The Value of Agricultural Census Data in Carbon Cycle Studies

by

Eric Chan

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Earth Sciences

Waterloo, Ontario, Canada, 2011

Author’s declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Agricultural census data have been identified to possess the potential to provide constraints on 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 and by how much constrain the crop data can provide, as exhibited by uncertainty reductions.

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 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.

Acknowledgements

I would like to sincerely thank Dr. John C. Lin, my supervisor, for offering me this great opportunity to further develop my skills and myself as a person. In addition, I would also like to show my immense appreciation for his inspiration, help, and most importantly, patience, over the last two years. Other than that, I would like to thank all my friends in Waterloo who have accompanied me and brought me joy during this time. And above all, I would like to thank my mom for her limitless love and support.

Table of Content

List of Figures

List of Tables

Chapter 1 Introduction

1.1 The importance of carbon cycle

Carbon (C) is a key element for all life forms on Earth and circulates through the planet's interior, land, ocean, and atmosphere. While carbon cycles through these individual reservoirs, the exchange is often in both directions. The carbon budget, an accounting of carbon within these various reservoirs, should be in quasi-equilibrium state under natural conditions. However, currently this is not the case because carbon is excessively accumulating in the form of atmospheric carbon dioxide (CO_2) and methane (CH_4) since the preindustrial era (circa 1750) [*IPCC*, 2007]. From a climatological standpoint, this is not desirable because it will lead to a disruption in the energy balance within the Earth's atmosphere.

Energy that drives climate and weather on Earth originates from the Sun. At the Earth's surface, this energy is absorbed, reflected, and redistributed by the atmospheric and oceanic circulation before it is radiated back to space as long-wave infrared radiation. For Earth as a whole, the incoming solar energy is balanced by the outgoing long-wave radiation. Therefore, any factor that alters this balance within Earth's atmosphere will affect the climate [*IPCC*, 2007]. In the atmosphere, major constituent gases such as nitrogen (N_2), oxygen (O_2) and argon (Ar) are transparent to the outgoing long-wave radiation yet trace gases like CO_2, CH_4, nitrous oxide (N_2O), water vapor and halocarbons (a group of gases containing fluorine, chlorine or bromine) are not. Consequently, increase in concentrations of the said trace gases, also known as greenhouse gases (GHGs), will reduce the efficiency at which the reflected long-wave

radiation transmits into space and thus cause the outgoing long-wave radiation from the surface to be absorbed by the atmosphere and re-emitted at higher altitude; eventually affecting the global energy budget and climate [*Crowley*, 2000; *Karl and Trenberth*, 2003].

The resulting change in energy balance is expressed as a positive radiative forcing which tends to warm the lower atmosphere and Earth's surface. This "greenhouse effect" depends on the amount of increase in the concentration of each GHG, the radiative properties of the gases involved, and the concentrations of other GHGs already present in the atmosphere [*IPCC*, 2007]. The long-lived CO_2, CH_4, N_2O and halocarbons can be produced from natural as well as anthropogenic activities. The combined radiative forcing due to increases in these GHGs is $+2.3 \pm 0.2$ $W \cdot m^{-2}$ in 2005 and the CO_2 radiative forcing had increased by 20% from 1995 to 2005 [*IPCC*, 2007]. Among all the GHGs, CO_2 has been the most important contributing factor to the increased forcing [*IPCC*, 2007]. Global atmospheric concentration of CO_2 is 379 ppm in 2005 and the annual CO_2 concentration growth rate was larger during 1995 to 2005 (average: 1.9 ppm per year) than during 1960 to 2005 (average: 1.4 ppm per year), although there is significant year-to-year variability in growth rates. Accordingly, the CO_2 radiative forcing has increased by 20% from 1995 to 2005, the largest change for any decade in at least the last 200 years [*IPCC*, 2007].

From an oceanic standpoint, increase in atmospheric CO_2 concentrations is also responsible for increased acidity in the surface ocean [*Caldeira and Wickett*, 2003], which in turn may lead to dreadful future consequences for corals and other marine organisms that build their skeletons and shells from calcium carbonate, a chemical compound that is dissolvable even in weak acid. Ocean acidification is therefore another

reason why we must study the carbon cycle and the accumulation of CO_2 in the atmosphere [*Orr et al*., 2005].

1.2 Carbon cycle and croplands

Primary production of a terrestrial ecological system refers to the rate at which autotrophic producers, mainly green plants, produce organic substances that are used to maintain and support growth through photosynthesis and chemosynthesis [*Odum*, 1971]. In both processes, CO_2 from the atmosphere is converted into carbohydrate (CH_2O) that is required for biochemical synthesis of new biomass in plants. The gross primary production (*GPP*) quantifies gross carbon fixed by vegetation in terrestrial ecosystems and is in effect the principal indicator of biospheric carbon fluxes. Given that the terrestrial biosphere is the prime bearer of plants, animals, soil microbes, and decomposing organic materials, it is therefore a large reservoir of organic carbon and facilitates the exchange of carbon between the biota and atmosphere.

Croplands play an important role in the terrestrial carbon cycle because they cover ~15 million km^2 of the planet [*Ramankutty 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.

1.3 Approaches to assess carbon budget

Biospheric models [*Potter et al*., 1993; *Ruimy et al*., 1996; *Sellers 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 spatio-temporal 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 [*Gerbig et al*., 2003a, 2003b; *Matross et al*., 2006]. However, inverse methods still require work in 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 (~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 dataset 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.

1.4 Project objective

Agricultural census data have been identified to possess the potential to provide constraints on carbon uptake by croplands at the regional scale. While numerous studies recognized and attempted to take advantage of this potential, they did not do it in a comprehensive manner as they neither compared the census data against multiple satellite land cover datasets 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 inter-annual fit between census data-derived estimates and satellite-based modeled fluxes by examining correlation coefficients. Neither study considered a priori uncertainties in either of these 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 build on previous efforts to further assess this potential quantitatively by using agricultural census data collected by the Canadian federal and provincial agencies to determine fractional cropland coverage (required for identifying where the field crops are grown) as well as 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 to obtain "optimal" posterior estimates (i.e. expected values of the posterior) for all sub-provincial jurisdictions in southwestern Ontario. The reduction in uncertainties, presented as the percentage differences between prior and posterior uncertainties, will be analyzed to determine if and by how much the agricultural data constrained the modeled fluxes.

Chapter 2 Area of interest, agricultural census data and data pre-processing

2.1 Area of interest

According to the 2006 Canadian Census of Agriculture, Ontario is one of the major field crop production 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. In this study, the area of interest is southwestern Ontario (Figure 1), the location where the majority of croplands are situated in the province.

2.2 Agricultural census data

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.

For comparing fractional cropland coverage, we used harvested area estimates

reported for all field crops listed on the OMAFRA field crop statistics webpage. 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 datasets such as moisture content, harvest index and root:shoot ratio of other field crops required for the conversion process (eq. (7)) 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].

Data for uncertainty analysis (Section 4.2.3) 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 sub-provincial level uncertainties. Every five 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].

County-level uncertainties associated with harvested area and production are calculated through downscaling the corresponding values reported at larger (higher level) jurisdictions (Section 4.2.3). 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.

