DATA-DRIVEN SPATIAL MODELING OF HISTORIC AND FUTURE LAND CHANGE AT GLOBAL SCALE

BY
PRASANTH MEIYAPPAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Atmospheric Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Atul K Jain, Chair and the Director of Research
Professor Stephen W Nesbitt
Professor Ryan L Sriver
Doctor Michael Dalton

ABSTRACT

Assessing the historic and future impacts of land-use and land-cover change (LULCC) on climate requires spatially and temporally explicit data sets on LULCC spanning several decades to centuries, because climate change is a long-term problem. Though remote sensing data provides a globally consistent picture of land cover, these data are only available from the past four decades. Therefore, existing LULCC reconstructions are modeled estimates that combine remote sensing data with relatively coarser-resolution inventory statistics that covers longer historical period. The uncertainties in modeling assumptions, and limited availability and inconsistencies across inventory datasets among other reasons introduce uncertainties in LULCC reconstructions. These uncertainties not only limit our ability to model future LULCC, but also translate as uncertainties in both historic and future environmental assessments.

The objectives of my PhD work are as follows: (1) systematically investigate the causes of uncertainties in existing historical LULCC datasets, (2) test the sensitivity of LULCC quantification uncertainty in estimating CO_2 emissions from LULCC (historic and future) using a process-based land-surface model, the Integrated Science Assessment Model (ISAM), (3) compare the relative uncertainties from various drivers (e.g. LULCC datasets, model processes e.g. nitrogen cycle, environmental factors such as climate) in estimating historic and future LULCC emissions, and (4) explore statistical techniques to model future LULCC that takes into account the uncertainties in quantifying the spatial and temporal patterns of LULCC, and (5) as a case-study, identify a key regional hotspot of historic LULCC quantification uncertainty (here, India), and reduce uncertainty through improved understanding of the dynamics and drivers of land change in the case-study region. I address the above goals by integrating land-surface modeling (ISAM), remote sensing and GIS, data collected through ground transects, and geospatial data on socioeconomics.

ISAM simulations show that the estimated net global emissions from LULCC (mean and range) across three different historical LULCC reconstructions are 1.88 (1.7 to 2.21) GtC/yr for the 1980's, 1.66 (1.48 to 1.83) GtC/yr for the 1990's, and 1.44 (1.22 to 1.65) for the 2000's. The estimates are higher than other published estimates that range from 0.80 to 1.5 GtC/yr for the

1990's and 1.1 GtC/yr for the 2000's. These results are higher than other published estimates because they include the effects of nitrogen limitation on regrowth of forests following wood harvest and agricultural abandonment. The estimated LULUC emissions for the tropics are 0.79±0.25 for the 1980's, 0.78±0.29 for the 1990's and 0.71±0.33 GtC/yr for the 2000's, and for the non-tropics regions are 1.08±0.52, 0.90±0.19 and 0.69±0.12 GtC/yr for the three decades. The model results indicate that failing to account for the nitrogen cycle underestimates LULCC emissions by about 40% globally (0.66 GtC/yr), 10% in the tropics (0.07 GtC/yr) and 70% in the non-tropics (0.59 GtC/yr). If LULCC emissions are higher than assessed, it means fossil fuel emissions would have to be even lower to meet the same mitigation target.

Extending ISAM simulations to the 21st century resulted in two key insights. First, nitrogen limitation of CO_2 uptake is substantial and sensitive to nitrogen inputs. In ISAM, excluding nitrogen limitation underestimated global total LULUC emissions by 34-52 PgC (~21-29%) during the 20th century and by 128-187 PgC (90-150%) during the 21st century. The difference increases with time because nitrogen limitation will progressively down-regulate the magnitude of CO_2 fertilization effect on regrowing forests, due to decreasing supply of plant-usable mineral nitrogen. Second, historically, the indirect effects of anthropogenic activity through environmental changes in land experiencing LULCC (indirect emissions) are small compared to direct effects of anthropogenic LULCC activity (direct emissions). As a result, including or excluding indirect emissions had a minor influence on the estimated total LULUC emissions historically. In contrast, the indirect LULCC emissions for the 21st century are a much larger source to the atmosphere, in simulations with nitrogen limitation. This is because of the gradual weakening of the photosynthetic response to elevated (CO_2) caused by nitrogen limitation. Therefore, what fluxes are including in LULCC emissions across different models is a crucial source of uncertainty in future LULCC emissions estimates.

A detailed investigation of the sensitivity of different global-scale LULCC modeling techniques show that land use allocation approaches based solely on previous land use history (but disregarding the impact of driving factor), or those based on mechanistically fitting models for the spatial processes of land use change do not reproduce well long-term historical land use patterns. With an example application to the terrestrial carbon cycle, I show that such inaccuracies in land use allocation can translate into significant implications for global

environmental assessments. In contrast to previous approaches, I present a statistical land use downscaling model and show that the model can reproduce the broad spatial features of the past 100 years of evolution of cropland and pastureland patterns. Therefore, the modeling approach and its evaluation provide an example that can be useful to the land use, Integrated Assessment, and the Earth system modeling communities.

ACKNOWLEDGEMENTS

I thank Prof. Atul Jain for giving me the opportunity to come to UIUC, and for being my advisor during the last five years of my PhD endeavor. I thank him very much for all his valuable time and efforts towards me.

To former members Dr. Rahul Barman and Dr. Bassil El-Masri, and current group members Dr. Yang Song, and Shije Shu, thank you all for significantly helping with my research as and when need be, and for being great friends to me.

I sincerely thank my thesis committee members: Prof. Ryan Sriver (UIUC), Prof. Stephen Nesbitt (UIUC), and Dr. Michael Dalton (NOAA), for their valuable time and guidance, and for sharing their thoughts with me. I especially offer my deepest gratitude to Dr. Dalton, who has been a tremendous mentor to me during my visits to NOAA. Since then, the background that Dr. Dalton provided me has helped me come a long way in the pursuit of my research work. I also offer my deepest gratitude to Dr. Sriver for taking sincere interest in my research and career, and offering support on these fronts.

I have also been extremely fortunate to work with Dr. Brian O'Neill and colleagues at NCAR, and Dr. Parth S. Roy and colleagues from India. I enjoyed working with both the teams, and I look forward to continued collaboration in the future. Dr. Roy's knowledge on remote sensing and societal aspects of land change in India are beyond amazing. I admire him both as a scientist and as an individual. He likely has spent more time discussing research with me that any of his PhD students cumulated over the past 2 years. I am grateful to him for his perennial support with my research.

Finally, my deepest regards towards my parents, who have devoted much of their lifetime in loving my brother, and me, and have ensured the best of opportunities for us. I dedicate this work to them.

TABLE OF CONTENTS

CHAPTER 1

Introduction

1.1 Overall Objectives and Content

Human activities have transformed natural ecosystems into managed areas in almost every part of the world. At present, nearly 40% of the Earth's ice-free land surface is used for agricultural activities, all of which had previously been covered by natural vegetation. Such large-scale changes in land cover affect regional and global climate through biogeophysical and biogeochemical pathways.

