What is the value of agricultural census data in carbon cycle studies?

Eric C. Chan[1] and John C. Lin[1]

Received 30 November 2010; revised 28 March 2011; accepted 15 April 2011; published 29 July 2011.

[1] Agricultural census data have been identified as possessing the potential to provide constraints on modeled carbon uptake by croplands at the regional scale. In this study, we build on previous efforts and further assess this potential quantitatively by comparing (1) fractional cropland coverage in southwestern Ontario, Canada, derived from agricultural statistics against three different remotely sensed land cover maps and (2) carbon uptakes determined from agricultural data with simulations generated by a satellite-data-driven biospheric model. In addition, we assimilated the census-data-derived carbon uptakes with modeled estimates in a Bayesian inverse approach to determine if the crop data can provide constraint, as exhibited by uncertainty reductions, and if so, how much. Uncertainties in census-data-derived gross primary production (GPP) estimates are carefully quantified using a Monte Carlo simulation. In general, results from the fractional cropland coverage comparison indicate significant value of the agricultural census data by revealing biases in the spatial distribution of croplands, as found in all three of the satellite land cover products. However, we find that the carbon uptake values derived from crop harvested records are still subject to significant uncertainties that have been underestimated or neglected altogether in past studies. The Monte Carlo simulation suggests that the largest source of uncertainty can be traced to errors in the growth efficiency, followed by harvest production records, and then the harvest index. As a result, attention must be paid to such errors when using the agricultural census data for carbon accounting purposes or to provide constraints to simulations of crop carbon uptake.

Citation: Chan, E. C., and J. C. Lin (2011), What is the value of agricultural census data in carbon cycle studies?, *J. Geophys. Res.*, *116*, G03012, doi:10.1029/2010JG001617.

1. Introduction

[2] An increase in atmospheric greenhouse gas (GHG) concentrations would enhance the radiative energy absorbed by the Earth's surface and hence influence the global energy budget and climate [*Crowley*, 2000; *Karl and Trenberth*, 2003]. Among the various kinds of GHGs, carbon dioxide (CO_2) is the most significant in terms of radiative forcing [*Intergovernmental Panel on Climate Change*, 2007]. Consequently, understanding how CO_2 circulate on Earth becomes necessary. The global carbon cycle is a biogeochemical pathway characterized by the movement of carbons between the Earth's biospheric, oceanic, and atmospheric pools. An assessment of this pathway is therefore critical when we wish to determine which system is functioning as a source or sink of CO_2.

[3] Croplands play an important role in the terrestrial carbon cycle because they cover ~15 million km^2 of the planet [*Monfreda et al.*, 2008]. Naturally, organic carbons fixed by vegetation through photosynthesis and subsequent storage into soils can be regarded as a repository for atmospheric CO_2. In particular, strategically managed agricultural lands that favor the accumulation of carbons in soils would be considered as an option for mitigating GHG emissions [*Lal*, 2004; *Desjardins et al.*, 2005; *Kroodsma and Field*, 2006; *Hutchinson et al.*, 2007]. In view of that, carbon fluxes associated with croplands must be properly accounted for when quantifying carbon sources and sinks over the landscape.

[4] Biospheric models [*Potter et al.*, 1993; *Ruimy et al.*, 1996; *Mahadevan et al.*, 2008] play a central role in helping us to learn and quantify flows of terrestrial carbon because they can account for the biophysical processes that determine the spatiotemporal variations of carbon fluxes [*Cramer et al.*, 1999]. In the context of carbon cycle studies, there are various choices of data for constraining these biospheric model outputs. For example, inverse ("top down") analyses of CO_2 budgets on regional scales can utilize measurements of atmospheric CO_2 concentrations on tall towers or by aircraft within the regions where sources and sinks are most active [*Bakwin et al.*, 1998; *Gerbig et al.*, 2003a, 2003b; *Matross et al.*, 2006]. However, inverse methods still require work in

[1]Waterloo Atmosphere-Land Interactions Research Group, Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada.

0148-0227/11/2010JG001617

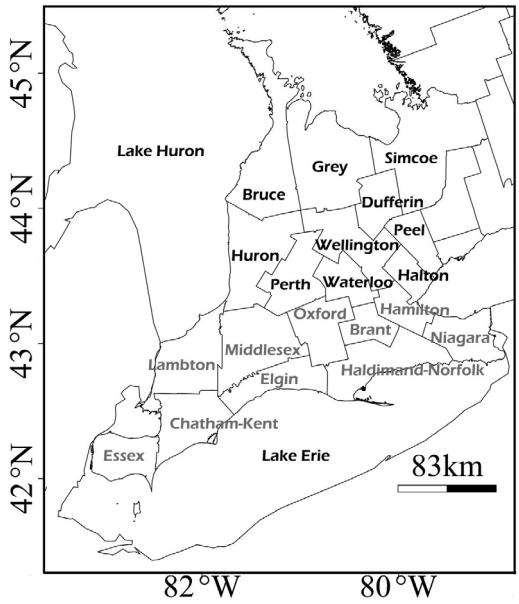

Figure 1. Map of census divisions in southwestern Ontario. Region 1 is southern Ontario (in gray); Region 2 is western Ontario (in black).

quantifying and understanding the error structure of both biospheric and atmospheric models [*Lin and Gerbig*, 2005; *Gerbig et al.*, 2006; *Lin et al.*, 2011]. Likewise, although eddy covariance measurements [*Baldocchi et al.*, 2001] offer details about the dynamics of terrestrial-atmospheric CO_2 exchange, their capability to estimate carbon sources and sinks at the regional scale is inadequate because the localized footprint ($\sim 10^1$ km^2 to 10^2 km^2) cannot be reliably scaled up to regional scales [*Jenkins et al.*, 2001; *Matross et al.*, 2006]. For scaling up flux estimates in space using satellite data, significant obstacles arise when field measurements are not available at the >1–10 km level scale for validation [*Lobell et al.*, 2002]. Agricultural census data can be regarded as a possible exception and thus poses as a valuable data set for constraining the biospheric models over the regional scale, as they provide two unique sources of information for carbon cycle science: (1) area covered by croplands at the >1–10 km scale and (2) a partial measure of total plant production throughout the growing season and hence potential carbon uptake.

[5] While numerous studies recognized and attempted to take advantage of the potential of agricultural census data, they have yet to do it in a comprehensive manner. They neither compared the census data against multiple satellite land cover data sets nor took into account the uncertainties of the census data. For example, *Ramankutty and Foley* [1998], *Frolking et al.* [1999], and *Hurtt et al.* [2001] compared cropland area estimates derived from ground-based census data against estimates created from one satellite-based land cover product developed using the 1 km Advanced Very High Resolution Radiometer (AVHRR). *Malmström et al.* [1997] and *Lobell et al.* [2002] used agricultural census data to test satellite-based biospheric models, but both studies only quantified interannual fit between census-data-derived estimates and satellite-based modeled fluxes by examining correlation coefficients. Neither study considered a priori uncertainties in either the agricultural data. When *Prince et al.* [2001] and *Bolinder et al.* [2007] converted agricultural data into primary production, the authors only took into account the error sources in the root:shoot ratio and harvest index and neglected uncertainties contributed by other parameters that are required in the conversion process. In this study, we intend to take a step beyond these previous efforts to highlight and examine the value and information content that agricultural census data bring to carbon cycle studies.

[6] The region of focus in this study is southwestern Ontario, Canada (Figure 1), because it is one of the major field crop producing regions in the country, accounting for 55.2% of the total national production of corn and 60% of winter wheat [*Statistics Canada*, 2006]. Given such significance, understanding how Ontario's croplands contribute to the regional carbon cycle is of great importance to resource managers and policy makers when managing risks and opportunities arising from climate change. For this study, we will use agricultural census data collected by the Canadian federal and provincial agencies to determine the fractional cropland coverage and primary production estimates. With the cropland fractions, we will compare them against three sets of satellite-based vegetation classification maps for verification. Then, in estimating the primary production, we will apply a Monte Carlo simulation to calculate their associated uncertainties and their sensitivity to the assumed parameters required for their estimation. These carbon uptake estimates and their uncertainties will later be assimilated with outputs generated by a biospheric model in a Bayesian inversion approach to obtain "optimal" posterior estimates (i.e., expected values of the posterior) for all subprovincial jurisdictions in southwestern Ontario. The reduction in uncertainties, presented as the percentage differences between prior and posterior uncertainties, will be analyzed to determine if/how much the agricultural data constrained the modeled fluxes.

2. Method and Data

2.1. Field Crop Statistics From Agricultural Census Data

[7] To determine cropland fraction and carbon uptake, annual harvested field crop area and yield production information of each county in southwestern Ontario are retrieved from the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) (http://www.omafra.gov.on.ca/english/stats/crops/index.html). According to OMAFRA, these data are compiled from telephone, mail-in, and enumerative surveys of farmers, with additional information supplied by government field officers, agribusiness personnel, and farm marketing boards.

[8] For comparing fractional cropland coverage, we used harvested area estimates reported for all field crops listed on the OMAFRA field crop statistics Web page. However, when determining crop GPP, we only focused on the dominant field crops such as winter wheat, spring wheat, grain corn, fodder corn, soy, hay and barley because relevant data sets such as moisture content, harvest index and

root:shoot ratio of other field crops required for the conversion process (equation (6)) are not available. Nevertheless, the selected field crops represent over 90% of the total harvested area and over 95% of the total harvested production among all reported field crops in Ontario during 2004 [*Statistics Canada*, 2004].

[9] Data for uncertainty analysis (section 2.5) are extracted from other Canadian government publications like the Field Crop Reporting Series [*Statistics Canada*, 2004] and the Census of Agriculture [*Statistics Canada*, 2006], because these records included essential data accuracy evaluations that are needed to determine subprovincial level uncertainties. Every 5 years, the Census of Agriculture collects information on agricultural operations across Canada, including institutional farms, community pastures, Indian reserves, etc. The Census of Agriculture provides a list of farms and their crop areas from which a probability sample for production estimates is selected [*Statistics Canada*, 2004].

