C.; Zhao, M.; Running, S.W. Usefulness and limits of MODIS GPP for estimating wheat yield. *Int. J. Remote Sens.***2005**, *26*, 1403–1421.

19. Claverie, M.; Demarez, V.; Duchemin, B.; Hagolle, O.; Ducrot, D.; Marais-Sicre, C.; Dejoux, J.-F.; Huc, M.; Keravec, P.; Béziat, P. Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sens.Environ.* **2012**, *124*, 844–857.

20. Launay, M.; Guerif, M. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agric. Ecosyst. Environ.* **2005**, *111*, 321–339.

21. Mishra, V.; Cruise, J.F.; Mecikalski, J.R.; Hain, C.R.; Anderson, M.C. A remote-sensing driven tool for estimating crop stress and yields. *Remote Sens.***2013**, *5*, 3331–3356.

22. Fumoto, T.; Kobayashi, K.; Li, C.; Yagi, K.; Hasegawa, T. Revising a process-based biogeochemistry model (DNDC) to simulate methane emission from rice paddy fields under various residue management and fertilizer regimes. *Glob. Chang. Biol.* **2008**, *14*, 382–402.

23. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT cropping system model. *Eur. J. Agron.* **2003**, *18*, 235–265.

24. Keating, B.A.; Carberry, P.S.; Hammer, G.L.; Probert, M.E.; Robertson, M.J.; Holzworth, D.; Huth, N.I.; Hargreaves, J.N.G.; Meinke, H.; Hochman, Z.; *et al.* An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **2003**, *18*, 267–288.

25. Krobel, R.; Smith, W.N.; Grant, B.B.; Desjardins, R.L.; Campbell, C.A.; Tremblay, N.; Li, C.S.; Zentner, R.P.; McConkey, B.G. Development and evaluation of a new Canadian spring wheat sub-model for DNDC. *Can. J. Soil. Sci.* **2011**, *91*, 503–520.

26. Tonitto, C.; David, M.B.; Drinkwater, L.E.; Li, C.S. Application of the DNDC model to tile-drained Illinois agroecosystems: model calibration, validation, and uncertainty analysis. *Nutr. Cycl. Agroecosyst.* **2007**, *78*, 51–63.

27. Monteith, J.L. Solar-radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* **1972**, *9*, 747–766.

28. Monteith, J.L. Climate and efficiency of crop production in Britain. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **1977**, *281*, 277–294.

29. Potter, C.S.; Randerson, J.T.; Field, C.B.; Matson, P.A.; Vitousek, P.M.; Mooney, H.A.; Klooster, S.A. Terrestrial ecosystem production—A process model-based on global satellite and surface data. *Glob. Biogeochem. Cy.* **1993**, *7*, 811–841.

30. Ruimy, A.; Saugier, B.; Dedieu, G. Methodology for the estimation of terrestrial net primary production from remotely sensed data. *J. Geophys. Res.: Atmos.* **1994**, *99*, 5263–5283.

31. Running, S.W.; Coughlan, J.C. A general-model of forest ecosystem processes for regional applications. 1. Hydrologic balance, canopy gas-exchange and primary production processes. *Ecol. Model.* **1988**, *42*, 125–154.

32. Running, S.W.; Justice, C.O.; Salomonson, V.; Hall, D.; Barker, J.; Kaufmann, Y.J.; Strahler, A.H.; Huete, A.R.; Muller, J.P.; Vanderbilt, V.; *et al.* Terrestrial remote-sensing science and algorithms planned for EOS MODIS. *Int. J. Remote Sens.* **1994**, *15*, 3587–3620.

33. Liu, J.; Pattey, E.; Miller, J.R.; McNairn, H.; Smith, A.; Hu, B. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. *Remote Sens. Environ.* **2010**, *114*, 1167–1177.

34. Lobell, D.B.; Asner, G.P.; Ortiz-Monasterio, J.I.; Benning, T.L. Remote sensing of regional crop production in the Yaqui Valley, Mexico: Estimates and uncertainties. *Agric. Ecosyst. Environ.* **2003**, *94*, 205–220.

35. Lobell, D.B.; Hicke, J.A.; Asner, G.P.; Field, C.B.; Tucker, C.J.; Los, S.O. Satellite estimates of productivity and light use efficiency in United States agriculture, 1982–98. *Glob. Chang. Biol.* **2002**, *8*, 722–735.

36. Heinsch, F.A.; Reeves, M.; Votava, P.; Kang, S.; Milesi, C.; Zhao, M.; Glassy, J.; Jolly, W.M.; Loehman, R.; Bowker, C.F. *User's Guide GPP and NPP (MOD17A2/A3) Products NASA MODIS Land Algorithm*; MODIS Land Team: Washington, DC, USA, 2003; pp. 1–57.

37. Zhao, M.; Heinsch, F.A.; Nemani, R.R.; Running, S.W. Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sens. Environ.* **2005**, *95*, 164–176.