Assessing the historic and future impacts of land-use and land-cover change (LULCC) on climate requires spatially and temporally explicit data sets on LULCC spanning several decades to centuries, because climate change is a long-term problem. Though remote sensing data provides a globally consistent picture of land cover, these data are only available from the past four decades. Therefore, existing LULCC reconstructions are modeled estimates that combine remote sensing data with relatively coarser-resolution inventory statistics that covers longer historical period. The uncertainties in modeling assumptions, and limited availability and inconsistencies across inventory datasets among other reasons introduce uncertainties in LULCC reconstructions. These uncertainties not only limit our ability to model future LULCC, but also translate as uncertainties in both historic and future environmental assessments. For example, quantifying CO_2 emissions from historical LULCC is the major source of uncertainty in the global carbon budget. Therefore, a systematic understanding of the sources of uncertainties in existing LULCC datasets is crucial to: (1) better assess the utility of a dataset (or its subset) to particular scientific application, (2) draw better-informed conclusions when used as an input for environmental assessments, and (3) to further reduce uncertainty in quantifying LULCC through iterative process.

Motivated by the above reasons, the goal of this study is to: (1) systematically investigate the causes of uncertainties in existing historical LULCC datasets, (2) test the sensitivity of LULCC quantification uncertainty in estimating CO2 emissions from LULCC (historic and future) using a process-based land-surface model, the Integrated Science Assessment Model

(ISAM), (3) compare the relative uncertainties from various drivers (e.g. LULCC datasets, model processes e.g. nitrogen cycle, environmental factors such as climate) in estimating historic and future LULCC emissions, and (4) explore statistical techniques to model future LULCC that takes into account the uncertainties in quantifying the spatial and temporal patterns of LULCC, and (5) as a case-study, identify a key regional hotspot of historic LULCC quantification uncertainty (here, India), and reduce uncertainty through improved understanding of the dynamics and drivers of land change in the case-study region. I address the above goals by integrating land-surface modeling, remote sensing and GIS, data collected through ground transects, and geospatial data on socioeconomics.

Broadly, the contents of this dissertation can be sub-divided into three parts. In the first part consisting of Chapters 2—4, I used existing global land use datasets (historic and future) to understand and quantify the sources of uncertainty among them (Chapter 2), and how these uncertainties translate as uncertainties in modeling CO_2 emissions from land-use and land-cover change (Chapters 3, 4). Having developed an understanding of the overall data and model uncertainties in these chapters, in the second part (Chapters 5) we (with collaborators) developed statistical models to predict the spatial patterns of land-use change. This is the first study to demonstrate the ability of a land change model to reproduce the past 100 years of evolution of spatial changes in agriculture at global scale. The model is currently being applied for predicting the future spatial land use patterns within the Integrated Assessment Modeling (IAM) group at NCAR, and in other multi-model comparison projects such as LUC4C. In the third part (Chapter 6), I take a more spatially detailed, but regional perspective to understand the dynamics and drivers of spatial patterns of LULCC in India. I presented each of these chapters as self-contained units, containing individual abstract, introduction, methods, results, discussion, and conclusions. Specific contents and objectives of individual chapters are as follows:

1. Chapter 2: Comparison of three different historical land use datasets (cropland, pastureland, wood harvest, and urban land) and developing algorithms to translate them into changes in land cover, consistent with the land surface representation of ISAM.
2. Chapter 3: Applying the three LULCC reconstruction developed in Chapter 2 within ISAM to examine the uncertainties in modeling CO_2 emissions due to uncertainties in

quantifying historical LULCC. Here, we not only explore LULCC data uncertainty, but also uncertainties in modeling key terrestrial processes, including the nitrogen cycle.

3. Chapter 4: I extend the work of Chapter 2 to estimate LULCC (consistent with land surface representation in ISAM) between 2005-2100 under various Representative Concentration Pathways (RCPs) of the IPCC CMIP5. Then by extending Chapter 3, I drive ISAM with these future LULUC datasets to quantify uncertainties in future CO_2 emissions from LULCC resulting from (1) differences in scenarios, (2) different LULCC activities represented in the model, and the sensitivity to the method of representation, (3) key structural and parameter uncertainty in model, including the representation of a nitrogen cycle, (4) uncertainties in modeling environmental factors (especially climate), and (5) different terminologies of what "LULCC emission" implies.
4. Chapter 5: I present the description and historical evaluation of the development of a spatial model of agricultural land use change at global scale. This analysis extends the work presented in Chapters 2 and 3, by evaluating the sensitivity of different land use reconstruction methodologies to estimating CO_2 emissions from LULCC.
5. Chapter 6: Here, I take a regional focus to improve our understanding of LULCC in India. There are two motivations to focus on India. First, India is a region where the average human pressure on land resources much exceeds the global average. The pressure is expected to further intensify in the future, thus being a global hotspot of land change. Second, from analysis presented in Chapter 2, we find that uncertainties in historical (late 20^{th} century) LULCC in India are much greater than other regions in South Asia. Therefore, there remains a potential to improve our understanding of historical LULCC in India (thereby reducing uncertainties).

In this chapter, I present estimates of various land-cover conversions in India at national scale between 1985 and 2005, based on a wall-to-wall analysis of high-resolution Landsat imageries. Using high-resolution biophysical and socioeconomic datasets combined with statistical models, I also investigated the drivers of key land-cover conversions in India. This understanding is essential to model LULCC at higher resolution (typically 1km lat/long) required for regional environmental assessments and land use planning. Current global datasets (as presented in Chapter 2) typically available at ~10km lat/long or coarser resolution do not

adequately capture the heterogeneity and fragmentation of India's landscapes. Note that drivers of LULCC vary with resolution due to scale dependencies.

Finally, in Chapter 7, I provided an overall summary, and the future direction of research presented in this dissertation.

Chapters 2-5 have already been published in peer-reviewed journals (see table in next page). Chapter 6 is currently under review for Regional Environmental Change.

Note on Supplementary/Appendix: For brevity, no supplementary/appendix material has been included with the dissertation. The supplementary text/figure/table numbers cited in each chapter corresponds to the online supplementary material (open-access) of the journal publication of respective chapters (see next table for chapter-wise journal publication).

List of journal and data publications categorized by chapter (… indicates two or more coauthors in between; Bold Chapter # indicates substantial contribution to the work).