[10] County-level uncertainties associated with harvested area and production are calculated through downscaling the corresponding values reported at larger (higher level) jurisdictions (section 2.5). For uncertainties associated with crop production, we used the Canadian national level uncertainty reported in the November 2004 issue of the Field Crop Reporting Series [*Statistics Canada*, 2004]. For uncertainties related to harvested area we used the provincial level uncertainty reported in the Census of Agriculture 2006 [*Statistics Canada*, 2006]. In the latter case, we assumed no significant differences in data quality between 2004 and 2006.

2.2. Satellite-Based Land Cover Maps and Cropland Fractional Coverage

[11] Cropland area derived from the agricultural census data is used to evaluate the performance of three satellite-based land cover products: (1) a reclassified International Geosphere-Biosphere Programme Data and Information System land cover (IGBP-DISCover) map [*Loveland et al.*, 2000; *Mahadevan et al.*, 2008], (2) the SYNMAP [*Jung et al.*, 2006], and (3) the Ontario Land Cover (OLC) database [*Ontario Ministry of Natural Resources*, 2002]. In particular, two of the data sets (IGBP-DISCover and SYNMAP) are commonly used in terrestrial carbon modeling studies. A brief description for each product is provided below, and readers are encouraged to refer to the original publications for further details.

2.2.1. IGBP-DISCover Data Set

[12] The original IGBP-DISCover data set, consisting of 17 classes, was developed from the monthly global normalized difference vegetation index (NDVI) composites taken from the 1 km AVHRR data covering 1992–1993 [*Loveland et al.*, 2000]. During the creation of a new satellite-based biospheric model for estimating terrestrial carbon fluxes, *Mahadevan et al.* [2008] reclassified the original data set into 12 classes because the calibration and validation data for that model were not available for each of the 17 classes. Accordingly, an important feature with the *Mahadevan et al.* [2008] land cover map was the lumping of the "cropland/natural vegetation mosaic" class into the "grasslands" class. In this study, the reclassified IGBP-DISCover map is used because we use the *Mahadevan et al.* [2008] model (VPRM) in another section. For brevity's sake we would refer to this land cover data set as "IGBP" from this point on.

2.2.2. SYNMAP

[13] SYNMAP is an enhanced 1 km land cover data set that was developed through synthesizing the Global Land Cover Characterization Database (GLCC), the Global Land Cover 2000 (GLC2000) and the MODIS land cover product [*Jung et al.*, 2006]. According to *Jung et al.* [2006], SYNMAP is a desirable data set because it should be more accurate than existing global land cover products since it blends and makes use of their individual strengths to achieve a possibly better signal-to-noise ratio. Owing to its possible better performance over IGBP, we included SYNMAP in this study.

2.2.3. Ontario Land Cover Database

[14] Finally, the Ontario Land Cover (OLC) database was initiated by the Ontario Ministry of Natural Resources (OMNR) and was derived from digital, multispectral LANDSAT Thematic Mapper (TM) data recorded on a range of dates between 1986 and 1997. It consists of 28 land cover classes mapped across the entire province of Ontario [*Ontario Ministry of Natural Resources*, 2002]. We included it in this study because it has a 30 m resolution, much higher than that of the previous two products, and should provide the best land cover description of Ontario.

[15] To calculate cropland fractions with the census data, we first took the harvested areas from the census records and directly mapped them onto a 1/96° by 1/96° spatial resolution grid, in which cell areas are determined by using the great circle formulation. Then, cropland fractions are simply calculated through dividing the harvested areas by the grid cell areas. In the end, we compared the various fractional cropland coverage estimates by aggregating all maps into a spatial resolution of 1/6° (latitude) by 1/4° (longitude). The census-data-derived cropland fractions map becomes the benchmark to evaluate the performance of the satellite products.

2.3. Calculating GPP From Crop Attributes and Agricultural Census Data

[16] Several studies [*Prince et al.*, 2001; *Lobell et al.*, 2002; *Hicke and Lobell*, 2004; *Huang et al.*, 2007; *Bolinder et al.*, 2007] have used agricultural census data and crop attributes to study the spatial and temporal variations of primary production in croplands. In particular, an approach followed by *Prince et al.* [2001] and *Hicke and Lobell* [2004] translated harvested production and area to net primary production (NPP):

$$\mathrm{NPP} = \sum_{i=1}^{N} \frac{P_i \times (1 - \mathrm{MC}_i) \times C}{\mathrm{HI}_i \times f_{\mathrm{AG}i} \times \mathrm{HA}_i}, \tag{1}$$

where index i denotes different crops. P_i represents the harvested production (g) and HA_i is the harvested area (m^2). The crop attributes are as follows: MC_i is the typical harvest moisture content (mass water/mass harvest; g g^{-1}); HI_i is the harvest index which specifies the ratio of yield mass to aboveground biomass (unitless); C converts harvested mass to carbon mass (approximately 0.45 g C g^{-1}) and is assumed constant between different crops; $f_{\mathrm{AG}i}$ is the fraction of

Table 1. Crop-Specific Attributes Used for Calculating Gross Primary Production (GPP) Estimates and Their Associated Uncertainties

Crop	Attributes	Values	Source
Winter wheat	Moisture content (%)	13–23	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.32–0.6	*de Jong et al.* [2001], *Falk et al.* [2007]
	f_{AG}	0.82–0.87	*Bolinder et al.* [2007]
	Growth efficiency	0.37–0.79	*Amthor* [1989]
Spring wheat	Moisture content (%)	13–23	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.32–0.6	*de Jong et al.* [2001], *Falk et al.* [2007]
	f_{AG}	0.82–0.87	*Bolinder et al.* [2007]
	Growth efficiency	0.37–0.79	*Amthor* [1989]
Grain corn	Moisture content (%)	11–22	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.45–0.55	*O'Neill* [2005], *Bolinder et al.* [2007]
	f_{AG}	0.89–0.92	*Bolinder et al.* [2007]
	Growth efficiency	0.31–0.83	*Amthor* [1989]
Fodder corn	Moisture content (%)	65–70	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.55–1	*O'Neill* [2005], *Bolinder et al.* [2007]
	f_{AG}	0.89–0.92	*Bolinder et al.* [2007]
	Growth efficiency	0.31–0.83	*Amthor* [1989]
Barley	Moisture content (%)	15–20	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.32–0.54	*Bolinder et al.* [2007]
	f_{AG}	0.63–0.69	*Bolinder et al.* [2007]
	Growth efficiency	0.51–0.52	*Amthor* [1989]
Soy	Moisture content (%)	14	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	0.23–0.4	*Morrison et al.* [1999], *Rollefson et al.* [2004], *Bolinder et al.* [2007]
	f_{AG}	0.68–0.89	*Bolinder et al.* [2007]
	Growth efficiency	0.25–0.55	*Amthor* [1989]
Hay	Moisture content (%)	12–18	*Ontario Ministry of Agriculture, Food and Rural Affairs* [2009]
	Harvest index	1	*Hicke and Lobell* [2004]
	f_{AG}	0.52–0.77	*Bolinder et al.* [2007]
	Growth efficiency	0.51–0.67	*Amthor* [1989]

production allocated aboveground (unitless). Since f_{AGi} is defined as the ratio shoot:(root + shoot), it is closely related to the root:shoot ratio:

$$f_{AG} = \frac{\text{shoot}}{\text{root} + \text{shoot}} = \frac{1}{(\text{root} : \text{shoot} + 1)}. \quad (2)$$

By definition, NPP refers to the rate at which primary producers capture and accumulate phytomass, minus the rate at which they respire for growth and maintenance (autotrophic respiration, R_a):

$$\text{NPP} = \text{GPP} - R_a. \quad (3)$$

Because VPRM, the biospheric model used in this study, partitions net ecosystem exchange between gross ecosystem exchange (GEE) and total respiration (autotrophic + heterotrophic; section 2.4), when we compare the agricultural census-derived production estimates against simulated values we must take an extra step to convert NPP into GPP.

[17] Growth efficiency (GE) is a parameter that quantifies the role of respiration to a crop's carbon balance by accounting for the loss of CO_2 during crop growth and maintenance [*Amthor*, 1989]:

$$\text{GE} = 1 - \frac{R_a}{\text{GPP}}. \quad (4)$$

By combining equations (3) and (4), GPP can be expressed as a function of NPP and GE:

$$\text{GPP} = \frac{\text{NPP}}{[1 - (R_a/\text{GPP})]} = \frac{\text{NPP}}{\text{GE}}. \quad (5)$$

Combining equations (5) and (1), we have

$$\text{GPP} = \sum_{i=1}^{N} \frac{P_i \times (1 - \text{MC}_i) \times C}{\text{HI}_i \times f_{AGi} \times \text{HA}_i \times \text{GE}_i}. \quad (6)$$

The agricultural census data provide input for P and HA, and equation (6) serves as the basis on which to quantify GPP. The values of crop attributes (MC, HI, f_{AG}, and GE) are taken from the literature for crops grown in Canada and are summarized in Table 1.

[18] VPRM simulates hourly values of gross ecosystem exchange (GEE), details of which are given in section 2.4. GPP estimates derived from equation (6) can thus be compared against the model by summing GEE over an entire year:

$$\begin{aligned} \text{GPP} &= \text{NPP} + R_a \\ &\approx \sum^{1\text{year}} \text{GEE}_t \end{aligned}. \quad (7)$$

2.4. Vegetation Photosynthesis and Respiration Model

[19] The Vegetation Photosynthesis and Respiration Model (VPRM) is a data-driven, diagnostic biospheric carbon flux model developed for regional to global-scale inverse analysis [*Mahadevan et al.*, 2008]. It assimilates satellite information with meteorological data and eddy flux measurements to model variations in atmosphere-terrestrial biosphere carbon flux exchange. VPRM is chosen in this study because its simple structure required less parameterizations and parameters that vary with space and time,

making it a straightforward yet useful tool for upscaling and regional carbon prediction [*Mahadevan et al.*, 2008].