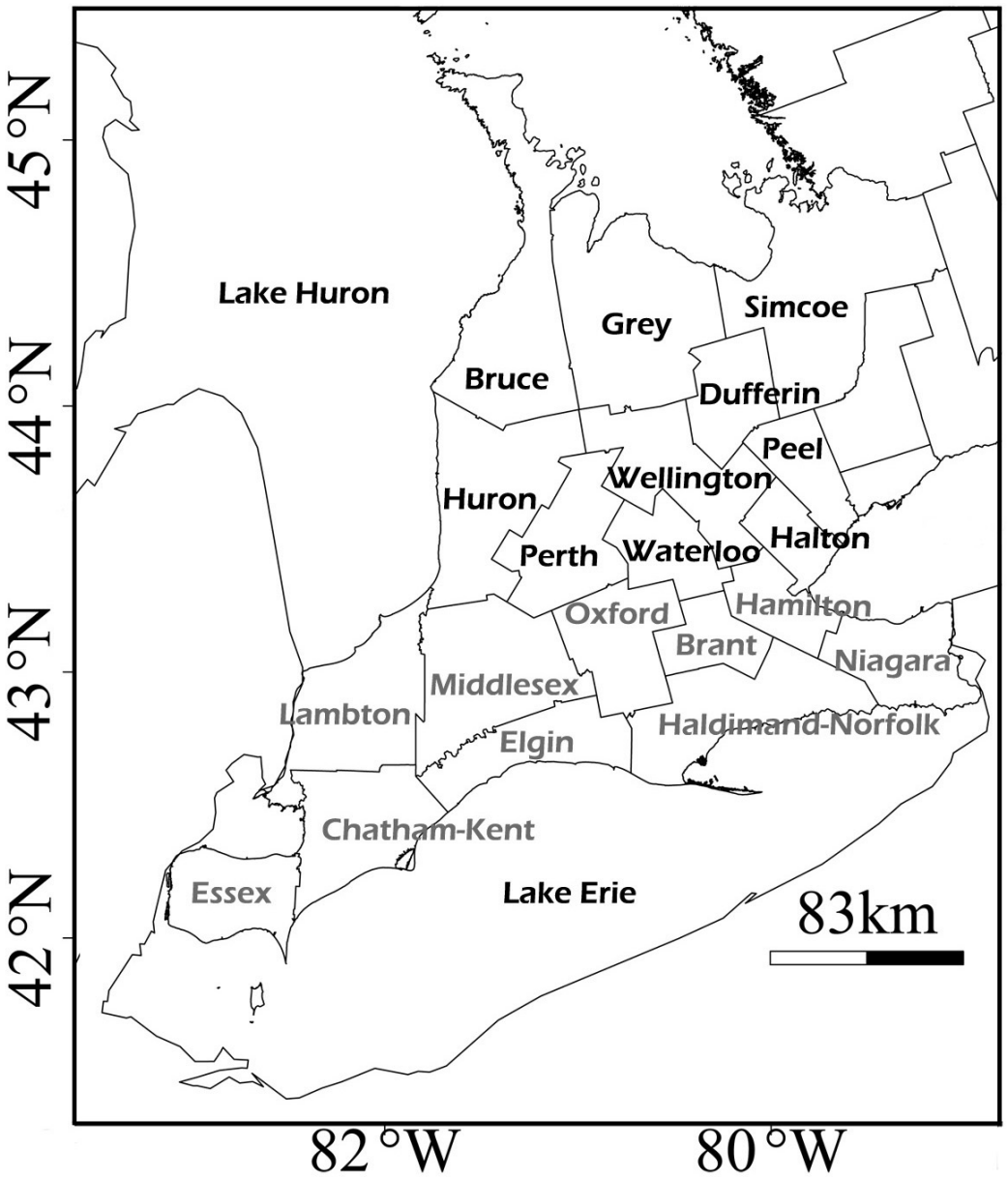

Figure 1. Map of census divisions in southwestern Ontario. Region 1 - southern Ontario (in gray); Region 2 - western Ontario (in black).

2.3 Data pre-processing

2.3.1 Creation of a raster map

The final benchmark map that is responsible for constraining the satellite-based land cover maps is a raster image of Ontario at a resolution of 1/6°(latitude) by 1/4°(longitude) and is created by aggregating the pixels of a prior image of 1/96° by 1/96° through averaging. We created the initial finer image based on the OMAFRA agricultural land use dataset [*OMAFRA*, 1988] because it contains unique identification (UID) numbers for each municipality (census subdivision / subcounty (CSD)) and its parent legislated area (census division / county (CD)) that are described in geographic co-ordinates. Such features, when incorporated in a raster image, facilitate the mapping of any census data because we can easily replace the spatially defined UIDs with the corresponding census data in the matrix that describes the image. In constraining the land cover maps, the UIDs are eventually replaced by cropland fractions. Table 1 is an excerpt of the cross-reference table that we used in creating the final map. The UIDs in the original dataset are seven-digits long and are cumbersome to work with. For that reason, we substituted them with another set of UIDs that are one- to two-digits long.

Table 1. Excerpt of the cross-reference table used in creating the raster map of cropland fraction

CSDNAME	CSDUID	COUNT	CSDArea [sqkm]	CDNAME	CDCODE	CDUID	CDArea [sqm]	CDHarvArea [sqm]	HarvAreaFrac
South Glengarry	1	645	609.98	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
Akwesasne (Part) 59	2	16	15.18	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
South Stormont	3	477	451.97	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
Cornwall	4	63	59.73	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
South Dundas	5	549	521.32	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
North Dundas	6	532	503.97	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
North Stormont	7	548	518.05	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
North Glengarry	8	686	646.85	Stormont, Dundas and Glengarr	STOR	1	3327032970	1437443401	0.43
East Hawkesbury	9	264	248.23	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Hawkesbury	10	13	12.20	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Champlain	11	236	221.70	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Alfred and Plantagenet	12	435	408.75	Prescott and Russell	PRES	2	2062102030	864813217	0.42
The Nation / La Nation	13	701	660.57	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Clarence-Rockland	14	326	306.70	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Casselman	15	4	3.77	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Russell	16	212	200.18	Prescott and Russell	PRES	2	2062102030	864813217	0.42
Ottawa	17	3061	2889.21	Ottawa	OTTA	3	2889206000	635356458	0.22
Edwardsburgh/Cardinal	18	328	312.06	Leeds and Grenville	LEED	4	3600557086	549158417	0.15
Augusta	19	332	316.37	Leeds and Grenville	LEED	4	3600557086	549158417	0.15
Prescott	20	5	4.77	Leeds and Grenville	LEED	4	3600557086	549158417	0.15
Elizabethtown-Kitley	21	593	565.53	Leeds and Grenville	LEED	4	3600557086	549158417	0.15
Brockville	22	21	20.06	Leeds and Grenville	LEED	4	3600557086	549158417	0.15
Front of Yonge	23	136	130.09	Leeds and Grenville	LEED	4	3600557086	549158417	0.15

2.3.2 Estimating the total land area of rasterized municipalities and counties

A rasterized municipality's and county's total land area is estimated by the total grid cell area that a set of identical UIDs occupy. In turn, one grid cell area is simply the product of the lengths in latitude and longitude that a grid cell covers. In this study, we employed a simple geometric reasoning for determining the lengths: latitudinal distance (in kilometers) between two points is approximated as the radius of the Earth times the latitudinal difference in radians:

$$d_{LAT} = R_{EARTH} \times (y_2 - y_1) \times \frac{\pi}{180} \quad (1)$$

where y_1 and y_2 are in degrees. Likewise, the longitudinal distance is approximated as:

$$d_{LON} = R_{EARTH} \times \cos\left[\frac{(y_2 + y_1)}{2} \times \frac{\pi}{180}\right] \times (x_2 - x_1) \times \frac{\pi}{180} \quad (2)$$

where x_1 and x_2 are also in degrees and R_{EARTH} is assumed to be 6371 km.

Approximated grid cell areas are spatially defined and arranged in a matrix form that is identical to that of the UIDs. Therefore, to estimate the total grid cell area that a specific county take up, we can first overlap the two matrices to count all the grid cells that have the same UIDs and identify their corresponding latitudes and longitudes. Then we can use the identified co-ordinates from the UIDs matrix to set a boundary within which matching grid cell areas in the grid cell area matrix can be picked out. Finally, we sum up the selected grid cell areas and get the total land area.