Chapter #	Reference
2	**Meiyappan, P**., & Jain, A. K. (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Frontiers of Earth Science*, *6*(2), 122-139.
2	**Meiyappan, P**. Historical Land-Cover Change and Land-Use Conversions Global Dataset. Accessible from *NOAA"s National Climatic Data Center* (http://www.ngdc.noaa.gov). Dataset Identifier: gov.noaa.ncdc:C00814
3	Jain, A. K., **Meiyappan, P**., Song, Y., & House, J. I. (2013). CO_2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data. *Global Change Biology*, 19, 2893-2906. Nature News & Views | Science Daily
3	Jain, A. K., **Meiyappan, P**., & Richardson, T. (2013). Carbon emissions from land-use change: model estimates using three different data sets. *Land Use and the Carbon Cycle: Advances in Integrated Science, Management, and Policy*, Cambridge University Press, 241-258.
4	**Meiyappan, P**., Jain, A. K., & House, J. I. Increased influence of nitrogen limitation on CO_2 emissions from future land use and land-use change. *Global Biogeochemical Cycles*, 29, 1524-1548. Science Daily | U of I News Bureau
5	**Meiyappan, P**., Dalton, M., O'Neill, B. C., & Jain, A. K. Spatial modeling of agricultural land-use change at global scale. *Ecological Modelling*, 291, 152-174.
5	Alexander, P.,Prestle, R., Verburg, P.,…, Meiyappan, P. et al. Assessing uncertainties in land cover projections. (In review for *Nature Climate Change*).
5	Ren, X.,…, Meiyappan, P. et al. Avoided economic impacts of climate change on agriculture: Integrating a land-surface model (CLM) with a global economic model (iPETS). (Invited article for *Climatic Change*).

5	Lawrence, P.,…, Meiyappan, P. System description for the toolbox for human-earth system integration and scaling version 1.0 (THESIS 1.0). National Center for Atmospheric Research (NCAR) Technical Note.
6	**Meiyappan, P**., Roy, P. S., Sharma, Y. et al. Dynamics and determinants of land change in India: linking remote sensing to village socioeconomic. (In review for *Regional Environmental Change*).
6	**Meiyappan, P**., et al. Village level spatial database on Indian census for 1991 and 2001: A compendium of 200 variables from PCA and village directory. (In process for data dissemination through NASA Socioeconomic Data and Applications Center)
6	Roy, P. S.,…, Meiyappan, P., et al. Development of decadal (1985-1995-2005) land use and land cover database for India. *Remote Sensing*, 7, 2410-2430.
7	**Meiyappan, P**., Gao, J., Roy, P. S. et al. Estimating India's land change from 1950 to 2011 at 1km resolution. (Target: *Global Change Biology*).

CHAPTER 2

Three distinct global estimates of historical land-cover change and land-use conversions over a period of 200 years

2.1 Abstract

Earth's land cover has been extensively transformed over time due to both human activities and natural causes. Previous global studies have focused on developing spatial and temporal patterns of dominant human land-use activities (e.g. cropland, pastureland, urban land, wood harvest). Process-based modeling studies adopt different strategies to estimate the changes in land cover by using these land-use data sets in combination with a potential vegetation map, and subsequently use this information for impact assessments. However, due to unaccounted changes in land cover (resulting from both indirect anthropogenic and natural causes), heterogeneity in land-use/cover (LUC) conversions among grid cells, even for the same land-use activity, and uncertainty associated with potential vegetation mapping and historical estimates of human land-use result in land cover estimates that are substantially different compared to results acquired from remote sensing observations. Here we present a method to implicitly account for the differences arising from these uncertainties in order to provide historical estimates of land cover that are consistent with satellite estimates for recent years. Due to uncertainty in historical agricultural land use, we use three widely accepted global estimates of cropland and pastureland in combination with common wood harvest and urban land data sets to generate three distinct estimates of historical land-cover change and underlying LUC conversions. Hence, these distinct historical reconstructions offer a wide range of plausible regional estimates of uncertainty and the extent to which different ecosystems have undergone changes. The annual land cover maps and LUC conversion maps are reported at 0.5°×0.5° resolution and describe the area of 28 land cover types and respective underlying land-use transitions. The reconstructed data sets are relevant for studies addressing the impact of land-cover change on biogeophysics, biogeochemistry, water cycle, and global climate.

2.2 Introduction

Human activities have transformed natural ecosystems into managed areas in almost every part of the world (Foley et al., 2005; 2011). At present, nearly 40% of the Earth's ice-free land surface is being used for agricultural activities, all of which had previously been covered by natural vegetation (Ramankutty et al., 2008; Ellis et al., 2010). Such large-scale changes in land cover affect regional and global climate through biogeophysical (Bonan et al., 1992; Pielke et al. 2002; 2011; Feddema et al. 2005; Brovkin et al., 2006; Bala et al. 2007; Pitman et al. 2009, 2011; Findell et al., 2009) and biogeochemical (Jain and Yang 2005; Canadell et al. 2007; Bonan 2008; Jain et al. 2009; Pongratz et al., 2009; Shevliakova et al. 2009; Houghton et al., 2012) pathways.

Assessing the historical impacts of land-use/cover change (LUCC) at global scale (e.g. biogeophysical, biogeochemical, and climate effects) requires spatially and temporally explicit data sets on land cover and land-use/cover (LUC) conversions (replacement of one land cover type by another) spanning several hundred years. Though remote sensing data provides a globally consistent picture of land cover, these data are only available from the past four decades (Houghton et al. 2012). Hence, several studies (e.g. Ramankutty and Foley, 1999; Klein Goldewijk, 2001; Klein Goldewijk et al., 2006; Hurtt et al., 2006; 2011; Olofsson and Hickler, 2008; Pongratz et al., 2008; Klein Goldewijk et al., 2010; Klein Goldewijk et al., 2011;) have adopted different approaches in order to reconstruct spatially explicit data sets of dominant land-use activities (e.g. cropland, pastureland, urban land, wood harvest) covering several centuries. Typically, process-based modeling studies combine one or more of these land-use data sets with a map of potential vegetation (representing primary land cover in the absence of human activities) to estimate the changes in land cover. The method adopted to replace potential vegetation varies from simple proportional clearing (e.g. Jain and Yang, 2005; Pitman et al., 2009) to a rule-based approach based on several logical assumptions and prioritizations that best describe the trends associated with historical LUCC (e.g. Hurtt et al., 2006; 2011).

Hurtt et al. (2006) developed a Global Land-use Model (GLM) to provide historical estimates of LUCC and LUC conversions due to expansion of cropland and pastureland, shifting cultivation and wood harvest at 1° spatial resolution. An updated version of GLM framework has recently been used in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment

Report (AR5) to provide estimates of LUCC and LUC conversions among five simple classes (cropland, pastureland, urban land, primary land, and secondary land) at 0.5°×0.5° resolution annually from 1500 to 2100 (AD) (Hurtt et al., 2011). This includes historical input data covering the period 1500–2005 and data for the four Representative Concentration Pathways (RCP) scenarios (Moss et al., 2010) for the future (2005–2100). The LUCC and LUC conversion estimates are usually translated to the specific land cover classes suitable for use in a process-based model and subsequently used for impact assessments (e.g. Lawrence et al., 2012). Jain and Yang (2005) used a much simpler technique of superimposing the historical cropland data (based on Ramankutty and Foley 1999) on a 0.5°×0.5° potential vegetation map (with each grid cell occupied by one potential vegetation) to estimate the changes in land cover. Similar but varying methods for superimposing a common cropland and pastureland were adopted by each of the seven climate models that participated in an inter-comparison study aimed at understanding the historical impact of land-cover change (Pitman et al., 2009). These estimates have been used as inputs to terrestrial carbon models, dynamic vegetation models, and earth system models to assess the impacts of LUCC (e.g. Shevliakova et al., 2009; Yang et al., 2010; Lawrence et al., 2012) on biogeophysics and/or biogeochemistry. However, most of these previous studies have not considered land-cover change arising due to indirect anthropogenic (e.g. climate driven land-cover change) or natural disturbances like fires, blowdowns, and insect outbreaks. Several local- to national-scale studies have demonstrated their importance and ecological significance (e.g. Giglio et al., 2010; van der Werf et al., 2010; also see Lambin et al., 2003 and Foley et al., 2003). For example, according to Forest Resources Assessment (FAO 2006), 104 million hectares of forest on average were reported to be significantly affected each year by forest fire, pests (insects and disease), or climatic events such as drought, wind, snow, ice, and floods, with many countries missing this crucial information. In addition to differences arising from unaccounted land-cover change (indirect anthropogenic and natural causes), significant uncertainties could also arise due to heterogeneity associated with LUCC at temporal and spatial scales which cannot be captured using a rule-based approach of converting vegetation generalized at a regional or global scale. As a result, the global land cover estimated by most of the previous studies does not match estimates based on remote sensing data, a valuable tool in detecting several types of land-cover changes and land-cover modifications (subtle changes in land cover) that are difficult to map using other methods. For example, a comparison of forest area in 2005

from Hurtt et al. (2011) (estimated by combining information on primary and secondary land with a basemap which classifies each grid cell as either forest or non-forest based on potential vegetation biomass, as provided by Hurtt et al. 2006) and 500 m resolution Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 5 land cover data (Friedl et al. 2010) following International Geosphere-Biosphere Programme (IGBP) classification scheme (Loveland and Belward 1997) indicates pronounced differences in magnitude and spatial distribution (Figures 2.1(a) and (b)). Globally, Hurtt et al. (2011) estimated forest area was about 8×10^6 km^2 higher than the MODIS estimated value of 31.5×10^6 km^2 in 2005. Similarly, other studies also overestimated the global forest extent for the recent past at similar magnitudes (refer to Sect. 2.5). It is essential to reconcile such differences in estimates, especially in the context of studies addressing the biogeophysical impacts of land-cover change.

The objective of this study is to build upon and extend the approaches of previous studies in order to provide estimates of historical land-cover change (and underlying LUC conversions) that are consistent with satellite observations. We use a rule-based approach to assign priorities for converting land cover due to various human land-use activities. Multiple years of satellite data sets are used to quantify the differences in estimates that may be arising due to unaccounted land-cover change and heterogeneity associated with LUCC that cannot be captured using simple rules for clearing vegetation. These differences are used to constrain and accordingly adjust the priorities for changing land cover, thereby producing land cover maps consistent with satellite observations for recent years. The work presented here takes into account land-cover change due to four major land-use activities: 1) cropland expansion and abandonment, 2) pastureland expansion and abandonment, 3) urbanization, and 4) regrowth due to wood harvest. Due to uncertainties associated with historical agricultural land-use, we have used three global historical data sets of cropland and pastureland (refer step 1 in Sect. 2.3) in combination with a common data set for historical wood harvest and urban land, to produce three distinct estimates. The core products we generated were annual maps (at 0.5°×0.5° resolution) of land cover and LUC conversions starting from the pre-industrial year of 1765 until 2010 or before (based on the ending time of the three cropland and pastureland data sets). The annual land cover data sets are reported as area fractions of 28 land cover types (Table 2.1) for each 0.5°×0.5° grid cell and the annual LUC conversion maps are reported as the area converted for each of the 92 unique conversions possible (refer supplementary Table S1) among the 28 land cover types. The results

are compared with other recently published model results and data-based studies. Finally, the sources of uncertainties in the present study are discussed.

2.3 Methods

The method used to characterize historical land-cover change can be described in five steps: 1) Historical land-use change data sets are processed to suit this study; 2) Land cover map for the year 1765 are generated by combining potential vegetation map, cropland, pastureland, and urban land map for that year; 3) Land-cover change and LUC conversions starting with the 1765 land cover map are estimated using a rule-based approach for prioritizing LUC conversion for each of the four land-use activities; 4) Estimates from the previous step are compared with satellite data. Priorities are accordingly adjusted to correct for the differences; 5) Grassland, pastureland, and cropland estimates are separated into C_3/C_4 photosynthetic pathways.

Step 1: processing of historical land-use change data sets

The three different data sets on cropland and pastureland are based on: 1) HYDE 3.1 (Historical Database of the Global Environment) (Klein Goldewijk et al., 2011), 2) New pastureland estimates and updated cropland estimates based on Ramankutty and Foley (1999) (N. Ramankutty, personal communication, 2011), and 3) Regional estimates based on Houghton (2008). These three agricultural land-use data sets are henceforth referred to as HYDE, RF, and HH data, respectively. The urban land data set is from Klein Goldewijk et al. (2010). Historical wood harvest data are based on annual wood harvesting rates from Hurtt et al. (2011). RF and HH data are at an annual time scale. The decadal time resolution HYDE data was linearly interpolated to yield annual maps. All these data sets except HH data are gridded data sets at 0.5°×0.5° or finer resolution. Finer resolution data were aggregated to 0.5°×0.5° resolution. The HH data set provides the annual rate of deforestation/reforestation due to cropland, pastureland, wood harvest and shifting cultivation for ten regions (defined in Houghton et al., 1983) covering the entire globe, rather than by geographic details. HH regional data sets for cropland and pastureland resulting from deforestation were converted to gridded estimates using the LUC conversion estimates derived based on RF data. Additional details on the method used to spatialize HH data, details and processing of other data sets are available in supplementary text. The three land-cover change and LUC conversion estimates generated from this study

(henceforth referred to as ISAM-HYDE, ISAM-RF and ISAM-HH) based on three agricultural data sets (HYDE, RF, and HH) utilized the same wood harvest and urban land data. ISAM-HYDE, ISAM-RF, and ISAM-HH estimates extend to the year 2010, 2007 and 2005, respectively.