[20] VPRM calculates the net ecosystem exchange (NEE) for 12 main vegetation classes as the difference between GEE and total ecosystem respiration (R): NEE = −GEE + R. Note that R here includes both the autotrophic and heterotrophic components. Following sign convention, negative fluxes represent the removal of CO_2 from the atmosphere by vegetation [*Matross et al.*, 2006]. GEE is assumed to be a function of the MODIS Enhanced Vegetation Index (EVI):

$$\mathrm{GEE} = \lambda \times T_{\mathrm{scale}} \times P_{\mathrm{scale}} \times W_{\mathrm{scale}} \times \frac{1}{\left[1 + \left(\mathrm{PAR}/\mathrm{PAR}_0\right)\right]} \times \mathrm{PAR} \times \mathrm{EVI}, \tag{8}$$

where PAR is the photosynthetically active radiation; PAR_0 is the half-saturation value; λ represents the overall light use efficiency of the ecosystem; T_{scale}, P_{scale} and W_{scale} are scalars ranging in value between 0 and 1 that signify the effect of temperature, leaf phenology and canopy water content, respectively.

[21] R is represented as a linear function of the ambient temperature given by

$$R = \alpha \times T + \beta, \tag{9}$$

where α captures the dependence of R on air temperature when air temperatures are above a minimum temperature T_{min}; β represents the basal respiration rate; $T = T_{\mathrm{low}}$ when $T \leq T_{\mathrm{low}}$ to account for the persistence of soil respiration in winter when air temperatures are very cold but soils remain warm.

[22] The resulting full VPRM model equation is

$$\mathrm{NEE} = -\lambda \times T_{\mathrm{scale}} \times P_{\mathrm{scale}} \times W_{\mathrm{scale}} \times \frac{1}{\left[1 + \left(\mathrm{PAR}/\mathrm{PAR}_0\right)\right]} \times \mathrm{PAR} \times \mathrm{EVI} + \alpha \times T + \beta. \tag{10}$$

The a priori estimates of the four calibration parameters λ, PAR_0, α and β, one set per vegetation type, were calibrated by optimizing against un-gap-filled eddy covariance NEE measurements taken from their 11 corresponding AmeriFlux and Fluxnet-Canada sites, with a turbulent intensity filter applied to eliminate unrepresentative observations. Specifically, they were optimized via nonlinear least squares (Newton-Raphson, tangent linear approximation) and estimated confidence intervals assuming Gaussian error for both model and tower data [*Mahadevan et al.*, 2008]. At each calibration site, *Mahadevan et al.* [2008] generated hourly data from the smoothed time series of vegetation indices (EVI and LSWI) and obtained measurements of air temperature and PAR from the tower sites. For calibration results of the model parameters, please refer to *Mahadevan et al.* [2008, Table 2]. For validation results, please refer to *Mahadevan et al.* [2008, Table 3]. In general, cropland parameters were calibrated at a maize-soybean agroecosystem from Mead, Nebraska and modeled NEE estimates were validated at a maize-soybean cropland from Champaign, Illinois [*Mahadevan et al.*, 2008].

[23] NEE, GEE, and R estimates generated by the VPRM are gridded on a grid of 1/4° longitude by 1/6° latitude for every hour [*Matross et al.*, 2006]. Sub-grid-scale contributions from different vegetation types are preserved as grid-scale carbon fluxes by weighing contributions from various vegetation types k:

$$\mathrm{NEE} = -\mathrm{GEE} + R = \sum_k f_k(-\mathrm{GEE}_k + R_k), \tag{11}$$

where f_k is the fractional areal coverage by vegetation type k and is currently determined from the IGBP 1 km resolution land cover scheme (section 2.2). In this study, we extracted carbon fluxes solely from croplands in VPRM and ignored flux contributions from other vegetation types by considering croplands as the only land cover input.

2.5. Uncertainty Analysis of GPP Estimated From Crop Attributes and Agricultural Census Data

[24] In calculating carbon uptake in crops from agricultural census data and crop attributes with equation (6), we adopted a Monte Carlo approach to assess the GPPs' associated uncertainties and their sensitivity to the assumed parameters in equation (6), as recommended by the IPCC Good Practice Guidance [*Intergovernmental Panel on Climate Change*, 2000]. For each parameter we assigned a probability density function (PDF) from which 10,000 random values are selected and used to calculate the carbon uptakes for each county in southwestern Ontario. The underlying PDF was chosen depending on our prior knowledge: if we can identify a range of probable values for a parameter but cannot decide which value is most likely to occur, a uniform distribution was assumed to maximize information entropy (i.e., minimal prior knowledge) [*Jaynes*, 1968]. As such, HI, f_{AG}, MC and GE are assigned uniform distributions because we do not have adequate information on the standard deviations associated with the crop-specific attributes to appropriately generate Gaussian random values. Table 1 summarizes the ranges of values for the uniform distributions for each crop-specific attribute.

[25] Eventually, after generating 10,000 random values for each parameter and calculating the GPP estimates, an average is determined along with the standard deviation that quantifies the associated uncertainty.

[26] For HA and P, their uncertainties are associated with errors in gathering the agricultural census data. Since we do not have uncertainty estimates at the county level, our only option is to scale the associated variance of the provincial estimates down to the county levels through the following steps. To begin, we considered that the harvested area (HA) for county i is a multiple of an arbitrary unit harvested area (a):

$$\mathrm{HA}_i = n_i a, \tag{12}$$

where n_i is the number of unit harvested areas in a particular county. Applying the same logic, the total provincial harvested area ($\mathrm{HA}_{\mathrm{prov}}$) can be considered as a sum of numerous county harvested areas:

$$\mathrm{HA}_{\mathrm{prov}} = \sum_i \mathrm{HA}_i = \sum_i n_i a = Na, \tag{13}$$

where the total provincial harvested area is composed of N units. Since a larger agricultural area is more difficult for the census to capture exhaustively, it is reasonable to expect

that the estimation error increases with the area. If the uncertainty-induced variance for a unit area can be characterized statistically by σ^2, and the uncertainties from unit to unit are statistically independent, then the individual variances would simply add [*Taylor*, 1997], and the resultant variance for HA_i would be

$$\sigma^2_{\mathrm{HA}_i} = n_i\sigma^2, \tag{14}$$

while the variance at the provincial level would be

$$\sigma^2_{HA_{prov}} = \sum_i \sigma^2_{HA_i} = \sum_i n_i\sigma^2 = N\sigma^2. \tag{15}$$

Combining equations (14) and (15), we have

$$\sigma^2_{\mathrm{HA}_i} = \frac{n_i}{N}\sigma^2_{\mathrm{HA}_{\mathrm{prov}}}, \tag{16a}$$

$$\sigma_{\mathrm{HA}_i} = \left(\sqrt{\frac{n_i}{N}}\right)\sigma_{\mathrm{HA}_{\mathrm{prov}}}. \tag{16b}$$

Thus, the uncertainty in harvested area at the county level is simply the uncertainty at the provincial level multiplied by the square root of the ratio in the areas. For the value of $\sigma_{\mathrm{HA}_{\mathrm{prov}}}$, we used the uncertainty estimate reported in the Census of Agriculture [*Statistics Canada*, 2006], which stated the standard error of farmland area in Ontario as 0.3%.

[27] For harvested production P, we first assumed that the provincial level uncertainty, reported as the coefficient of variation (CV), is equivalent to the national level uncertainty reported in the November 2004 issue of Field Crop Reporting Series [*Statistics Canada*, 2004]. According to Statistics Canada, the potential error introduced by probabilistic sampling for crop production estimates was given by the CV, which ranged from 1% to 5% for the major crops at the national level [*Statistics Canada*, 2004]. Since the CV is calculated by $\sigma_{\mathrm{Prod}_{\mathrm{prov}}}/\bar{x}_{\mathrm{Prod}_{\mathrm{prov}}}$, where $\bar{x}_{\mathrm{Prod}_{\mathrm{prov}}}$ is taken to be the provincial crop production, the uncertainty (standard deviation) of the county level crop production is therefore

$$\sigma_{\mathrm{Prod}_i} = \left(\sqrt{\frac{n_i}{N_{\mathrm{natl}}}}\right)\sigma_{\mathrm{Prod}_{\mathrm{prov}}}. \tag{17}$$

where $\sigma_{\mathrm{Prod}_{\mathrm{prov}}} = \mathrm{CV} \times \bar{x}_{\mathrm{Prod}_{\mathrm{prov}}}$. In this study, an average of 3% error was used, and equation (17) follows the same logic that led to equation (16b) above.