2.3.3 Creation of the K matrix

In Chapters 3 and 4, a matrix denoted as **K** (following the notation used in Rodgers, 2000)

is used extensively and plays as an important role during the analyses. According to Rodgers [2000], under different circumstances the **K** matrix may also be referred to as the Jacobian, the kernel (hence **K**), the sensitivity kernel, the tangent linear model or the adjoint. In this study, **K** of size $m \times n$ is responsible for transforming gridded results such as cropland fraction and biospheric model generated carbon uptake into "discrete observations" as seen at the separate census divisions; *m* represents the 20 individual census divisions within southwestern Ontario and *n* represents the total number of grid cells in a 1/6° by 1/4° within some desired domain. Accordingly, each element along *m* represents the fraction of cells of a particular census division, originally gridded at 1/96° by 1/96°, that is now present in a single 1/6° by 1/4° cell.

To help visualize this, consider three different regions (Region 1, 2 and 3) that are initially gridded within some domain at a resolution of 1/96° by 1/96° (Figure 2). Region 1 is made up of 336 finer grid cells, Region 2 is made up of 812 finer grid cells and Region 3 is made up of 388 finer grid cells. If we conceptually alter the scale of resolution of the entire domain and look at it as if it is gridded at 1/6° by 1/4°, the domain would be characterized by four coarser grid cells. Now suppose we determined that each of the four coarser grid cells is associated with a cropland fraction estimate, defined as vector **x**: for Cell 1 (top left corner) it is 0.60, Cell 2 (top right corner) it is 0.50, Cell 3 (bottom left corner) it is 0.70 and Cell 4 (bottom right corner) it is 0.90; then in order to deduce the actual cropland fraction in each of the three regions, defined as vector **y**, we can simply multiply **K** by **x** as follow:

$$\begin{bmatrix} 303/336 & 33/336 & 0 & 0 \\ 10/812 & 351/812 & 67/812 & 384/812 \\ 71/388 & 0 & 317/388 & 0 \end{bmatrix} \begin{bmatrix} 0.60 \\ 0.50 \\ 0.70 \\ 0.90 \end{bmatrix} = \begin{bmatrix} 0.59 \\ 0.72 \\ 0.70 \end{bmatrix} \begin{matrix} \leftarrow \text{For Region 1} \\ \leftarrow \text{For Region 2} \\ \leftarrow \text{For Region 3} \end{matrix}$$

$$\mathbf{K} \qquad \mathbf{x} \qquad \mathbf{y}$$

In this example, each element in **K** is equal to the number of finer grid cells representing a specific county located within one coarser grid cell divided by the total number of finer grid cells representing a specific county.

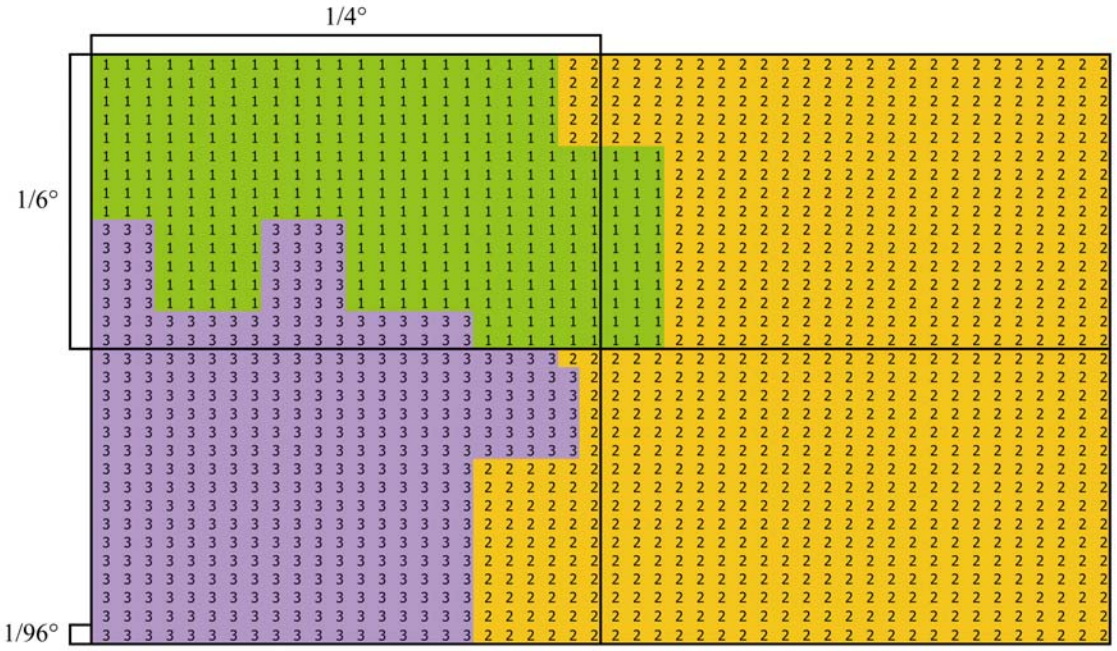

Figure 2. Gridded Region 1, 2 and 3. Top left corner - Coarser Cell 1; Top right corner - Coarser Cell 2; Bottom left corner - Coarser Cell 3; and Bottom right corner - Coarser Cell 4.

In Chapter 3, **K** is applied to calculate cropland fractions estimated by four different datasets and in Chapter 4, **K** is used in a Bayesian inversion to help determine the value of agricultural census data.

Chapter 3 Validating satellite-based land cover classification maps

3.1 Introduction

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 datasets (IGBP-DISCover and SYNMAP) are commonly used in terrestrial carbon cycle modeling studies. A brief description for each product is provided below. Readers are encouraged to refer to the original publications for further details.

3.1.1 IGBP-DISCover dataset

The original IGBP-DISCover dataset, 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 dataset 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 dataset as "IGBP" from this point on.

3.1.2 SYNMAP

SYNMAP is an enhanced 1-km land cover dataset that was developed through synthesizing the Global Land Cover Characterization Data Base (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 dataset 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. Due to its possible better performance over IGBP, we included SYNMAP in this study.

3.1.3 Ontario Land Cover database

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 meters resolution, much higher than that of the previous two products, and should provide the best land cover description of Ontario.

3.2 Method

3.2.1 Homogenization of the satellite-based land cover maps

Developed from differing approaches and raw data, the three satellite products map vegetation in diverse resolutions. Therefore, in order to facilitate comparison of the different maps, we need to have a homogenized resolution among the satellite-based datasets. As mentioned earlier, since the outputs of the Mahadevan et al [2008] biospheric model are used and analyzed in another chapter, the model's spatial resolution thus becomes the standard that the resizing of others are based upon. In this study, all satellite-based land cover maps are eventually resized to a spatial resolution of 1/6° (latitude) by 1/4° (longitude) and clipped to a subset with 96 rows and 104 columns to center in southwestern Ontario.

3.2.2 Mapping census data-derived cropland fraction

As mentioned in Chapter 2 (Section 2.3.1), a raster image of Ontario was prepared and census data can be mapped onto it by simply replacing the spatially gridded UIDs with the corresponding census data in the matrix that describes the image. To map cropland fraction, we first calculated the ratio by dividing the county-level harvested area of a particular county by its total land area, approximated by the total grid cell area it covers (Section 2.3.1). Subsequently, we changed the CDUIDs of that county, already mapped on a 1/96° by 1/96° resolution raster image, into estimated cropland fraction. In the end, we aggregated the resolution of this image into 1/6° by 1/4° through averaging the pixels in order to make it comparable with all other satellite-based land cover maps. Now, the census data-derived cropland fraction map becomes the benchmark to evaluate the

performance of the satellite products. Agreement between datasets is evaluated by comparing the derived cropland fractions among the various maps and each map is multiplied with the K matrix (Section 2.3.3) to determine the estimated cropland fraction for each census division.

3.3 Results and discussions

Table 2 tabulates the comparison of approximated and actual surveyed census division land areas given by Statistics Canada [*Statistics Canada*, 2006]. Results from this table indicate that our approximations of the land area are quite precise and thus our estimation of cropland fraction should be reasonable. Pictorially, Figure 3a displays the cropland distributions in southwestern Ontario as described by the agricultural census data. Uncertainties associated with these cropland fractions (σ_{Frac} in Table 3) are calculated by dividing the σ_{HA} (Section 4.2.3) 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 3).