Step 2: land cover map of 1765

A land cover map for the year 1765 was generated as a reference map to track land-cover change and LUC conversions. We started with the global map of potential vegetation derived at 5 min spatial resolution by Ramankutty and Foley (1999). Fourteen of the 15 vegetation classes present in the potential vegetation map directly correspond to the potential land cover types used in this study (Table 2.1). The land cover classification used in this study is chosen to be consistent with the land cover types required for the Integrated Science Assessment Model (ISAM) (Jain and Yang, 2005; Yang et al., 2009; Yang et al., 2010) for which we originally produced these data sets. Mixed forest (which is not part of our land cover classification) from the potential vegetation map was reclassified into any one of the seven forest types by searching for dominant (greater than 70% of the area considered) forest type within a 4°×4° resolution window around the grid cell. The window size was increased until the requirements for dominant forest type were satisfied. Savanna (usually defined as tropical grasslands) present outside tropical regions was reclassified to other herbaceous types, using the method adopted for reclassifying mixed forest. Ramankutty and Foley (1999) assigned single potential vegetation to each 5 min grid cell from 1km DISCover satellite-based global land cover data (Loveland and Belward, 1997) even in grid cells where anthropogenic land cover was absent. In such grid cells, we used MODIS data (Friedl et al., 2010) for the year 2005 classified under IGBP classification scheme to reassign the grid cell area (currently occupied by either 100% forest or non-forest) to fractional area of forest and non-forest. The forest and non-forest types were determined using a combination of MODIS land cover data (Friedl et al., 2010) and the method adopted to reclassify mixed forest. This reduced the total area of forest in the potential vegetation map from 55.2×10^6 km^2 to about 48.6×10^6 km^2. An additional land cover class (water-covered areas) map was derived at 5 min resolution using MODIS land cover data (Friedl et al., 2010) for the year 2005, and was included in the potential vegetation map by proportional adjustment of potential vegetation areas.

Next, we aggregate the 5 min resolution potential vegetation map to 0.5°×0.5° resolution to yield the fractional areas of 15 land cover types within each grid cell. Hence, each grid cell in our potential vegetation map can be occupied by more than one type of natural vegetation. We assume water-covered areas to be constant for every year.

Finally, we derive the land cover map for the year 1765 by including the 1765 cropland and pastureland maps from RF and the urban land map (Klein Goldewijk et al., 2010); the 0.5°×0.5° resolution potential vegetation map is generated by simple proportional adjustments to the area of potential vegetation presents within each grid cell. The map was used as a starting point to produce all three estimates of land-cover change and LUC conversions. We also assume all forest in the 1765 land cover map as primary forest. At this stage, we do not distinguish between C_3/C_4 types for grassland, pastureland, and cropland. Classification to C_3/C_4 pathways is accomplished in the final step.

Step 3: estimating historical LUCC and LUC conversions

To derive the LUCC and LUC conversion estimates, we define a set of rules to characterize each of the four land-use activities. These rules impose a logical sequence and priority order in which land cover is modified. Based on these rules, a priority factor is assigned to each land cover type within each grid cell, corresponding to each of the four land-use activities (Figure 2.2). The priority factor for a land cover type indicates the probability of that vegetation being altered due to that particular land-use activity. The priority factor for an individual land cover type within each grid cell varies from 0 to 1.0, and the sum of priority factors for all land cover types corresponding to each land-use activity sums up to 1.0.

The rules that determine the priority factors for a land-use activity depend on the magnitude of that land-use activity for that year, the land cover map from the previous year, and the potential vegetation map. For example, for an increase in cropland area between two consecutive years in a grid cell, a priority factor is assigned to each land cover type (except for water, pastureland and urban land for which priority factor is assumed to be 0), which is proportional to the total area of natural vegetation in that grid cell. The increase in cropland area is accounted by converting each land cover type to cropland based on its designated priority factor. In the case of cropland abandonment (characterized by decrease in cropland area with

time), the abandoned land reverts back to the potential vegetation level present in that grid cell. In such cases, the potential vegetation map was used to determine the priority factors. Usually, grasses and other herbaceous land cover types are faster colonizers than forests (Arora and Boer, 2006). They invade the abandoned land initially, while woody vegetation grows later. However, the rationale here is that a one-year time gap is sufficient for woody vegetation to reappear. This method provides a simple representation of successions. LUCC treatment due to urbanization is similar to that described for cropland, with the exception that in case of decrease in urban land area with time, the decreased area is reverted to grasses (i.e. priority factor for grasses was assigned as 1.0), irrespective of the potential vegetations present within that grid cell. For wood harvest, preference is given to primary forests over secondary forests. Priority factors were assigned proportional to the area of each of the seven primary forests within that grid cell. In cases where total primary forest was insufficient to account for wood harvest, clearing was done from secondary forests following a similar approach. For an expansion of pastureland, clearing of grassland is preferred (Houghton, 1999). In cases where grassland is insufficient, we followed the method adopted for increase in cropland area. In case of decrease in pastureland area, the abandoned area was reverted back to grassland.

There are a few exceptions to these rules. In cases where cropland is abandoned and pastureland/urban land concurrently increase with time, a part of the abandoned area was considered a source for pastureland/urban land. The fraction of abandoned cropland area used as a source of pastureland/urban land is determined by the likelihood that the other vegetations present in the grid cell are sources for the growth in pastureland/urban land. For example, a grid cell dominated by forest is more likely to have a higher fraction of abandoned cropland area to be used as a source of pastureland than a grid cell dominated by grassland. Similar treatment exists for decrease in pastureland area accompanied by increase in cropland/urban land, in which a part of cleared pastureland area is considered a source for cropland/urban land. It should be noted that in case of succession, forest returns as secondary forest (vegetation numbers 16 to 22 in Table 2.1), whereas we have not differentiated herbaceous land cover types as primary/secondary. Because the data sets for four land-use activities considered in this study come from more than one source, certain cases exist where the desired conversations, as determined by the assigned prioritization factor, could not be carried out for all four land-use activities. In such cases, we assign the following order of preference to modify land cover: urban

land, cropland, wood harvest, and pastureland. This order of preference was chosen considering the uncertainties in magnitude, spatial distribution, and definitions associated with each land-use activity. Hence, the cropland and pastureland areas in ISAM-RF, ISAM-HYDE, and ISAM-HH will be slightly less compared to the original RF, HYDE, and HH data sets in certain grid cells.

The land cover map of 1765 derived from step 2 (step 2 in Sect. 2.3) is used as the initial condition from which we move forward in time, modifying land cover by superimposing the year-to-year land-use activities following the method described above.

Step 4: Calibration using satellite data

Historically, substantial land-cover changes have occurred due to climate feedbacks (Parry et al., 2007) and through natural disturbances like forest fires (Giglio et al., 2010; van der Werf et al., 2010), blowdowns, and insect outbreaks (Foley et al., 2003; Lambin et al., 2003). Due to the unavailability of information on the magnitude and spatial extent to which these effects have altered land cover historically at a multi-centennial time scale, their impacts on land cover have been excluded from the rule-based approach for estimating historical land-cover change. Additionally, the rule-based approach is a simplified representation of general trends associated with historical land-cover change due to human land-use activities, which is subject to variations at the regional and grid cell levels. Due to the factors discussed above, there exist differences between satellite observations and estimates from the rule-based approach (Step 3; Sect. 2.3). For example, the total forest area estimated using rule-based approach (Figure 2.3(a)) differs from satellite estimates (Figure 2.1(a)) for certain grid cells. Our estimated forest area varies from 36.7 to 39.4 $\times 10^6$ km^2 among the three estimates, compared to 30.7 $\times 10^6$ km^2 (after changing to the land mask used in this study) estimated using MODIS land cover data (Friedl et al., 2010) classified under IGBP classification scheme. We implicitly account for these differences by calibrating with satellite data.