2.6. Bayesian Inversion to Optimize GPP Estimates

[28] In order to acquire insights on how the agricultural census-data-derived GPP improves knowledge about the VPRM-modeled GPP, we performed a Bayesian optimization [*Rodgers*, 2000]. Census-data-derived GPP can be related linearly to the VPRM-modeled GPP through

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \varepsilon, \tag{18}$$

where $\mathbf{y}$ is the vector of census-data-derived GPP values, with each entry representing a value for each census division (Figure 1); $\mathbf{x}$ is the vector of VPRM GPP, with each entry signifying the modeled GPP estimate in each 1/6° by 1/4° grid cell; and $\mathbf{K}$ is the matrix that relates the census-data-based GPP to VPRM's GPP. Each entry in $\mathbf{K}$ denotes the fraction of a census division covered by each 1/6° by 1/4° grid cell; ε is an error vector accounting for uncertainties in both the census-based estimates and VPRM simulations. The optimal posterior GPP estimates are then obtained by minimizing the cost function J using a standard least squares formulation [*Rodgers*, 2000; *Raupach et al.*, 2005]:

$$J(\mathbf{x}) = (\mathbf{y} - \mathbf{K}\mathbf{x})^{\mathrm{T}}\mathbf{S}_{\varepsilon}^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}) + \left(\mathbf{x} - \mathbf{x}_{\mathrm{prior}}\right)^{\mathrm{T}}\mathbf{S}_{\mathrm{prior}}^{-1}\left(\mathbf{x} - \mathbf{x}_{\mathrm{prior}}\right), \tag{19}$$

where $\mathbf{S}_{\varepsilon}$ is the error covariance matrix of inventory-based estimates with diagonals corresponding to variances calculated from the Monte Carlo simulation for each census division (section 2.5, Table 4); $\mathbf{x}_{prior}$ refers to the a priori GPP estimates simulated by the VPRM; and $\mathbf{S}_{prior}$ is the associated error covariance matrix of the VPRM estimates in which diagonal elements are derived by multiplying average percentage uncertainties between modeled and observed fluxes for soy and corn reported by *Mahadevan et al.* [2008, Table 3] and the VPRM-generated primary productions in model grid cells. Off-diagonal elements of both error covariance matrices are assumed to be zero.

[29] The expected values of the posterior "optimal" GPP estimates expressed as $\mathbf{x}_{\mathrm{post}}$ is calculated by

$$\mathbf{x}_{\mathrm{post}} = \left(\mathbf{K}^T\mathbf{S}_{\varepsilon}^{-1}\mathbf{K} + \mathbf{S}_{\mathrm{prior}}^{-1}\right)^{-1}\left(\mathbf{K}^T\mathbf{S}_{\varepsilon}^{-1}\mathbf{y} + \mathbf{S}_{\mathrm{prior}}^{-1}\mathbf{x}\right) \tag{20}$$

and accordingly, the uncertainty of $\mathbf{x}_{\mathrm{post}}$, expressed as the posterior error covariance matrix $\mathbf{S}_{\mathrm{post}}$, is given by

$$\mathbf{S}_{\mathrm{post}} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\varepsilon}^{-1}\mathbf{K} + \mathbf{S}_{\mathrm{prior}}^{-1}\right)^{-1}. \tag{21}$$

3. Results and Discussion

3.1. Comparison of Cropland Fractions Against Satellite Land Cover Maps

[30] Figure 2a displays the cropland distributions in southwestern Ontario as described by the agricultural census data. Uncertainties associated with these cropland fractions (σ_{Frac} in Table 2) are calculated by dividing the σ_{HA} (section 2.5) by the total area of each county. From the values of σ_{Frac}, it is evident that the uncertainties in the cropland fractions derived from the agricultural census data are small. According to OMAFRA, counties including Chatham-Kent, Essex, Elgin, Middlesex, Lambton and Oxford of southern Ontario together with Huron, Perth and Waterloo of western Ontario can be regarded as agriculturally productive, since cropland fractions within these counties are at least 0.5 or above (Table 2).

[31] However, this is not the case in the IGBP "croplands" map (Figure 2b), which shows severe underestimation in cropland areas in all of western Ontario and most of southern Ontario, except for Chatham-Kent, where a small overestimation is observed. Overall, the IGBP "croplands" map has an average of 70% underestimation in cropland fractions (Table 2). This outcome is likely due to the fact that "cropland/natural vegetation mosaic" was combined into the "grasslands" class by *Mahadevan et al.* [2008] rather than the "croplands" class. In order to examine how

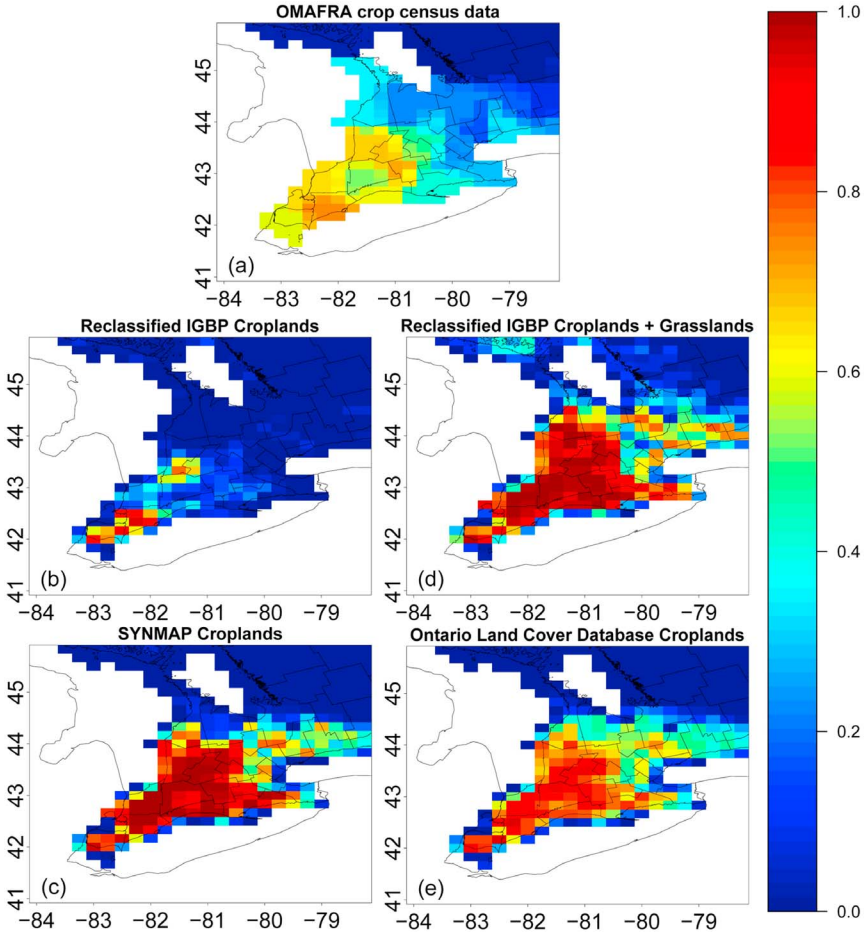

Figure 2. Fractional coverage of croplands and other vegetation derived from various data sets: (a) agricultural census data from Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA); (b) International Geosphere-Biosphere Programme Data and Information System land cover (IGBP-DISCover) "croplands"; (c) SYNMAP "croplands"; (d) IGBP-DISCover "croplands+grasslands" (which includes "cropland/natural vegetation mosaic"); and (e) Ontario Land Cover Database "croplands."

Mahadevan et al.'s [2008] reclassification scheme affected land cover representation (section 2.2.1), we merged the modified IGBP "croplands" class with the "grasslands" class to construct a new "croplands+grasslands" map. This new map (Figure 2d), when compared to the IGBP "croplands" map, seems to better capture the extent of cropland areas in the south and parts of the west but still significantly overestimates cropland fractions, by 76%, on average (Table 2).

[32] SYNMAP's "croplands" map (Figure 2c) shows an improved performance over that of IGBP as it is capable of displaying which counties in the region are relatively more cultivated. Nevertheless, significant overestimations of cropland fractions also exist in this land cover map, on average by 62%. Finally, although the OLC database (Figure 2e) gives the most comparable pattern of cropland distribution among all the examined satellite land cover data sets, it too overestimates cropland fractions by 50%.

[33] Discrepancies between the satellite-based land cover maps and the census data can be attributed to numerous sources. For the IGBP "croplands" map, disagreements are likely a result of losing "croplands" pixels when the "cropland/natural vegetation mosaic" class, the least reliable land cover category [*Jung et al.*, 2006], was originally incorporated within the "grassland" class. Noises caused by the presence of aerosols and clouds may also interfere with the data retrieval procedure. *Loveland et al.* [2000] identified IGBP North American croplands to be highly affected by "noise contamination," with dryland croplands and irrigated crop fields having ~27% and ~42% such contamination. Likewise, *Vogelmann et al.* [2001] proposed that hardware calibration errors might lead to the degradation of radiometric accuracy of the LANDSAT Thematic Mapper (TM) satellite data and cause the OLC database, which was derived from the LANDSAT data covering dates between 1986 and 1997, to have an accuracy of approximately 85% for agricultural lands [*Ontario Ministry of Natural Resources*, 2002]. The use of obsolete retrieved data might also lead to deviations in the fraction estimates as remotely sensed data acquired at different times may not correctly present the needed land cover status [*Jung et al.*, 2006].