38. Turner, D.P.; Urbanski, S.; Bremer, D.; Wofsy, S.C.; Meyers, T.; Gower, S.T.; Gregory, M. A cross-biome comparison of daily light use efficiency for gross primary production. *Glob. Chang. Biol.* **2003**, *9*, 383–395.

39. Rembold, F.; Atzberger, C.; Savin, I.; Rojas, O. Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sens.***2013**, *5*, 1704–1733.
40. Turner, D.P.; Gower, S.T.; Cohen, W.B.; Gregory, M.; Maiersperger, T.K. Effects of spatial variability in light use efficiency on satellite-based NPP monitoring. *Remote Sens. Environ.* **2002**, *80*, 397–405.
41. Li, A.N.; Liang, S.L.; Wang, A.S.; Qin, J. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques. *Photogramm. Eng. Remote Sens.***2007**, *73*, 1149–1157.
42. Myneni, R.B.; Hoffman, S.; Knyazikhin, Y.; Privette, J.L.; Glassy, J.; Tian, Y.; Wang, Y.; Song, X.; Zhang, Y.; Smith, G.R.; *et al.* Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens.Environ.* **2002**, *83*, 214–231.
43. USGS Earth Resources Observation and Science (EROS) Center. Available online: http://eros.usgs.gov/ (accessed on 1 November 2013).
44. USGS Land Processes Distributed Active Archive Center (LP DAAC). Available online: https://lpdaac.usgs.gov/ (accessed on 1 November 2013).
45. National Agricultural Statistics Service (NASS) Quick Stats Database. Available online: http://www.nass.usda.gov/Quick_Stats/ (accessed on 1 November 2013).
46. Han, W.G.; Yang, Z.W.; Di, L.P.; Mueller, R. CropScape: A web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Comput. Electron. Agric.* **2012**, *84*, 111–123.
47. Pervez, M.S.; Brown, J.F. Mapping irrigated lands at 250-m scale by merging MODIS data and national agricultural statistics. *Remote Sens.***2010**, *2*, 2388–2412.
48. Running, S.W.; Nemani, R.R.; Heinsch, F.A.; Zhao, M.S.; Reeves, M.; Hashimoto, H. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience* **2004**, *54*, 547–560.
49. Running, S.W.; Thornton, P.E.; Nemani, R.; Glassy, J.M. Global terrestrial gross and net primary productivity from the earth observing system. *Method. Ecosyst. Sci.* **2000**, *5*, 44–57.
50. de Wit, A.; Duveiller, G.; Defourny, P. Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations. *Agric. For. Meteorol.* **2012**, *164*, 39–52.
51. Lindquist, J.L.; Arkebauer, T.J.; Walters, D.T.; Cassman, K.G.; Dobermann, A. Maize radiation use efficiency under optimal growth conditions. *Agron. J.* **2005**, *97*, 72–78.
52. Sinclair, T.R.; Muchow, R.C. Radiation use efficiency. *Adv. Agron.* **1999**, *65*, 215–265.
53. Singer, J.W.; Meek, D.W.; Sauer, T.J.; Prueger, J.H.; Hatfield, J.L. Variability of light interception and radiation use efficiency in maize and soybean. *Field Crop. Res.* **2011**, *121*, 147–152.
54. DeLucia, E.; Drake, J.E.; Thomas, R.B.; Gonzalez-Meler, M. Forest carbon use efficiency: Is respiration a constant fraction of gross primary production? *Glob. Chang. Biol.* **2007**, *13*, 1157–1167.
55. Liang, S.; Zheng, T.; Liu, R.; Fang, H.; Tsay, S.-C.; Running, S. Estimation of incident photosynthetically active radiation from Moderate resolution imaging spectrometer data. *J. Geophys. Res.: Atmos.* **2006**, *111*, 1–13.

56. Prince, S.D.; Haskett, J.; Steininger, M.; Strand, H.; Wright, R. Net primary production of US Midwest croplands from agricultural harvest yield data. *Ecol. Appl.* **2001**, *11*, 1194–1205.

57. Kiniry, J.R.; Williams, J.R.; Vanderlip, R.L.; Atwood, J.D.; Reicosky, D.C.; Mulliken, J.; Cox, W.J.; Mascagni, H.J.; Hollinger, S.E.; Wiebold, W.J. Evaluation of two maize models for nine US locations. *Agron. J.* **1997**, *89*, 421–426.

58. Schapaugh, W.T.; Wilcox, J.R. Relationships between harvest indexes and other plant characteristics in soybeans. *Crop. Sci.* **1980**, *20*, 529–533.

59. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84.

60. Suyker, A.E.; Verma, S.B.; Burba, G.G.; Arkebauer, T.J. Gross primary production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season. *Agric. For. Meteorol.* **2005**, *131*, 180–190.