We first classify the 28 land cover classes into two broad categories: forest and non-forest. Medium resolution satellite data captures forest extent/type with high accuracy compared to other herbaceous types (Jung et al., 2006; Friedl et al., 2010). The basic aim is to reconcile these in a way that will make the magnitude and spatial patterns of present-day forest estimates as close as possible to satellite estimates.

We compare estimated forest area for the year 2005 with estimates from 500 m resolution MODIS land cover data (Friedl et al., 2010) for the year 2005 classified under IGBP land classification scheme, aggregated at 0.5°x0.5° resolution. An overestimation of forest area in a grid cell indicates that higher priority factor should be assigned to forest land cover types for clearing than previously assumed. Similarly, an underestimation of forest area in a grid cell indicates a lesser priority factor should be assigned to forest land cover types for clearing. To modify the priority factor for each land cover type in a grid cell for a particular year, we determine a "correction factor" using a combination of information from the potential vegetation map, the land cover map for the year 2005, the land cover map of the historical year for which the priority factor is to be adjusted, and the magnitude of underestimation/overestimation of forest area estimated in comparison to satellite data. The correction factor for each land cover type is chosen such that the estimated area of forest matches with satellite data when the correction factor is multiplied by the priority factor estimated in step 3. The value of the correction factor is > 1 for land cover types with increased priority and < 1 for land cover types with decreased priority. An additional constraint is imposed so that the sum of the correction factor multiplied by priority factor for all land cover types, corresponding to each land-use activity, add up to 1.0, a basic criteria described in step 3 (step 3 in Sect. 2.3). For the grid cells where land-use data indicates the absence of anthropogenic land cover types, a simple linear interpolation is used to adjust the area of natural vegetations between the starting and ending reference years, in order to make the present-day estimates consistent with satellite data. A similar approach was applied to grid cells where the magnitude of historical land-use was small and correction factor alone is insufficient for effecting the changes needed to match satellite estimates. The changes effected through linear interpolation are reflected in annual land cover maps, but are not recorded as LUC conversions. Hence, our estimates of LUC conversions are only attributable to the four direct human land-use activities. To avoid underestimation of forest area from satellite data, which may result due to the exclusion of regrowing forest, we also use four additional years of MODIS land cover data (Friedl et al., 2010) covering the period 2001-2004 to estimate the 'correction factor'. This method results in a close match between MODIS forest distribution (Figure 2.1(a)) and our estimated forest distribution (Figure 2.3(b)).

Step 5: Separation of grassland, pastureland, and cropland to C_3/C_4 types

We only classify grassland, pastureland, and cropland to C_3/C_4 types in annual land cover maps, not annual LUC conversion maps. To separate the grassland and pastureland area fractions into C_3 and C_4 types, we followed the modified approach of Still et al. (2003). If there is at least one month in a year when temperature is above crossover temperature (the temperature at which the C_3 quantum yield equals C_4 quantum yield) and rainfall is concurrently above 25 mm, it is assumed that the C_4 grass fraction is equal to the number of months where C_4 photosynthesis is favored relative to the number of growing season months with a temperature greater than 5°C. Mathematically,

C_4 fraction = (number of months with T_{air} > crossover temperature and rain > 25mm) / (number of months with T_{air} > 5°C)

We use the monthly air temperature (T_{air}) and precipitation data at 0.5°×0.5° resolution based on CRU TS 3.0 (updated based on Mitchell and Jones, 2005), covering the period 1901–2006; a 10-year moving average was calculated for both variables, to avoid sudden fluctuations. For the years 1765 to 1900, average monthly precipitation and temperature values from 1901 to 1910 were used. For the period 2007–2010, the same values were assigned as for the year 2006. For each year, we calculated the crossover temperature following Collatz, Berry and Clark (1998), based on global CO_2 concentration values from 1765 to 2010 (Meinshausen et al. 2011). The calculated crossover temperature varies from 18.2°C in 1765 to 24.1°C in 2010. The C_4 fraction generated for the period 1765–2010 was combined with annual pastureland and grassland estimates from step 4 (step 4 in Sect. 2.3), to separate them into C_3 and C_4 fractions.

To separate the annual croplands into C_3 and C_4 fractions, we use the estimates of harvested areas of 175 different crops across the world at 5 min by 5 min spatial resolution for the year 2000 (Monfreda, Ramankutty and Foley 2008). C_3 and C_4 designations were assigned to each crop type based on known pathway characterizations. A map indicating the fractional coverage of C_4 croplands was generated at 0.5°×0.5° resolution. In grid cells where there were no crops present, 100% of the grid cell was assigned to C_3 croplands. This map was used to separate annual historical cropland estimates into C_3 and C_4 types.

2.4 Results

Comparison of cropland and pastureland estimates among 3 data sets

Comparison of global cropland statistics of RF and HYDE data averaged over the period 2001-2005 shows similar levels of cropland area, varying from 14.3 to 15.3 × 10^6 km^2, with HYDE estimates being 8% higher than RF estimates (Table 2.2). However, this global picture varies regionally. The most pronounced differences are found in Pacific Developed region and China, where the cropland areas estimated by HYDE data are 70% and 23% higher than RF data, respectively. The major differences between the two data sets result from the fact that these data sets adopted different methods (Refer supplementary text) and agricultural inventory data sets. While HYDE inventory data was based on FAO (2008), RF estimates relied more on national-level census statistics, along with FAO estimates for recent years (Ramankutty et al., 2008). Houghton's estimates of both global and regional croplands are lower than RF and HYDE estimates. This is because he considers only croplands that were created or abandoned on lands originally covered by forests.

While global cropland statistics estimated based on RF and HYDE data match reasonably well with each other, pastureland statistics globally show substantial disagreement, with even more regional disagreement. This is because the global pastureland area estimated by the census report used in RF itself is significantly lower than FAO (2008) estimates of pastureland used in HYDE data. Globally, HYDE data estimates of pastureland are 26% higher than the RF estimated value of 26.3 × 10^6 km^2 average over the period 2001–2005. Major disagreement is found over 'North Africa and the Middle East' where pasture area estimates for HYDE are 83% higher than RF for 2005. While the percentage difference is highest for 'North Africa and the Middle East', a large difference in pastureland area is found in the Pacific Developed region and China, where the estimated pastureland area averages from 2001 to 2005 for HYDE are 1.5 × 10^6 km^2 (~53%) and 1.7 × 10^6 km^2 (~43%) higher than RF data, respectively. HH data estimates of pastureland are zero for all regions except Latin America, because Houghton (2008) assumes that all pastures are derived from grasslands, with the exception of Latin America, where significant clearance of forest area for pastureland has taken place due to extensive cattle ranching (Lambin and Giest, 2003).