[34] Further, in heterogeneous landscapes where croplands are highly intermixed with other vegetation types such as grasslands or scrublands, the target pixel might not be correctly classified [*Wulder et al.*, 2004; *Jung et al.*, 2006;

Table 2. Comparison of Cropland Fractions Derived From Different Sources[a]

	OMAFRA		IGBP Croplands		IGBP Croplands+Grasslands		SYNMAP		Ontario Land Cover Database	
Counties	Cropland Fractions	σ_{Frac}	Cropland Fractions	%Δ From Census Data	Cropland Fractions	%Δ From Census Data	Cropland Fractions	%Δ From Census Data	Cropland Fractions	%Δ From Census Data
BRAN	0.45	1.01E-02	0.17	−62	0.9	100	0.93	107	0.72	60
CHAT	0.72	7.58E-03	0.75	4	0.83	15	0.83	15	0.78	8
ELGI	0.59	8.83E-03	0.26	−56	0.89	51	0.88	49	0.73	24
ESSE	0.6	7.88E-03	0.6	0	0.73	22	0.72	20	0.7	17
HALD	0.42	6.20E-03	0.09	−79	0.82	95	0.81	93	0.64	52
HAMI	0.3	7.80E-03	0.07	−77	0.66	120	0.69	130	0.56	87
LAMB	0.65	6.94E-03	0.29	−55	0.83	28	0.86	32	0.75	15
MIDD	0.55	6.33E-03	0.23	−58	0.97	76	0.96	75	0.82	49
NIAG	0.29	6.41E-03	0.01	−97	0.68	134	0.58	100	0.47	62
OXFO	0.67	9.43E-03	0.15	−78	0.99	48	0.98	46	0.85	27
BRUC	0.35	4.68E-03	0	−100	0.62	77	0.46	31	0.45	29
DUFF	0.33	7.22E-03	0.02	−94	0.5	52	0.59	79	0.57	73
GREY	0.25	3.26E-03	0	−100	0.67	168	0.26	4	0.46	84
HALT	0.26	7.75E-03	0.01	−96	0.54	108	0.57	119	0.5	92
HURO	0.64	7.06E-03	0.25	−61	0.88	38	0.9	41	0.78	22
PEEL	0.25	7.18E-03	0.01	−96	0.45	80	0.39	56	0.46	84
PERT	0.66	8.89E-03	0.38	−42	0.97	47	0.99	50	0.89	35
SIMC	0.21	3.27E-03	0.01	−95	0.46	119	0.32	52	0.39	86
WATE	0.53	9.76E-03	0.1	−81	0.91	72	0.87	64	0.73	38
WELL	0.47	6.76E-03	0.05	−89	0.8	70	0.84	79	0.69	47
Average	0.46	7.17E-03	0.17	−70.60	0.76	76.00	0.72	62.10	0.65	49.55

[a]The first four letters of the counties' names, as indicated in Figure 1, are used as abbreviations. Value σ_{Frac} is the uncertainty in cropland fractions (section 3.1).

Herold et al., 2008; *Wu et al.*, 2008]. *Loveland and Belward* [1998, p. 689] pointed out that in the IGBP data set, a pixel would be classified as "cropland/natural vegetation mosaic" when "no [single vegetation] component comprises more than 60% of the landscape." As a result, some croplands may be disregarded and thus give an underestimated total cropland area. Such inability to correctly classify a target pixel that contains several vegetation types may be viewed as a general problem of optical remote sensing [*Jung et al.*, 2006], and accordingly, we can expect higher accuracies for areas in with homogeneous cropland coverage and lower accuracies in areas with sparse patches of cropland [*Jung et al.*, 2006; *Wu et al.*, 2008]. Indeed, this phenomenon is evident in the current study, as the magnitude of the relative error in the cropland fraction decreases as the cropland fraction increases for all four satellite land cover maps examined here (Figure 3).

3.2. Monte Carlo Uncertainty Analysis for Agricultural Census-Derived GPP Estimates

[35] Several sources of uncertainties are present when estimating GPP from agricultural census data and crop attributes. Figure 4 illustrates results from a Monte Carlo analysis. In this analysis, we randomized one attribute at a time to evaluate the sensitivity of the estimated GPP to uncertainties in the individual variables. Finally, all of the variables are randomized to examine the aggregate effect of their errors. The average error resulting from all of the variables was ~20% of the GPP value (Table 4). As seen in Figure 4, the growth efficiency (GE), harvested production (P), and harvest index (HI), in decreasing order of significance, are the three largest sources of uncertainties in estimating GPP. Here we discuss each in turn.

[36] *Amthor* [1989] pointed out that estimating GE requires accurate but difficult measurements of respiration and photosynthesis (or changes in dry phytomass). For instance, measuring GE on the basis of ^{14}C labeling is dependent on the time of day the labeling takes place because crops would refixate respired CO_2 depending on time. In addition, the crop's stage of development matters as well [*Amthor*, 1989, p. 114]: "Labeling with ^{14}C later in the development of a crop will tend to underestimate the actual loss of carbon to respiration over the course of a growing season while labeling early in the season will tend to overestimate total respiratory losses." Under such limitations, large discrepancies in GE values can therefore be expected. When randomizing GE alone in equation (6), the average uncertainty in GPP across all counties is 127 g C m^{-2} yr^{-1}.

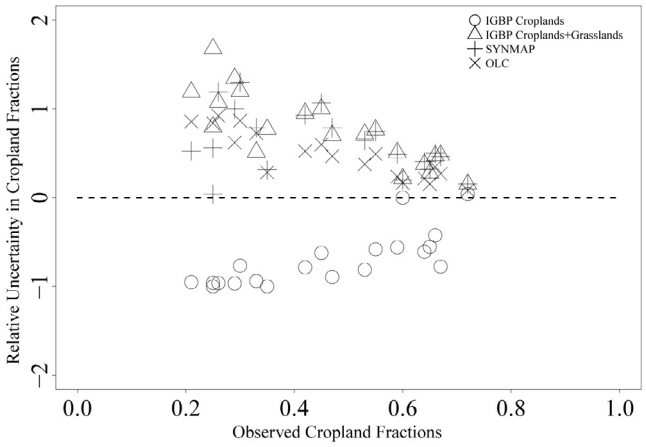

Figure 3. Comparison of relative uncertainties in cropland fractions. Relative uncertainties per county are calculated by dividing the cropland fractions' differences (i.e., satellite-derived fractions minus census-data-derived fractions) by the census-data-derived fractions.

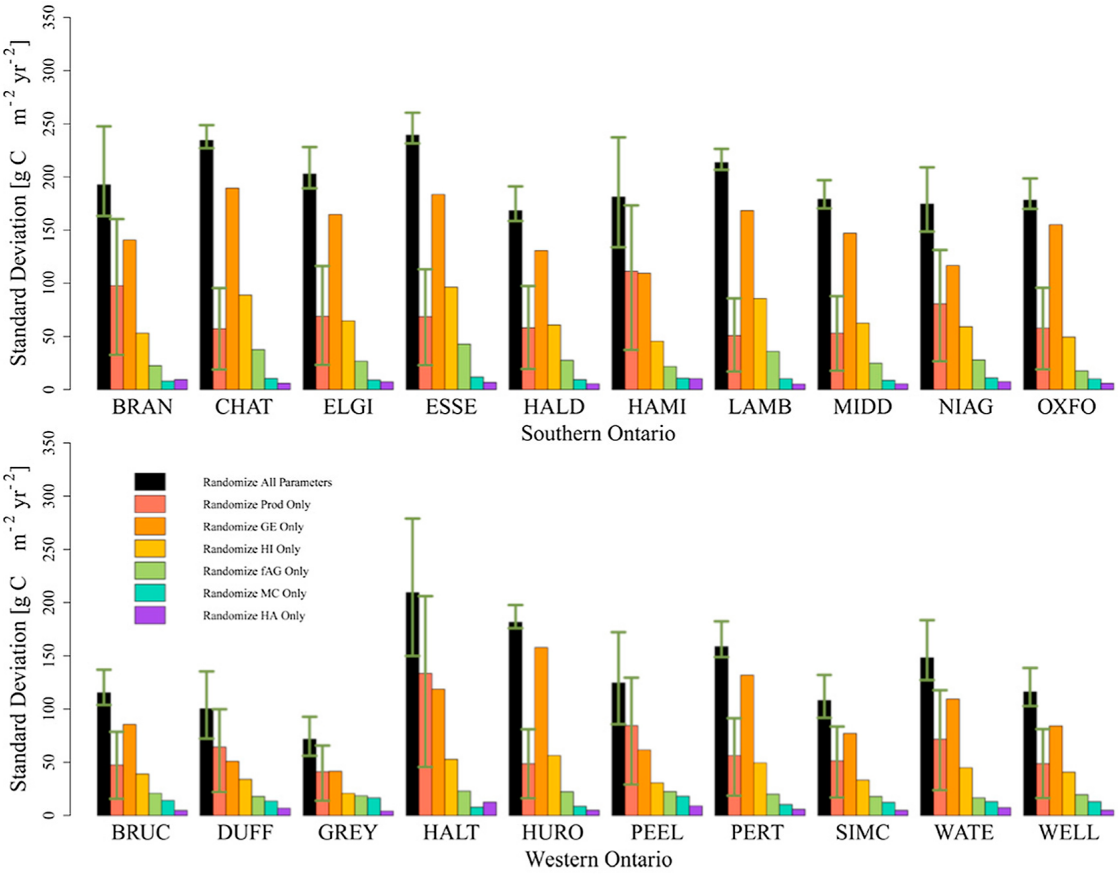

Figure 4. Results of Monte Carlo analysis of uncertainties in census-data-derived GPP for southwestern Ontario. Different colors illustrate how randomizing different parameters in the crop-attribute-based model influences the uncertainties in GPP estimates. Green error bars explain how uncertainties in GPP would change when only the uncertainty in provincial harvested production changes from 1% to 5% discretely. The lower limit of the green error bar illustrates the associated uncertainty when standard errors were at 1%; the upper limit of the green error bar illustrates the associated uncertainty when standard errors were at 5%. Prod, harvested production; GE, growth efficiency; HI, harvest index; fAG, fraction allocated above ground; MC, moisture content; and HA, harvested area. The first four letters of the counties' names, as indicated in Figure 1, are used as abbreviations.

[37] As for *P*, potential errors are introduced by probabilistic sampling and are described by the coefficient of variation (CV), which ranges from 1% to 5% for all major crops at the national level [*Statistics Canada*, 2004]. In this study, when randomizing *P* alone while adopting CV = 3% (halfway between 1% and 5%), the mean uncertainty in GPP is 70 g C m^{-2} yr^{-1}.

[38] By definition, HI is equal to the seed yield divided by aboveground biological yield [*Donald and Hamblin*, 1976], where biological yield "includes the total dry matter the plant produces above ground" [*Prince et al.*, 2001, p. 1196]. Physiologically, variability in HI depends on several factors such as cultivar and the geographical features surrounding the growing crop [*Hay*, 1995; *Hay and Gilbert*, 2001; *Prince et al.*, 2001; *de Jong et al.*, 2001; *Falk et al.*, 2007; *Bolinder et al.*, 2007]. For instance, small changes in HI can be seen during stressed conditions as decreases in yield are generally accompanied by reductions in crop biomass [*Prince et al.*, 2001]. When randomizing HI alone, the average uncertainty in GPP over all counties is 54 g C m^{-2} yr^{-1}.