However, this is not the case in the IGBP "croplands" map (Figure 3b), 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 3). 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 Mahadevan et al.'s reclassification scheme affected land cover representation (Section 3.1.1), we merged the modified IGBP "croplands" class with the "grasslands" class to construct a new "croplands+grasslands" map. This new map (Figure 3d), 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 3).

Table 2. Comparison of approximated and actual surveyed census division land areas

Census Division	Approximated Land Area [km^2]	Surveyed Land Area [km^2]	%Δ
BRAN	1101.72	1092.95	0.01
CHAT	2480.08	2470.66	0.00
ELGI	1877.00	1880.84	0.00
ESSE	1866.03	1851.34	0.01
HALD	2912.12	2894.15	0.01
HAMI	1149.22	1117.21	0.03
LAMB	3064.22	3001.70	0.02
MIDD	3338.57	3317.15	0.01
NIAG	1880.61	1854.17	0.01
OXFO	2049.18	2039.46	0.00
BRUC	4137.93	4079.17	0.01
DUFF	1494.88	1485.58	0.01
GREY	4542.94	4508.12	0.01
HALT	968.16	967.17	0.00
HURO	3407.48	3396.68	0.00
PEEL	1250.01	1242.40	0.01
PERT	2218.29	2218.41	0.00
SIMC	5323.37	4840.56	0.10
WATE	1389.70	1368.64	0.02
WELL	2692.68	2656.66	0.01

Table 3. Comparison of cropland fractions derived from different sources. The first 4 letters of the counties' names, as indicated in Figure 1, are used as abbreviations. σ_{Frac} is the uncertainty in cropland fractions (Section 3.1).

Counties	OMAFRA		IGBP Croplands		IGBP Croplands+Grasslands		SYNMAP		Ontario Land Cover database	
	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**

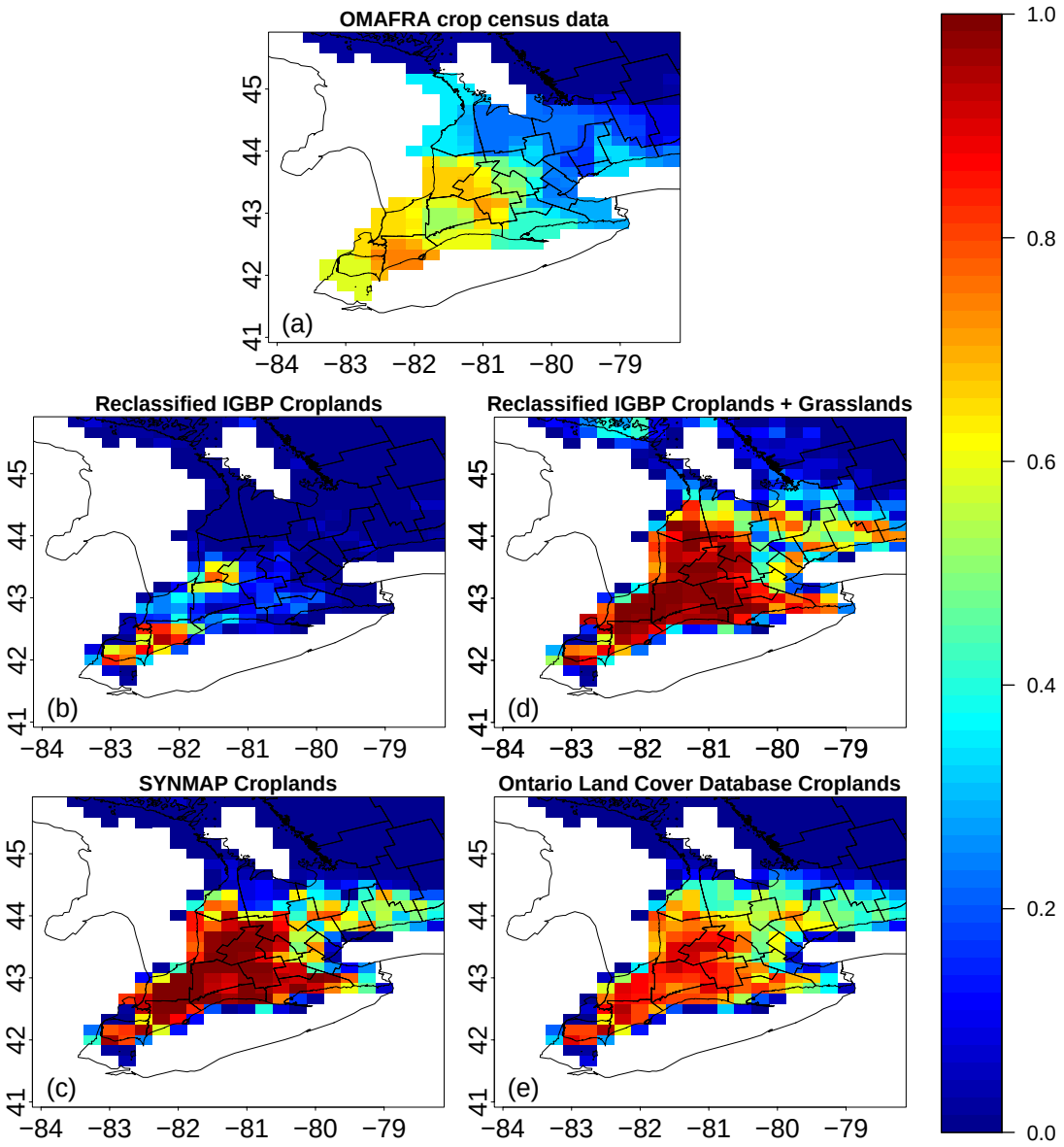

Figure 3. Fractional coverage of croplands and other vegetation derived from various datasets: (a) agricultural census data from OMAFRA; (b) IGBP-DISCover "Croplands"; (c) SYNMAP "Croplands"; (d) IGBP-DISCover "Croplands+Grasslands"

SYNMAP's "croplands" map (Figure 3c) 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 3e)

gives the most comparable pattern of cropland distribution amongst all the examined satellite land cover datasets, it too overestimates cropland fractions by 50%.

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].

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; *Herold et al*., 2008; *Wu et al*., 2008]. *Loveland and Belward* [1998] pointed out that in the IGBP dataset, 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 4).