Land-cover change estimates during 1765–2005

The 28 land cover classes have been combined into a broader category for the purpose of analysis (Table 2.3), and the values are presented in the text as range among three estimates (ISAM-HYDE, ISAM-RF and ISAM-HH). Globally, the total area of forest has decreased from 45.5×10^6 km^2 (~36% of the total land area) to about 29–30 million km^2 during this period, a one-third decrease. Of this, human land-use activities have contributed to a net decrease in forest area of about 6.5–8.4 million km^2 (Table 2.4), while the rest is attributed to indirect anthropogenic and natural causes. Total deforestation amounts to 14.5–14.7 million km^2, and forest regrowth ranges between 6 and 8 million km^2. Forest area in North America shrunk by 3–3.5 million km^2 (~35%–40%) and Tropical Africa shrunk by 2.3–2.6 million km^2 (~43%–49%) (Table 2.4). Total forest area in Europe decreased by 44%–52% from its initial value of 2.5×10^6 km^2. Estimates of forest area in China and South and South-East Asia (SSEA) regions show the largest difference among the three estimates. Forest area in China and SSEA decreased by 40%–52% and 47%–66%, respectively. Such large differences in these regions are mainly due to uncertainty in estimates of cropland (see Ramankutty et al. 2008; Liu and Tian 2010). North America, the former USSR and Tropical Africa show a large amount of net forest loss attributed to indirect anthropogenic and natural causes. Total forest regrowth due to human land-use activities is about 6–8 million km^2. During 2005, roughly 24%–28% of the total forests present are secondary forests (Figure 2.4 and Table 2.3). North America contains about 26% of global secondary forest whereas the former USSR contains 17-23% of global secondary forest (Figure 2.5). ISAM-RF estimates show higher secondary forest in all regions due to more abandonment of croplands present in RF data compared to HYDE data.

Global area of savannas shrunk by 5.4–7.1 million km^2 (i.e. 38%–50%) and shrublands decreased by 6.8–8.9 million km^2 (i.e. 40%–53%) (Table 2.3). The area of grassland and pastureland combined increased by about 19.7–24 million km^2 (i.e. 83%–101%). However, regional comparisons show more disagreement than global estimates of change (Figure 2.5). For a single time snap during 2005, ISAM-HYDE estimates show 57% less shrubland area in the Pacific Developed region compared to ISAM-RF estimates. Except for North America, ISAM-RF shows more cropland expansion in regions that were originally shrublands, compared to

ISAM-HYDE. The area of grassland in ISAM-RF is higher than ISAM-HYDE for all regions because of lower pastureland estimates by RF data compared to HYDE data. As we have considered only deforestation and reforestation statistics due to agricultural activities from HH data, they have been excluded in the discussion relating to comparison of herbaceous land cover types.

LUC conversions during 1765–2005

Globally 6.6–6.8 million km^2 of forest loss (~45% of human-caused forest loss) has occurred due to cropland expansion, whereas only 2.7–2.9 million km^2 was due to pastureland expansion (Supplementary Table S1). SSEA contributes to 25%–30% (1.6–2.1 million km^2) of forest loss occurring due to conversion to cropland, followed by North America (1.1–1.5 million km^2; 16%–23%). Although the cropland estimates for Latin America by RF, HYDE, and HH are in close range of 1.4–1.6 × 10^6 km^2 for the early 2000s (Table 2.2), their pathways of expansion are very different. ISAM-HYDE estimates only 0.65 × 10^6 km^2 of forest loss in Latin America due to cropland expansion, whereas ISAM-RF shows almost double the forest loss estimated by ISAM-HYDE (Supplementary Table S2). Because HH data was spatialized using ISAM-RF estimates, the trend exhibited by ISAM-HH cannot be considered independent from ISAM-RF estimates. Roughly 47%–58% (~7.4–9.6 million km^2) of cropland expansion has occurred due to conversion of non-forested land (Supplementary Table S1).

About 49%–62% of forest loss due to human land use in Latin America occurred due to conversion to pastureland, compared to 29%–36% caused due to cropland expansion (from Supplementary Table S2 and Table 2.4). Globally, 28.5–31.8 million km^2 of non-forested land was used for pastureland expansion, the majority of which consisted of grasslands. It is interesting to note that though the areas of cropland and pastureland estimated by RF are about 1 × 10^6 km^2 and 6.7 × 10^6 km^2 lower than HYDE estimates, respectively, for the period 2001–2005 (Table 2.2), ISAM-RF estimates show substantially more clearing (and regrowth) of forested and non-forested land compared to ISAM-HYDE (Supplementary Table S2 and S4). This is because HYDE data show a consistently increasing trend in cropland and pastureland area over time, compared to RF data which show substantially more abandonment (and thus more regrowth of natural vegetation), leading to more gross conversions by ISAM-RF. ISAM-

RF and ISAM-HH estimates show ~42% contribution of total secondary forest regrowth due to cropland abandonment, whereas ISAM-HYDE show a contribution of only about 23%.

Implications of partitioning to C_3/C_4 type

A significant amount of land in North America and Europe estimated as C_4 grasslands in 1765 was classified as C_3 grasslands for present-day conditions when changes in CO_2 concentration were taken into account in the simplified method of Still et al., (2003) (Refer to supplementary Figure S1). About 10% of the grassland and 22.4% of pastureland from ISAM-HYDE was classified as C_4-type for 2010 (Table 2.3). Combining the same grassland and pastureland estimates for 2010 from ISAM-HYDE with the C_4 fraction map for the year 1765 resulted in 18.4% and 32% classified as C_4-type, respectively. Both ISAM-HYDE and ISAM-RF estimates show about 23% of the total cropland area as C_4-type throughout the historical period.

2.5 Comparison with other studies

We compared our estimates of forest for the year 1990 with other studies (Table 2.5). The year 1990 was chosen for comparison because it is the farthest year from present for which many gridded estimates were available that would facilitate regional comparisons. All the previous modeling (Klein Goldewijk, 2001; Hurtt et al., 2006; Yang et al., 2010; Hurtt et al., 2011) studies show good agreement with one another, even regionally. However, global total forest area estimates from ISAM-HYDE, ISAM-RF, and ISAM-HH are about 10×10^6 km^2 less than previous studies. Major disagreements occur in North America and the former USSR, where our estimates of forest area are reduced by 3.3 and 4.7 million km^2, respectively. Because our forest estimates are a reflection of estimates from satellite-based land cover data, the differences in estimates arising in these regions can be attributed to unaccounted land-cover change, assuming the rule-based approach accurately captures land-cover change occurring due to all major land-use activities.

We compared our estimates with FAO forest statistics for 1990 (FAO, 2010). Due to difference between the definition of forest used in FAO (see FAO, 2001; 2006; 2010) and this study, we performed a 'test case' wherein we repeated the entire calculations using a potential vegetation map derived from MODIS land cover data from the year 2005 (Friedl et al., 2010) classified under the University of Maryland (UMD) classification scheme (Hansen et al., 2000).