3.3. Comparison of Estimated GPP

[39] In general, GPP estimates determined from both the agricultural census data and VPRM fall into the lower end of the range of results produced from other studies (Table 3).

[40] Comparisons between the agricultural-data-derived GPP and VPRM-modeled GPP are shown in Table 4 and Figure 5. In Figure 5a, the small Pearson correlation coefficient relating the two sets of GPP values signifies some capability for the VPRM to model agricultural GPP in southwestern Ontario. A few probable explanations for the overestimation of GPP by VPRM include the following: (1) Model parameters such as λ, PAR_0, α, and β may not be applicable to Ontario croplands because they are calibrated among crop fields (soy and corn) in the American Midwest. (2) The current version of land cover data set used by the model is not representative of the landscape. (3) The types of crops used in calibrating VPRM parameters may not fully resemble those grown on Ontario's croplands.

Table 3. Reported GPP Estimates in Oilseeds and Grain Crops From This Study and From Previously Published Comparable Studies

Source	Study Location	Study Period	Crop Types	Approach Applied	Average GPP (g C m^{-2} yr^{-1})
			This Study		
	Ontario, Canada[a]	2004	Grain corn, hay, soybean, winter wheat	Integrated surface flux modeling, VPRM	~1122 (range: 295–1627)
	Ontario, Canada[a]	2004	Grain corn, hay, soybean, winter wheat, spring wheat, barley, fodder corn	Crop inventory conversion	~828 (range: 620–1102)
			Other Studies		
Xiao et al. [2010]	Conterminous United States	2001–2006	Corn, soybean	Integrated regression modeling	~1500
Yan et al. [2009]	Shandong, China	2003–2004	Winter wheat, corn	Eddy covariance flux measurements; integrated surface flux modeling, VPM	EC flux measurements: ~1410 [2003], ~2132 [2004]; model estimates: ~1624 [2003], ~1745 [2004]
Moureaux et al. [2008]	Namur, Belgian	2004–2005	Winter wheat	Eddy covariance flux measurements	~1580
Suyker et al. [2005]	Nebraska, United States	2002	Corn, soybean	Eddy covariance flux measurements	~1744 [corn], ~966 [soybean]
Turner et al. [2005]	Illinois, United States	2000	Corn, soybean	Integrated surface flux modeling, Biome-BGC; remote sensing measurements, MODIS MOD17	Model estimate: >900; MODIS estimate: ~600
Gilmanov et al. [2003]	Oklahoma, United States	1997	Winter wheat	Eddy covariance flux measurements	~2853

[a]Only southwestern Ontario was considered in this study. Census divisions in southwestern Ontario (Figure 1) include the following: Hamilton, Niagara, Haldimand-Norfolk, Brant, Oxford, Elgin, Chatham-Kent, Essex, Lambton, Middlesex, Peel, Dufferin, Wellington, Halton, Waterloo, Perth, Huron, and Simcoe. Bruce and Grey were removed owing to a lack of croplands in the IGBP-DISCover satellite land cover map linked to the VPRM (Figure 2b).

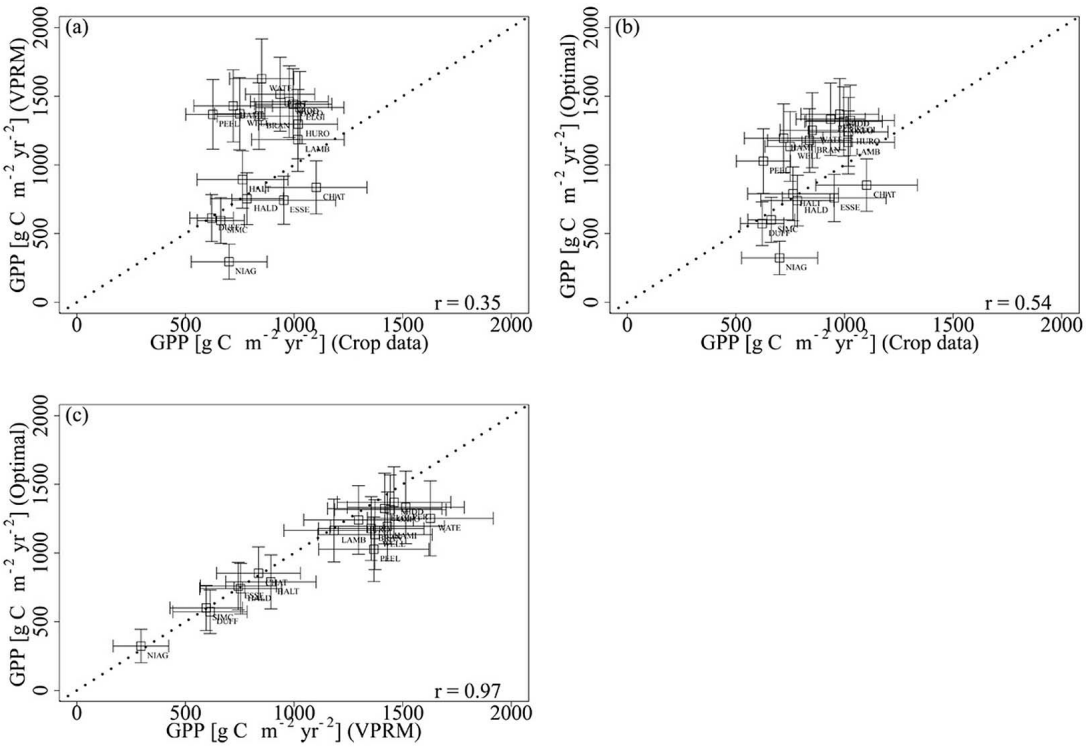

Figure 5. Comparison of GPP values between (a) a priori VPRM simulations and crop census data, (b) Bayesian-optimized VPRM simulations versus crop census data, and (c) optimized versus a priori VPRM simulations. The correlation coefficient *r* is also shown. Error bars associated with the crop-census-derived GPP are constructed from Monte Carlo analysis (section 3.2, Figure 4). The error bars associated with VPRM simulations are constructed via a method described in section 2.6.

Table 4. Comparison of GPP Estimates and Their Associated Uncertainties Over Southwestern Ontario[a]

	GPP (g C m^{-2} yr^{-1})			Associated Uncertainties (g C m^{-2} yr^{-1})			
Counties	Crop Data	VPRM	Optimal	Crop Data	VPRM	Optimal	Δ% (Optimal Relative to VPRM)
BRAN	839	1354	1179	194	242	232	−4.53
CHAT	1102	836	853	233	193	191	−1.04
ELGI	1027	1416	1324	203	263	257	−2.28
ESSE	952	743	759	242	175	173	−1.14
HALD	782	753	740	169	188	184	−2.13
HAMI	720	1429	1196	182	262	250	−4.58
LAMB	1016	1184	1163	212	231	229	−0.87
MIDD	978	1460	1369	179	261	259	−0.77
NIAG	700	295	324	172	128	122	−4.69
OXFO	995	1442	1319	182	257	252	−1.95
BRUC	727	NA	NA	116	NA	NA	NA
DUFF	620	613	572	102	171	159	−7.02
GREY	506	NA	NA	73	NA	NA	NA
HALT	762	893	790	211	208	196	−5.77
HURO	1018	1297	1241	183	252	250	−0.79
PEEL	626	1367	1028	125	254	235	−7.48
PERT	936	1514	1335	161	269	264	−1.86
SIMC	662	596	601	107	167	164	−1.8
WATE	852	1627	1255	149	290	273	−5.86
WELL	749	1372	1135	116	263	254	−3.42
Average	828	1122	1010	166	226	219	−3.22

[a]Uncertainties in agricultural census-derived GPP were determined by randomizing all parameters in a Monte Carlo analysis and assuming 3% uncertainty in provincial harvest production. See section 2.5 of the main text for details. The first four letters of the counties' names, as indicated in Figure 1, are used as abbreviations. NA, removed owing to a lack of croplands in the IGBP-DISCover satellite land cover map linked to the VPRM (Figure 2b).

Table 5. Comparison of Average GPP Estimates When a Subset of Crops is Selected From the Crop Census Data[a]

Crop-Type Combinations in Calculating Crop-Data-Derived GPP	Average GPP (g C m^{-2} yr^{-1})				Average Associated Uncertainties (g C m^{-2} yr^{-1})			
	Crop Data	VPRM	Optimal	Δ% (Optimal Relative to VPRM)	Crop Data	VPRM	Optimal	Δ% (Optimal Relative to VPRM)
Corn+soy	1030	1122	1092	−1.64	266	226	223	−1.40
Winter wheat+corn+soy	999	1122	1080	−2.41	230	226	222	−1.76
Winter wheat+corn+soy+hay	825	1122	1016	−7.42	171	226	220	−3.01
All crops	828	1122	1010	−7.63	166	226	219	−3.22

[a]Note that VPRM parameters from the "cropland" class were fitted against eddy covariance fluxes conducted over just two crops: corn and soy [*Mahadevan et al.*, 2008].

[41] To investigate point 3, we examined the sensitivity of crop-based GPP on the crop types included in the calculation. As seen in Table 5, the highest GPP is found when only corn and soy are considered. This higher GPP is closer to VPRM's, likely because VPRM was calibrated against eddy covariance data collected over corn and soy. Winter wheat and hay lower the GPP further, to almost the same value as the "all crop" case. Clearly, the mismatch in crop types explains a significant portion of the observed bias in VPRM GPP.

[42] Therefore, we expect consideration of winter wheat and hay during parameter calibration for VPRM to improve its simulation of GPP in Ontario's croplands. However, eddy covariance flux measurements for these crops are lacking, and this observational gap will need to be addressed in the future.