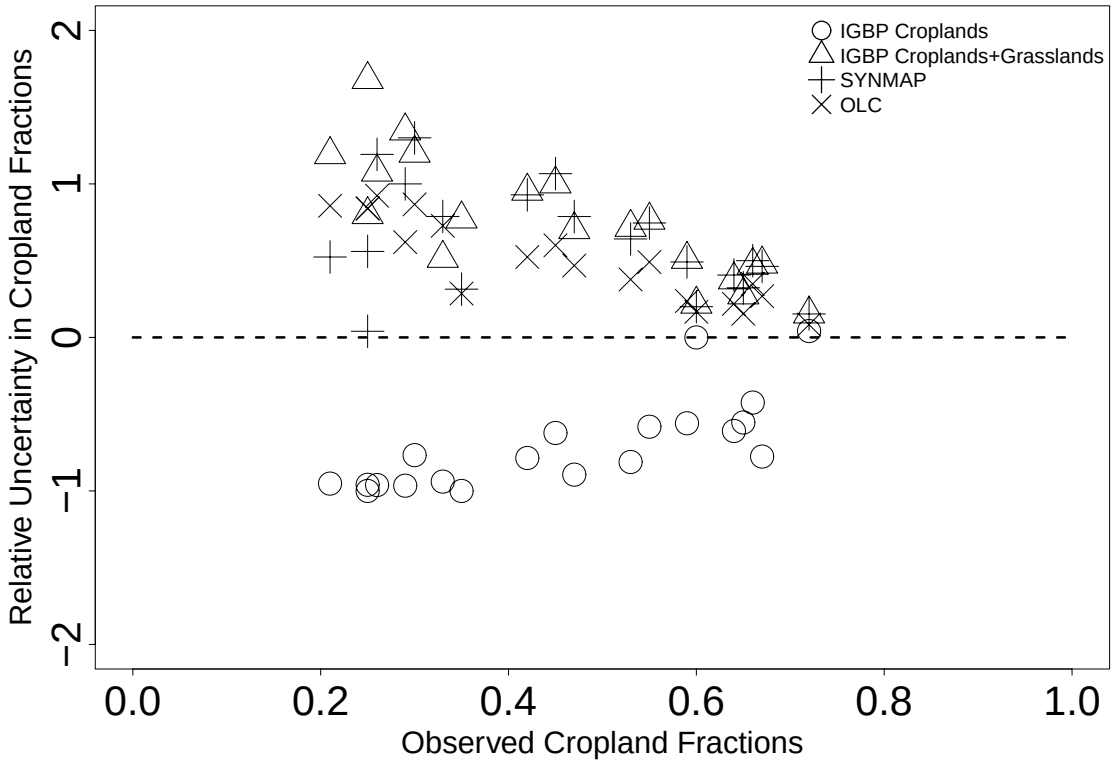

Figure 4. Comparison of relative uncertainties in cropland fractions. Relative uncertainties per county are calculated by dividing the differences in cropland fractions (i.e. satellite-derived fractions minus census data-derived fractions) with the census data-derived fractions.

Chapter 4 Constraining estimates generated by a biospheric model

4.1 Introduction

For this part of the study, carbon uptake estimates generated by a biospheric model are constrained by additional carbon cycle observations recorded by agricultural census data. Specifically, we shall focus on: i) the conversion of census data into additional source of carbon cycle observations; and ii) the inversion that examines the value of census data in constraining the initial modeled outputs in this chapter.

4.2 Method

4.2.1 Calculating *GPP* from crop attributes and agricultural census data

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*):

$$NPP = \sum_{i=1}^{N} \frac{P_i \times (1 - MC_i) \times C}{HI_i \times f_{AGi} \times HA_i} \tag{3}$$

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 above ground biomass (dimensionless); C converts harvested mass to carbon mass (approximately 0.45 g C g^{-1}) and is assumed constant between different crops; f_{AGi} is the fraction of production allocated above ground (dimensionless). Since f_{AGi} is defined as the ratio *shoot*:(*root* + *shoot*), it is closely related to the root:shoot ratio: $f_{AG} = \frac{shoot}{root + shoot} = \frac{1}{(root : shoot + 1)}$.

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):

$$NPP = GPP - R_a \tag{4}$$

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

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]:

$$GE = 1 - \frac{R_a}{GPP} \tag{5}$$

By combining eqs. (4) and (5), *GPP* can be expressed as a function of *NPP* and *GE*:

$$GPP = \frac{NPP}{(1 - \frac{R_a}{GPP})} = \frac{NPP}{GE} \tag{6}$$

Combining eqs. (6) and (3):

$$GPP = \sum_{i=1}^{N} \frac{P_i \times (1 - MC_i) \times C}{HI_i \times f_{AG\,i} \times HA_i \times GE_i} \tag{7}$$

The agricultural census data provide input for P and HA, and eq. (7) 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 4.

VPRM simulates hourly values of gross ecosystem exchange (GEE), details of which can be found in the next section. GPP estimates derived from eq. (7) can thus be compared against the model by summing GEE over an entire year:

$$\begin{aligned} GPP &= NPP + R_a \\ &\approx \sum^{1year} GEE_t \end{aligned} \tag{8}$$

Table 4. Crop specific attributes used for calculating *GPP* estimates and their associated uncertainties

Crop	Attributes	Values	Source
Winter wheat	Moisture content (*MC*)(%)	13 to 23	*OMAFRA*, 2009
	Harvest index (*HI*)	0.32 to 0.6	*de Jong et al*., 2001; *Falk et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.82 to 0.87	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.37 to 0.79	*Amthor*, 1989
Spring wheat	Moisture content (*MC*)(%)	13 to 23	*OMAFRA*, 2009
	Harvest index (*HI*)	0.32 to 0.6	*de Jong et al*., 2001; *Falk et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.82 to 0.87	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.37 to 0.79	*Amthor*, 1989
Grain corn	Moisture content (*MC*)(%)	11 to 22	*OMAFRA*, 2009
	Harvest index (*HI*)	0.45 to 0.55	*O'Neill*, 2005; *Bolinder et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.89 to 0.92	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.31 to 0.83	*Amthor*, 1989
Fodder corn	Moisture content (*MC*)(%)	65 to 70	*OMAFRA*, 2009
	Harvest index (*HI*)	0.55 to 1	*O'Neill*, 2005; *Bolinder et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.89 to 0.92	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.31 to 0.83	*Amthor*, 1989
Barley	Moisture content (*MC*)(%)	15 to 20	*OMAFRA*, 2009
	Harvest index (*HI*)	0.32 to 0.54	*Bolinder et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.63 to 0.69	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.51 to 0.52	*Amthor*, 1989
Soy	Moisture content (*MC*)(%)	14	*OMAFRA*, 2009
	Harvest index (*HI*)	0.23 to 0.4	*Morrison et al*., 1999; *Rollefson et al*., 2004; *Bolinder et al*., 2007
	Fraction of production allocated above ground (f_{AG})	0.68 to 0.89	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.25 to 0.55	*Amthor*, 1989

Table 4. Cont'd

Crop	Attributes	Values	Source
Hay	Moisture content (*MC*)(%)	12 to 18	*OMAFRA*, 2009
	Harvest index (*HI*)	1	*Lobell et al*., 2004
	Fraction of production allocated above ground (f_{AG})	0.52 to 0.77	*Bolinder et al*., 2007
	Growth efficiency (*GE*)	0.51 to 0.67	*Amthor*, 1989

4.2.2 Vegetation Photosynthesis and Respiration Model (VPRM)

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].

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):

$$GEE = \lambda \times T_{scale} \times P_{scale} \times W_{scale} \times \frac{1}{(1 + \frac{PAR}{PAR_0})} \times PAR \times EVI \tag{9}$$

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.

R is represented as a linear function of the ambient temperature given by:

$$R = \alpha \times T + \beta \tag{10}$$

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; $T = T_{low}$ when $T \leq T_{low}$ to account for the persistence of soil respiration in winter when air temperatures are very cold but soils remain warm.

The resulting full VPRM model equation is:

$$NEE = -\lambda \times T_{scale} \times P_{scale} \times W_{scale} \times \frac{1}{(1 + \frac{PAR}{PAR_0})} \times PAR \times EVI + \alpha \times T + \beta \tag{11}$$

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 Table 2 of the *Mahadevan et al*. [2008] paper. For validation results, please refer to Table 3 of the same paper. 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].

NEE, *GEE*, and *R* estimates generated by the VPRM are gridded on a grid of 1/4º

longitude by 1/6º 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*:

$$NEE = -GEE + R = \sum_k f_k (-GEE_k + R_k) \tag{12}$$

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 3.1.1). 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.

4.2.3 Uncertainty analysis of *GPP* estimated from crop attributes and agricultural census data

Monte Carlo simulation (MCS) is one of the most common stochastic methods used for uncertainty analysis [*IPCC*, 2000; *Winiwarter and Rypdal*, 2001; *Ogle et al*., 2003; *Ramirez et al*., 2008; *Del Grosso et al*., 2010]. In a simulation that generates results from a model, values of input parameters are randomly varied and selected according to each parameter's own unique probability distribution function (PDF). For many runs, each time with a different set of inputs, many possible output values can in turn be described by a PDF; and from which associated uncertainties of the model can be estimated. When compared to the conventional analytical approach for uncertainty analysis, the MCS has several additional attributes [*IPCC*, 2000]: i) it can deal with PDFs of any physically possible shape and width; ii) it can handle varying degrees of correlation and co-variation between parameters; and iii) it can be easily implemented for both simple or complex models.

In calculating carbon uptake in crops from agricultural census data and crop attributes with eq. (3), we adopted a Monte Carlo approach to assess the *GPP*s' associated uncertainties and their sensitivity to the assumed parameters in eq. (7), as recommended by the IPCC Good Practice Guidance [*IPCC*, 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 4 summarizes the ranges of values for the uniform distributions for each crop-specific attribute. 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.

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*):

$$HA_i = n_i a \tag{13}$$

where n_i is the number of unit harvested areas in a particular county. Applying the same

logic, the total provincial harvested area (HA_{prov}) can be considered as a sum of numerous county harvested areas.

$$HA_{prov} = \sum_i HA_i = \sum_i n_i a = Na \tag{14}$$

where the total provincial harvested area is comprised 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. Assuming that 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:

$$\sigma^2_{HA_i} = n_i \sigma^2 \tag{15}$$

while the variance at the provincial level would be:

$$\sigma^2_{HA_{prov}} = \sum_i \sigma^2_{HA_i} = \sum_i n_i \sigma^2 = N\sigma^2 \tag{16}$$

Combining eqs. (13) and (14):

$$\sigma^2_{HA_i} = \frac{n_i}{N} \sigma^2_{HA_{prov}} \tag{17a}$$

$$\sigma_{HA_i} = (\sqrt{\frac{n_i}{N}}) \sigma_{HA_{prov}} \tag{17b}$$

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 $\sigma_{HA_{prov}}$, 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%.

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 $\frac{\sigma_{Prod_{prov}}}{\overline{x}_{Prod_{prov}}}$, where $\overline{x}_{Prod_{prov}}$ is taken to be the provincial crop production, the uncertainty (standard deviation) of the county level crop production is, therefore:

$$\sigma_{Prod_i} = \left(\sqrt{\frac{n_i}{N_{natl}}}\right)\sigma_{Prod_{prov}} \qquad (18)$$

$$\text{where } \sigma_{Prod_{prov}} = CV \times \overline{x}_{Prod_{prov}}$$

In this study, an average of 3% error was used, and eq. (16) follows the same logic that led to eq. (17b) above.

4.2.4 Bayesian inversion to optimize *GPP* estimates

When modeling physical systems, model parameters and observations go hand in hand. If we are making predictions of what should be observed in the system using given parameters, we are dealing with a forward problem. In contrast, if we are making inferences about the model parameters using actual observations, we are working on an inverse problem [*Tarantola*, 1987]. To further clarify this concept, consider an outburst of a seismic event. As geophysicists, if we are interested in determining the arrival times of seismic waves at a particular seismic station, given that the location of the epicenter, the time of the event and the wave velocities are perfectly known, we are then tackling a

forward problem. On the other hand, if we wish to determine the location of the epicenter from the wave arrival times at a particular seismic station, we are then dealing with an inverse problem [*Tarantola*, 1987].

Rodgers [2000] pointed out that by applying the Bayesian approach in an inverse problem, in which we have some prior understanding about some quantity, we can then update our understanding in light of new information. Thus, in order to acquire insights on how the agricultural census data-derived *GPP* improves knowledge about the VPRM-modeled *GPP*, we performed a Bayesian inversion [*Rodgers,* 2000]. With notations borrowed from *Rodgers* [2000], census data-derived *GPP* can be linearly related to the VPRM-modeled *GPP* through:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon} \tag{19}$$

where **y** is the vector of census data-derived *GPP*, with each element representing a value for each census division; **x** is the vector of VPRM *GPP*, with each element signifying the modeled *GPP* estimated in each $1/6^{o}$ by $1/4^{o}$ grid cell; **K** is the matrix that relates the census data *GPP* with VPRM *GPP* and has elements that denotes the fraction of a census division covered by each $1/6^{o}$ by $1/4^{o}$ grid cell; **ε** is an error vector accounting for uncertainties in both the census-based estimates and VPRM simulations.

Following Bayes' theorem, generalized for the vector case, prior understanding of *GPP* from the VPRM and new information of *GPP* from the census data can be linked as:

$$P(\mathbf{x} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \tag{20}$$

If we assume that all probabilities above are Gaussian, then $P(\mathbf{y} \mid \mathbf{x})$, $P(\mathbf{x})$ and $P(\mathbf{x} \mid \mathbf{y})$ can be expressed as:

$$-2\ln P(\mathbf{y}\mid\mathbf{x}) = (\mathbf{y}-\mathbf{K}\mathbf{x})^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y}-\mathbf{K}\mathbf{x}) + c_1 \tag{21a}$$

$$-2\ln P(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_{prior})^T \mathbf{S}_{prior}^{-1}(\mathbf{x}-\mathbf{x}_{prior}) + c_2 \tag{21b}$$

$$-2\ln P(\mathbf{x}\mid\mathbf{y}) = (\mathbf{x}-\mathbf{x}_{post})^T \mathbf{S}_{post}^{-1}(\mathbf{x}-\mathbf{x}_{post}) + c_3 \tag{21c}$$

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 4.2.3, Table 4); $\mathbf{x}_{prior}$ refers to the a priori *GPP* estimates simulated by the VPRM; $\mathbf{S}_{prior}$ is the associated error covariance matrix of the VPRM estimates whose diagonals are derived by multiplying VPRM generated flux in each model grid cell to percentage uncertainties determined from comparing modeled and observed fluxes for soy and corn reported in Table 3 of *Mahadevan et al*. [2008]. Then an average is taken between the soy and corn percentage uncertainties and multiplied with the VPRM-generated primary productions to determine the diagonals of the $\mathbf{S}_{prior}$ matrix. Off-diagonal elements of both error covariance matrices are assumed to be zero.

Given that the probability of the observed census data-derived data is constant we can ignore $P(\mathbf{y})$ and determine $P(\mathbf{x} \mid \mathbf{y})$ alternatively by substituting eqs. (21a) and (21b) into eq.(20):

$$-2\ln P(\mathbf{x}\mid\mathbf{y}) = (\mathbf{y}-\mathbf{K}\mathbf{x})^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y}-\mathbf{K}\mathbf{x}) + (\mathbf{x}-\mathbf{x}_{prior})^T \mathbf{S}_{prior}^{-1}(\mathbf{x}-\mathbf{x}_{prior}) + c_4 \tag{22}$$

Now, if we relate eq. (21c) and eq. (22) by equating the quadratic terms $\mathbf{x}^T\mathbf{x}$ in $\mathbf{x}$, we would arrive at:

$$\mathbf{x}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}\mathbf{x} + \mathbf{x}^T\mathbf{S}_a^{-1}\mathbf{x} = \mathbf{x}^T\mathbf{S}_{post}^{-1}\mathbf{x} \tag{23}$$

$$\mathbf{S}_{post} = (\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{S}_{prior}^{-1})^{-1}$$

Likewise, if we relate eq. (21c) and eq. (22) via equating $\mathbf{x}^T$ in $\mathbf{x}$, we would get:

$$(-\mathbf{Kx})^T \mathbf{S}_{\varepsilon}^{-1}(\mathbf{y}) + (\mathbf{x})^T \mathbf{S}_{\mathrm{a}}^{-1}(-\mathbf{x}_{prior}) = \mathbf{x}^T \mathbf{S}_{post}^{-1}(-\mathbf{x}_{post}) \quad (24)$$

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

Eventually, the expected values of the posterior *GPP* estimates expressed as $\mathbf{x}_{post}$ is therefore:

$$\mathbf{x}_{post} = (\mathbf{K}^T \mathbf{S}_{\varepsilon}^{-1}\mathbf{K} + \mathbf{S}_{prior}^{-1})^{-1}(\mathbf{K}^T \mathbf{S}_{\varepsilon}^{-1}\mathbf{y} + \mathbf{S}_{prior}^{-1}\mathbf{x}) \quad (25)$$

and accordingly, the uncertainty of $\mathbf{x}_{post}$, expressed as the posterior error covariance matrix $\mathbf{S}_{post}$, is eq.(23):

4.3 Results and discussions

4.3.1 Monte Carlo uncertainty analysis for agricultural census-derived *GPP* estimates

Several sources of uncertainties are present when estimating *GPP* from agricultural census data and crop attributes. Figure 5 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 5). As seen in Figure 5, 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.