3.4. Bayesian Inversion to Optimize GPP Estimates

[43] In order to acquire insights into how the agricultural census-data-derived GPP improves knowledge about the VPRM-modeled GPP, we performed a Bayesian inversion, and results of this exercise are displayed in both Table 4 and Figure 5. After the inversion, the average optimized GPP over all counties is 1010 g C m^{-2} yr^{-1}, and its associated mean uncertainty is 219 g C m^{-2} yr^{-1}. This translates to a mere 3% reduction when compared to the average uncertainty associated with the VPRM-modeled GPP (226 g C m^{-2} yr^{-1}). Furthermore, the optimized GPP values are almost identical to the prior values simulated by the VPRM (Figure 5c). This result is perhaps not surprising, given the fact that uncertainties in the census-data-derived GPP are comparable to those in VPRM (Figure 5 and Table 4). Hence the "measured" GPP values from the agricultural census provide only minimal constraint on the simulations. It is worth noting, however, that the exact constraint provided by the agricultural data depends upon the VPRM model's a priori uncertainties ($\mathbf{S}_{prior}$; section 2.6). We have already pointed out in section 3.3 specific deficiencies of the VPRM simulations used in this study. Owing to uncertainties in the land cover data set and the model parameters, the model's prior uncertainties should be larger, in which case there would be a greater uncertainty reduction after the inversion.

4. Summary and Conclusions

[44] The objective of this study is to highlight and examine the potential value and information content of agricultural census data for carbon cycle studies. Specifically, two pieces of information that can be extracted from agricultural census data are examined here: (1) the areal coverage of croplands and (2) carbon uptake by field crops.

[45] This paper has shown that the census data provide a valuable source of information on the areal coverage of crops. Cropland fractions calculated from the census data reveal that all three satellite-based land cover data sets provide severely biased estimates of cropland fraction over southwestern Ontario, Canada. Since relative errors between census-data-derived versus satellite-based cropland fraction decrease when the cropland cover increases (Figure 3), this suggests that the error may be attributed largely to the difficulty of satellite data in distinguishing croplands embedded within a heterogeneous landscape. As a result, synthetic maps developed from merging land cover maps and agricultural census data [*Cardille et al.*, 2002; *Kerr and Cihlar*, 2003; *Ramankutty et al.*, 2008] should be considered for use in providing more reliable maps of crop coverage.

[46] On the other hand, carbon uptake estimates (GPP) based on the agricultural census data are subject to large uncertainties. The biggest sources of uncertainty are (in order of significance) the growth efficiency (GE), harvested production (P), and harvest index (HI). Although adopting NPP rather than GPP removes a large source of uncertainty (GE), significant errors are still present owing to imperfect knowledge of HI and P (Figure 4).

[47] Previous studies that only addressed the uncertainties associated with HI and root:shoot ratio and their effects on the overall uncertainty in estimating primary production [*Prince et al.*, 2001; *Bolinder et al.*, 2007] have underestimated the errors because they did not consider the possible impact imposed by uncertainties associated with harvested production estimates from agricultural census data.

[48] If the agricultural data are subject to large uncertainties, they simply cannot provide tight constraints to biospheric carbon models. As the minimal reduction in uncertainty from the Bayesian inversion suggests, there is limited information content in the agricultural data. We point out, however, that this result may be partly attributed to deficiencies in the VPRM parameters and land cover data that were not considered in the model's a priori uncertainties. If the model uncertainties were in fact larger, one could expect a greater reduction in uncertainty and thus more "information gain" due to use of the agricultural data.

[49] Ultimately, the uncertainties in agricultural production data stem from the fact that primary production, whether NPP or GPP, is not directly measurable and must be estimated on the basis of a suite of measurements and various underlying assumptions [*Scurlock et al.*, 1999; *Clark et al.*, 2001]. Exactly how losses due to herbivory and diseases are accounted for is another potential source of uncertainty in these calculations [*Ciais et al.*, 2010].

[50] Undoubtedly, agricultural census data have a significant role in monitoring crop production, security and sustainability. In the context of carbon cycle studies, they are likewise important because these spatially explicit data are crucial in helping to identify the areas where the various crops are actually grown. In this way, they provide valuable validation data for satellite land cover maps to enhance cropland identification and classification [*Cardille et al.*, 2002]. However, if the census data were used for carbon accounting purposes, one must be aware of the sources of errors as well as the underlying assumptions necessary to estimate the carbon fluxes.

[51] **Acknowledgments.** E.C.C. was supported by the Ontario Early Researcher Award awarded to J.C.L. and the Canadian Carbon Program (CCP). The CCP research networks were enabled by funding from the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS). We thank M. Pejam for assistance with GIS software and David Burroughs as well as Laura MacKenzie at Statistics Canada for their comments and suggestions on the use of the agricultural census data.

References

Amthor, J. S. (1989), *Respiration and Crop Productivity*, Springer, New York.

Bakwin, P. S., P. P. Trans, H. D. Hurst, and C. Zhao (1998), Measurements of carbon dioxide on very tall towers: Results of the NOAA/CMDL program, *Tellus, Ser. B*, *50*, 410–415.

Baldocchi, D. D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*(11), 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Bolinder, M. A., H. H. Janzen, E. G. Gregorich, D. A. Angers, and A. J. VandenBygaart (2007), An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada, *Agric. Ecosyst. Environ.*, *118*, 29–42, doi:10.1016/j.agee.2006.05.013.

Cardille, J. A., J. A. Foley, and M. H. Costa (2002), Characterizing patterns of agricultural land use in Amazonia by merging satellite classifications and census data, *Global Biogeochem. Cycles*, *16*(3), 1045, doi:10.1029/2000GB001386.

Ciais, P., et al. (2010), The European carbon balance. Part 2: Croplands, *Global Change Biol.*, *16*, 1409–1428, doi:10.1111/j.1365-2486.2009.02055.x.

Clark, D. A., S. Brown, D. W. Kicklighter, J. Q. Chambers, J. R. Thomlison, and J. Ni (2001), Measuring net primary production in forests: Concepts and field methods, *Ecol. Appl.*, *11*(2), 356–370, doi:10.1890/1051-0761(2001)011[0356:MNPPIF]2.0.CO;2.

Cramer, W., et al. (1999), Comparing global models of terrestrial net primary productivity (NPP): Overview and key results, *Global Change Biol.*, *5*, suppl. 1, 1–15, doi:10.1046/j.1365-2486.1999.00009.x.

Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*, 270–277, doi:10.1126/science.289.5477.270.

de Jong, R., K. Y. Li, A. Bootsma, T. Huffman, G. Roloff, and S. Gameda (2001), A crop yield and variability under climate change and adaptative crop management scenarios, final report for Climate Change Action Fund Project A080, East. Cereal and Oilseed Res. Cent., Agric. and Agric.-Food Can., Ottawa.

Desjardins, R. L., W. Smith, B. Grant, C. Campbell, and R. Riznek (2005), Management strategies to sequester carbon in agricultural soils and to mitigate greenhouse gas emissions, *Clim. Change*, *70*, 283–297, doi:10.1007/s10584-005-5951-y.

Donald, C. M., and J. Hamblin (1976), The biological yield and harvest index of cereals as agronomic and plant breeding criteria, *Adv. Agron.*, *28*, 361–405, doi:10.1016/S0065-2113(08)60559-3.

Falk, D. E., S. Li, and D. L. Rinker (2007), Value adding wheat straw, *ONWheat Mag. Res. Ed.*, pp. 16–17, Ont. Wheat Prod. Market. Board, Guelph, Ont., Canada.

Frolking, S., X. M. Xiao, Y. H. Zhung, W. Salas, and C. S. Li (1999), Agricultural land-use in China: A comparison of area estimates from ground-based census and satellite-borne remote sensing, *Global Ecol. Biogeogr.*, *8*(5), 407–416, doi:10.1046/j.1365-2699.1999.00157.x.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003a), Towards constraining regional scale fluxes of CO_2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, *J. Geophys. Res.*, *108*(D24), 4756, doi:10.1029/2002JD003018.

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger (2003b), Towards constraining regional scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, *108*(D24), 4757, doi:10.1029/2003JD003770.

Gerbig, C., J. C. Lin, J. W. Munger, and S. C. Wofsy (2006), What can tracer observations in the continental boundary layer tell us about surface-atmosphere fluxes?, *Atmos. Chem. Phys.*, *6*, 539–554, doi:10.5194/acp-6-539-2006.

Gilmanov, T. G., S. B. Verma, P. L. Sims, T. P. Meyers, J. A. Bradford, G. G. Burba, and A. E. Suyker (2003), Gross primary production and light response parameters of four Southern Plains ecosystems estimated using long-term CO_2-flux tower measurements, *Global Biogeochem. Cycles*, *17*(2), 1071, doi:10.1029/2002GB002023.

Hay, R. K. M. (1995), Harvest index: A review of its use in plant breeding and crop physiology, *Ann. Appl. Biol.*, *126*, 197–216, doi:10.1111/j.1744-7348.1995.tb05015.x.

Hay, R. K. M., and R. A. Gilbert (2001), Variation in the harvest index of tropical maize: Evaluation of recent evidence from Mexico and Malawi, *Ann. Appl. Biol.*, *138*, 103–109, doi:10.1111/j.1744-7348.2001.tb00090.x.

Herold, M., P. Mayaux, C. E. Woodcock, A. Baccini, and C. Schmullius (2008), Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets, *Remote Sens. Environ.*, *112*, 2538–2556, doi:10.1016/j.rse.2007.11.013.

Hicke, J. A., and D. B. Lobell (2004), Spatiotemporal patterns of cropland area and net primary production in the central United States estimated from USDA agricultural information, *Geophys. Res. Lett.*, *31*, L20502, doi:10.1029/2004GL020927.