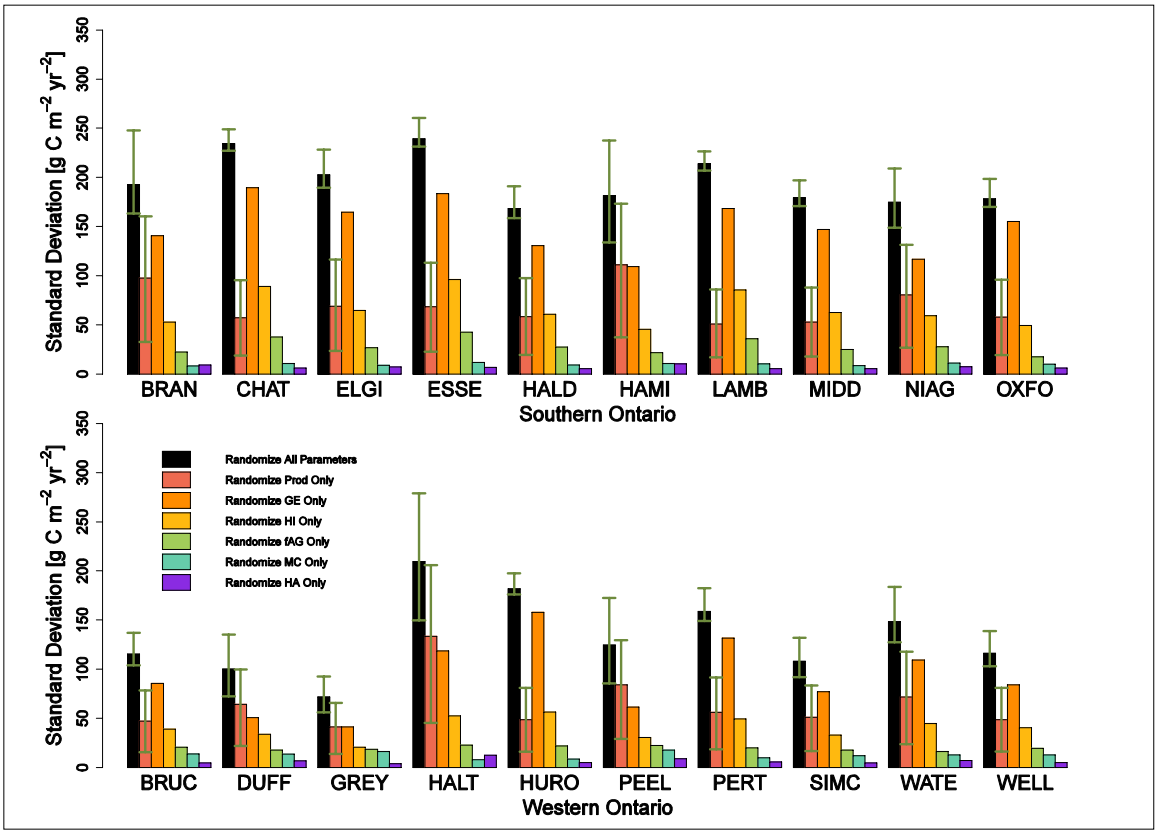

Figure 5. Results of Monte Carlo analysis of uncertainties in census data-derived *GPP* for southwestern Ontario. Different colors illustrate how randomizing different parameters in the crop attribute-based model influence the uncertainties in GPP estimates. "Prod" represents harvested production; "GE" represents growth efficiency; "HI" represents harvest index; "fAG" represents fraction allocated above ground; "MC" represents moisture content; and "HA" represents harvested area. 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%.

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* based on ^{14}C-labelling is dependent on the time of day the labeling takes place because crops would re-fixate respired CO_2 depending on time. In addition, the crop's stage of development matters as well. "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" [*Amthor*, 1989]. Under such limitations, large discrepancies in *GE* values can therefore be expected. When randomizing *GE* alone in eq. (3), the average uncertainty in *GPP* across all counties is 127 g C m^{-2} yr^{-1}.

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}.

By definition, *HI* is equal to the seed yield divided by above ground biological yield [*Donald and Hamblin*, 1976], where biological yield "includes the total dry matter the plant produces above ground" [*Prince et al*., 2001]. 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}.

4.3.2 Comparison of estimated *GPP*

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

6).

Comparisons between the agricultural data-derived *GPP* and VPRM-modeled *GPP* are shown in Table 5 and Figure 6. In Figure 6a, 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: a) 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; b) the current version of land cover dataset used by the model is not representative of the landscape; and c) the types of crops used in calibrating VPRM parameters may not fully resemble those grown on Ontario's croplands.

To investigate point c), we examined the sensitivity of crop-based *GPP* on the crop types included in the calculation. As seen in Table 7, 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*.

Table 5. Comparison of *GPP* estimates and their associated uncertainties over southwestern Ontario. 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 4.2.3 of the main text for details. The first 4 letters of the counties' names, as indicated in Figure 1 are used as abbreviations.

Counties	GPP [g C m^{-2} yr^{-1}]			Associated uncertainties [g C m^{-2} yr^{-1}]			
	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**

*Removed due to a lack of croplands in the IGBP-DISCover satellite land cover map linked to the VPRM (Figure 3b)

Table 6. Reported *GPP* estimates in oilseeds and grain crops from this study and other studies.

Source	Study location	Study Period	Crop types	Approach applied	Average *GPP* [g C m^{-2} yr^{-1}]
This study					
-	Ontario, Canada†	2004	Grain corn, hay, soybean, winter wheat	Integrated surface flux modeling - VPRM	~1122 (range: 295-1627)
-	Ontario, Canada†	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]

Table 6. Cont'd

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

†Only southwestern Ontario was considered in this study. Census divisions in southwestern Ontario (Figure 1) include: 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 due to a lack of croplands in the IGBP-DISCover satellite land cover map linked to the VPRM (Figure 3b)

Table 7. Comparison of average *GPP* estimates when a subset of crops is selected from the crop census data. 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]

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

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.

4.3.3 Bayesian inversion to optimize *GPP* estimates

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 5 and Figure 6. 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 6c). This result is perhaps not surprising, given the fact that uncertainties in the census data-derived *GPP* are comparable to those in VPRM (Figure 6, Table 5). 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 4.2.2). We have already pointed out in the previous section specific deficiencies of the VPRM simulations used in this study. Due to uncertainties in the land cover dataset 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.

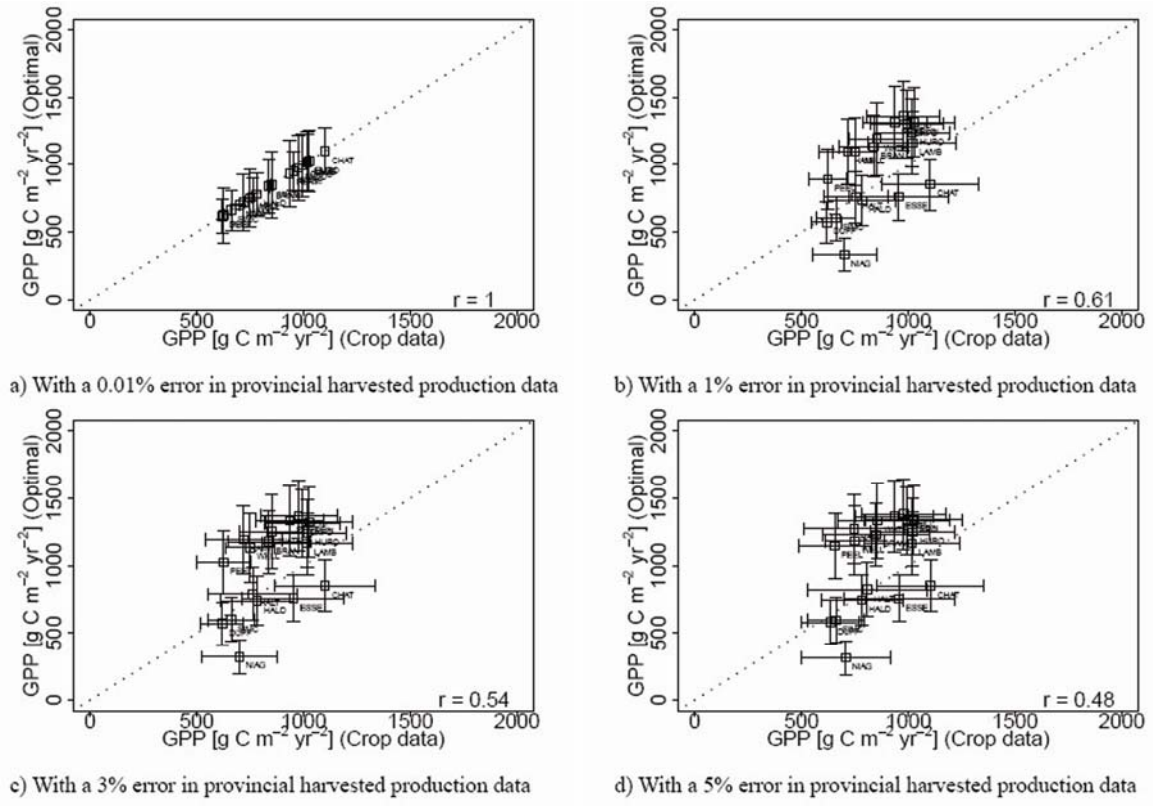

Figure 6. Comparison of *GPP* values with varying degrees of uncertainties in the harvested production data (a) With a 0.01% error; (b) With a 1% error; and (c) With a 3% error; and d) With a 5% error. The correlation coefficient (r) is also shown. Error bars associated with the crop census-derived *GPP* are constructed from Monte Carlo analysis (Section 4.2.3; Figure 5). The error bars associated with VPRM simulations are constructed via a method described in Section 4.2.2.

Chapter 5 Summary and future studies

5.1 Summary and conclusion

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.

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 datasets 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 [*Wulder et al*., 2004; *Jung et al*., 2006; *Herold et al*., 2008; *Wu et al*., 2008] 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.

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 due to imperfect knowledge of *HI* and *P* (Figure

4).

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.

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.

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 based on 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].

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 [*Carille 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.

5.2 Future studies

Based on what we learned from this study, issues below can be addressed in future follow-up studies:

- Assess how different prior PDFs would affect the inversion process;
- Re-calibrate and re-validate of the VPRM parameters within the Canadian landscape and re-run the model for a more in-depth terrestrial carbon cycle study;
- Run an array of regional-scale biospheric models that simulate the carbon exchange among Canadian cropland; and
- Determine the critical threshold uncertainty in the census data that best constrains outputs from the different models.

References

Amthor, J. S. (1989), Respiration and crop productivity, Springer Verlag, New York, N. Y.

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 50B,* 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), 2415-2435, 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.

Caldeira, K., and M. E. Wickett (2003), Oceanography: Anthropogenic carbon and ocean pH, *Nature, 425* (6956), 365, doi:10.1038/425365a.

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, Eastern Cereal and Oilseed Res. Centre (ECORC), Agric. and Agri-Food Can., Ottawa, Ont.

Del Grosso, S. J., S. M. Ogle, W. J. Parton, and F. J. Breidt (2010), Estimating uncertainty in N_2O emissions from US cropland soils, *Global Biogeochem. Cycles*, *24*, GB1009, doi: 10.1029/20099GB003544.

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: 10.1007/1-4020-4166-7_14.

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, in ONWheat Magazine Res. Ed., edited by E. Fletcher, pp. 16-17, Ont. Wheat Producers' Marketing Board, Guelph, Ont.

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, J. W. Munger, 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.

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 CO2 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 CO2 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.

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 CO2-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*(1), 103-109.

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.

IPCC (2000) Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories. Institute for Global Environmental Strategies (IGES), Kanagawa, Japan.

IPCC (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 University Press, Cambridge, UK and New York, NY, USA.

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, 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, *Global Ecol. Biogeogr., 12,* 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.

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*(1)*,* L01802, doi:10.1029/2004GL021127.

Lin, J. C., M. R. Pejam, E. Chan, S. C. Wofsy, E. W. Gottlieb, H. Margolis, and J. H. McCaughey (2011), How can uncertainties in simulating biospheric carbon fluxes be attributed to different factors, *Global Biogeochem. Cycles.*, doi:10.1029/2010GB003884, in press.

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*(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 CO2 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(5)*,* 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/agronj2000.924780x.

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.

Odum, E.P., (1971), Fundamentals of ecology, 3rd ed, Saunders, Philadelphia and London.

Ogle, S. M., F. J. Breidt, M. D. Eve, and K. Paustian (2003), Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agricultural lands between 1982 and 1997, *Global Change Biol.*, *9*, 1521-1542, doi: 10.1046/j.1529-8817.2003.00683.x.

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

Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) (1988), Land Use [computer file]

Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) (2009), *Agronomy Guide for Field Crops*, *Publication 811*, Ont. Ministry of Agric., Food and Rural Aff., Guelph, Ont.

Ontario Ministry of Natural Resources (2002), Ontario Provincial-Scale Land Cover, Ontario Min. of Nat. Res., Peterborough, Ont.

Orr, J. C., et al. (2005), Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature, 437*, 681-686, doi:10.1038/nature04095.

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., 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.

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..

Ramirez, A., C. de Keizer, J. P., Van der Sluijs, J. Olivier, and L. Brandes (2008), Monte Carlo analysis of uncertainties in the Netherlands greenhouse gas emission inventory for 1990-2004, *Atmos. Env.*, *42*, 8263-8272, doi: 10.1016/j.atmosenv.2008.07.059.

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, Natl. Res. Council Can., Ottawa, Ont.

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.

Sellers, P. J., L. Bounoua, G. J. Collatz, D. A. Randall, D. A. Dazlich, S. O. Los, J. A. Berry, I. Fung, C. J. Tucker, C. B. Field, and T.G. Jensen (1996), Comparison of radiative and physiological effects of doubled atmospheric CO_2 on climate. *Science*, *271*, 1402-1406, doi:10.1126/science.271.5254.1402.

Statistics Canada (2004), *Field crop reporting series, 83*(8), Govt. of Can., Ottawa, Ont.

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

Suyker, A. E., S. B. Verma, G. G. Burba, T. J. Arkebauer (2005), Gross primary

production and ecosystem respiration of irrigated maize and irrigated soybean during a growing season, *Agric. and For. Meteorol., 131,* 180-190, doi:10.1016/j.agrformet.2005.05.007.

Tarantola, A. (1987), Inverse problem theory: methods for data fitting and model parameter estimation. Elsevier Science Publishers B. V., Amsterdam, Netherlands

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*(4)*,* 666-684, doi:10.1111/j.1365-2486.2005.00936.x.

Vogelmann, J. E., D. Helder, R. Morfitt, M. J. Choate, J. W. Merchant, 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.

Winiwarter, W., and K. Rypdal (2001), Assessing the uncertainty associated with national greenhouse gas emission inventories: a case study for Austria, *Atmos. Env.*, *35*, 5425-5440.

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, L. Ediger (2009), Modeling gross primary productivity for winter wheat-maize double cropping system using MODIS time series and CO2 eddy flux tower data, *Agric. Ecosyst. Environ., 129,* 391-400*,* doi:10.1016/j.agee.2008.10.017.