Huang, Y., W. Zhang, W. J. Sun, and X. H. Zheng (2007), Net primary production of Chinese croplands from 1950 to 1999, *Ecol. Appl.*, *17*(3), 692–701, doi:10.1890/05-1792.

Hurtt, G. C., L. Rosentrater, S. Frolking, and B. Moore (2001), Linking remote-sensing estimates of land cover and census statistics on land use to produce maps of land use of the conterminous United States, *Global Biogeochem. Cycles*, *15*(3), 673–685, doi:10.1029/2000GB001299.

Hutchinson, J. J., C. A. Campbell, and R. L. Desjardins (2007), Some perspectives on carbon sequestration in agriculture, *Agric. For. Meteorol.*, *142*, 288–302, doi:10.1016/j.agrformet.2006.03.030.

Intergovernmental Panel on Climate Change (2000), Good practice guidance and uncertainty management in national greenhouse gas inventories, report, Geneva.

Intergovernmental Panel on Climate Change (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, New York.

Jaynes, E. T. (1968), Prior probabilities, *IEEE Trans. Syst. Sci. Cybern.*, *4*(3), 227–241, doi:10.1109/TSSC.1968.300117.

Jenkins, J. C., R. A. Birdsey, and Y. Pan (2001), Biomass and NPP estimation for the Mid-Atlantic region (USA) using plot-level forest inventory data, *Ecol. Appl.*, *11*(4), 1174–1193, doi:10.1890/1051-0761(2001)011[1174:BANEFT]2.0.CO;2.

Jung, M., K. Henkel, M. Herold, and G. Churkina (2006), Exploiting synergies of global land cover products for carbon cycle modeling, *Remote Sens. Environ.*, *101*, 534–553, doi:10.1016/j.rse.2006.01.020.

Karl, T. R., and K. E. Trenberth (2003), Modern global climate change, *Science*, *302*, 1719–1723, doi:10.1126/science.1090228.

Kerr, J. T., and J. Cihlar (2003), Land use and cover with intensity of agriculture for Canada from satellite and census data, *Glob. Ecol. Biogeogr.*, *12*(2), 161–172, doi:10.1046/j.1466-822X.2003.00017.x.

Kroodsma, D. A., and C. B. Field (2006), Carbon sequestration in California agriculture, 1980–2000, *Ecol. Appl.*, *16*(5), 1975–1985, doi:10.1890/1051-0761(2006)016[1975:CSICA]2.0.CO;2.

Lal, R. (2004), Agricultural activities and the global carbon cycle, *Nutr. Cycling Agroecosyst.*, *70*, 103–116, doi:10.1023/B:FRES.0000048480.24274.0f.

Lin, J. C., and C. Gerbig (2005), Accounting for the effect of transport errors on tracer inversions, *Geophys. Res. Lett.*, *32*, L01802, doi:10.1029/2004GL021127.

Lin, J. C., M. R. Pejam, E. Chan, S. C. Wofsy, E. W. Gottlieb, H. A. Margolis, and J. H. McCaughey (2011), Attributing uncertainties in simulated biospheric carbon fluxes to different error sources, *Global Biogeochem. Cycles*, *25*, GB2018, doi:10.1029/2010GB003884.

Lobell, D. B., J. A. Hicke, G. P. Asner, C. B. Field, C. J. Tucker, and S. O. Los (2002), Satellite estimates of productivity and light use efficiency in United States agriculture, 1982–98, *Global Change Biol.*, *8*, 722–735, doi:10.1046/j.1365-2486.2002.00503.x.

Loveland, T. R., and A. S. Belward (1998), The international geosphere biosphere programme data and information system global land cover data set (DIScover), *Acta Astronaut.*, *41*(4–10), 681–689.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, J. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, *21*, 1303–1330, doi:10.1080/014311600210191.

Mahadevan, P., S. C. Wofsy, D. M. Matross, X. Xiao, A. L. Dunn, J. C. Lin, C. Gerbig, J. W. Munger, V. Y. Chow, and E. W. Gottlieb (2008), A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochem. Cycles*, *22*, GB2005, doi:10.1029/2006GB002735.

Malmström, C. M., M. V. Thompson, G. P. Juday, S. O. Los, J. T. Randerson, and C. B. Field (1997), Interannual variation in global-scale net primary production: Testing model estimates, *Global Biogeochem. Cycles*, *11*(3), 367–392, doi:10.1029/97GB01419.

Matross, D. M., et al. (2006), Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus, Ser. B*, *58*, 344–358, doi:10.1111/j.1600-0889.2006.00206.x.

Monfreda, C., N. Ramankutty, and J. A. Foley (2008), Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, *Global Biogeochem. Cycles*, *22*, GB1022, doi:10.1029/2007GB002947.

Morrison, M. J., H. D. Voldeng, and E. R. Cober (1999), Physiological changes from 58 years of genetic improvement of short-season soybean cultivars in Canada, *Agron. J.*, *91*, 685–689, doi:10.2134/agronj1999.914685x.

Moureaux, C., A. Debacq, J. Hoyaux, M. Suleau, D. Tourneur, F. Vancutsem, B. Bodson, and M. Aubinet (2008), Carbon balance assessment of a Belgian winter wheat crop (Triticum aestivum L.), *Global Change Biol.*, *14*, 1353–1366, doi:10.1111/j.1365-2486.2008.01560.x.

O'Neill, D. W. (2005), Human appropriation of the products of photosynthesis in Nova Scotia, Canada, Masters thesis, Dalhousie Univ., Halifax, Nova Scotia, Canada.

Ontario Ministry of Agriculture, Food and Rural Affairs (2009), Agronomy guide for field crops, *Publ. 811*, Guelph, Canada.

Ontario Ministry of Natural Resources (2002), Ontario provincial-scale land cover, digital land cover map, Peterborough, Canada. (Available at http://www.lib.uwaterloo.ca/locations/umd/digital/olc.html.)

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993), Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochem. Cycles*, *7*(4), 811–841, doi:10.1029/93GB02725.

Prince, S. D., J. Haskett, M. Steininger, H. Strand, and R. Wright (2001), Net primary production of US Midwest croplands from agricultural harvest yield data, *Ecol. Appl.*, *11*(4), 1194–1205, doi:10.1890/1051-0761(2001)011[1194:NPPOUS]2.0.CO;2.

Ramankutty, N., and J. Foley (1998), Characterizing patterns of global land use: An analysis of global croplands data, *Global Biogeochem. Cycles*, *12*(4), 667–685, doi:10.1029/98GB02512.

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley (2008), Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, *Global Biogeochem. Cycles*, *22*, GB1003, doi:10.1029/2007GB002952.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. Defries, M. Heimann, D. S. Ojima, S. Quegan, and C. Schmullius (2005), Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, *Global Change Biol.*, *11*, 378–397, doi:10.1111/j.1365-2486.2005.00917.x.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci., River Edge, N. J.

Rollefson, J., G. Fu, and A. Chan (2004), Assessment of the environmental performance and sustainability of biodiesel in Canada, report, Natl. Res. Counc. Can., Ottawa.

Ruimy, A., G. Dedieu, and B. Saugier (1996), TURC: A diagnostic model of continental gross primary productivity and net primary productivity, *Global Biogeochem. Cycles*, *10*(2), 269–285, doi:10.1029/96GB00349.

Scurlock, J. M. O., W. Cramer, R. J. Olson, W. J. Parton, and S. D. Prince (1999), Terrestrial NPP: Toward a consistent data set for global model evaluation, *Ecol. Appl.*, *9*(3), 913–919.

Statistics Canada (2004), November estimate of production of principal field crops, Canada, 2004, *Field Crop Rep. Ser.*, vol. 83(8), Gov. of Can., Ottawa.

Statistics Canada (2006), 2006 Census of Agriculture, http://www.statcan.gc.ca/pub/95-629-x/2007000/4182415-eng.htm., Gov., of Can., Ottawa.

Suyker, A. E., S. B. Verma, G. G. Burba, and T. J. Arkebauer (2005), Gross primary production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season, *Agric. For. Meteorol.*, *131*, 180–190, doi:10.1016/j.agrformet.2005.05.007.

Taylor, J. R. (1997), *An Introduction to Error Analysis: The Study of Uncertainty in Physical Measurements*, 2nd ed., Univ. Sci. Books, Sausalito, Calif.

Turner, D. P., et al. (2005), Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring, *Global Change Biol.*, *11*, 666–684, doi:10.1111/j.1365-2486.2005.00936.x.

Vogelmann, J. E., D. Helder, R. Morfitt, M. J. Choate, J. W. Merchant, and H. Bulley (2001), Effects of Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper Plus radiometric and geometric calibrations and corrections on landscape characterization, *Remote Sens. Environ.*, *78*, 55–70, doi:10.1016/S0034-4257(01)00249-8.

Wu, W., R. Shibasaki, P. Yang, L. Ongaro, Q. Zhou, and H. Tang (2008), Validation and comparison of 1 km global land cover products in China, *Int. J. Remote Sens.*, *29*, 3769–3785, doi:10.1080/01431160701881897.

Wulder, M., B. Boots, D. Seemann, and J. White (2004), Map comparison using spatial autocorrelation: An example using AVHRR derived land cover of Canada, *Can. J. Remote Sens.*, *30*(4), 573–592.

Xiao, J., et al. (2010), A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data, *Remote Sens. Environ.*, *114*, 576–591, doi:10.1016/j.rse.2009.10.013.

Yan, H., Y. Fu, X. Xiao, H. Q. Huang, H. He, and L. Ediger (2009), Modeling gross primary productivity for winter wheat-maize double cropping system using MODIS time series and CO_2 eddy flux tower data, *Agric. Ecosyst. Environ.*, *129*, 391–400, doi:10.1016/j.agee.2008.10.017.

E. C. Chan and J. C. Lin, Waterloo Atmosphere-Land Interactions Research Group, Department of Earth and Environmental Sciences, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada. (erichanch@gmail.com)