Land cover classification in the UMD scheme is favorable for making direct comparisons with FAO estimates. In this case, the estimates seem to agree reasonably well with FAO statistics, with ISAM-HYDE estimates being at the high end for Tropical Africa and Pacific Developed regions. Similar trends were observed when MODIS-estimated forest area (UMD classification scheme) for the year 2005 was directly compared with FAO forest estimates for the same year (FAO, 2010).

2.6 Discussion and conclusions

This study focused on characterizing historical land-cover change and LUC conversions using annual maps of cropland, pastureland, wood harvest, and urban land as inputs. Due to uncertainties associated with estimates of historical land-use activities, three different data sets on agricultural extent were used to derive three different estimates, consistently using the same rule-based method of prioritizing and converting vegetation. Information from remote sensing data was used to constrain and modify the rule-based method to implicitly account for land-cover changes due to indirect anthropogenic or natural causes. The differences among the three estimates produced in this study can be largely explained by the spatial and temporal differences in estimates of cropland and pastureland areas among the three data sets. Therefore these data sets offer a wide range of plausible regional estimates of uncertainty and the extent to which different ecosystems have undergone changes historically.

The data sets produced in this study have several associated limitations. Since the annual cropland and pastureland maps reveal only the net changes in area, we could not calculate the effect of shifting cultivation in this study. Hurtt et al. (2006) performed a sensitivity test by assuming a standard land abandonment rate of 6.7% yr^{-1} due to shifting cultivation in the tropics, and showed that excluding shifting cultivation could lead to underestimation of secondary land created by agriculture. However, we chose not to include shifting cultivation in our study due to high uncertainty in the magnitude and spatial patterns (Hurtt et al., 2006) historically associated with shifting cultivation. In addition, our assumption that all forest on the land cover map for 1765 (starting year of analysis) was primary forest potentially underestimates the secondary forest area created due to wood harvest and cropland abandonment before 1765. The validity of this assumption is well established due to the fact that the aim of this study is to characterize land-cover change after the pre-industrial era.

There are three major sources of uncertainty. First, the potential vegetation map produced from satellite data is assumed to accurately represent the land cover that would have existed at present if human activities have been non-existent. Hence, the usage of potential vegetation map to represent pre-industrial land cover assumes that changes in environmental conditions have not changed the land cover. The second source of uncertainty arises from the rule-based approach to prioritize land-cover change used in this method; this is a simple representation of historical trends associated with various land-use activities that are not fully understood, and difficult to generalize at a global or regional scale. As shown, the rule-based approach leads to a land cover map that differs substantially compared to satellite estimates for recent years. However, we attribute the differences to unaccounted land-cover change and grid-cell level variations in land-use trends assumed in our rule-based approach. This difference is subsequently used to revise the rules at the grid-cell level to produce estimates close to satellite observations. Hence, the estimates provided here are largely dependent on the simplified representation of converting land cover assumed in this study. However, we have not performed a systematic sensitivity analysis of the different assumptions made to modify land cover. The third source of uncertainty arises due to land-use data sets used as inputs. Estimates of historical gridded wood harvest data were based on several assumptions, which are subject to uncertainty (Hurtt et al., 2006). As shown in this study and in other previous studies (Klein Goldewijk and Ramankutty, 2004; Jain and Yang, 2005), spatial and temporal patterns of historical cropland and pastureland have significant uncertainties. This is reflected in the distribution of non-forested land cover types as estimated using three agricultural data sets (Figure 2.5). As a result, only the total non-forested land as a single broad category matches with satellite estimates. The individual forest area, however, does seem to agree reasonably well between the three estimates, primarily due to the calibration carried out in step 4 (Sect. 2.3). Constraining each land cover type (especially herbaceous types) to be close to satellite estimates is impossible, as the cropland and pastureland estimates prescribed based on input data sets need to remain unaltered. In addition, medium/coarse-resolution satellite data have less accuracy in classifying herbaceous land cover types than trees or barren land (Friedl et al., 2010).

Several regional and national-level reconstructions using finer resolution census data have revealed significant differences in estimates of cropland and pastureland compared to older versions of RF and HYDE global data sets. For example, Li et al. (2010) found that RF data

overestimated cropland area in China by a factor of 21 for the year 1700 and 1.6 for 1990 when compared with the cropland data of Northeast China (Ye and Fang., 2011) reconstructed based on combining calibrated historical data from multiple sources. Similarly, they found significant differences in the spatial distribution of cropland in HYDE data for the 18th and 19th century. Historical reconstructions over Amazonia (Leite et al., 2011) using municipal-level census data with higher level of details also show considerable difference in spatial patterns and magnitude compared to RF data. The range of uncertainties in regional estimates is expected to have narrowed in most recent RF and HYDE data used in this study, but significant differences still exist. Because the three estimates produced in this study are directly dependent on these global land-use data sets, our global data sets should also be used with caution while drawing inferences from regional-level analysis. Since no single agricultural land-use data set used here can be pointed out as better as or worse than another, it is recommended to use all three estimates alongside one another to gain a better picture of the range of uncertainties.

The biggest source of uncertainty in the global C budget remains emissions due to LUCC, (Canadell, 2002) and these are estimated to be in the order of ± 0.5 GtC y^{-1} (Houghton, 2005; Houghton et al., 2012). Several multi-model comparison experiments have been performed to determine the uncertainty of LUCC in the global carbon budget (e.g. McGuire et al., 2001; Pitman et al., 2009; Reick et al., 2010). The LUCC uncertainty experiments involve using a common land-use data set (e.g. HYDE or RF) in each of the models and comparing the land-use fluxes. However, due to differences in the structure of each model, the method adopted to implement the common land-use data differs significantly between each model (e.g. see Pitman et al., 2009). As a result, it is impossible to attribute the estimated uncertainty to model-related uncertainty and uncertainties arising due to differences in the method of implementing land-use data between different models. However, driving the same model with multiple LUCC data sets derived consistently using same method, as discussed here, opens a new avenue for studying LUCC data-related uncertainty by eliminating the model-related uncertainty.

Certainly, indirect anthropogenic and natural effects have been dominant factors in historical land-cover change and have been poorly documented at a global scale (Lambin et al., 2003). Additionally, land-cover modifications like agricultural intensification have been thought to have a widespread impact on climate through altered surface attributes and changes in

biogeochemical cycles. Recent advances in remote sensing observations have provided a more accurate and globally consistent picture of more subtle changes in land cover (e.g. changes in tree height, vegetation biomass, and vegetation structure), in addition to capturing land-cover changes. Because globally consistent remote sensing observations are available only for the past four decades, we need to rely on other methods of reconstructing large time-scale changes in land cover. Monitoring all forms of land-cover change extensively and consistently at a global scale for the pre-satellite era, even at medium/coarse spatial and temporal resolution, was impractical. Hence, several assumptions need to be made to account for its impact on LUCC. Future research is required on monitoring long-term changes in all forms of land-cover change and land-cover modifications at higher spatial and temporal resolutions through remote sensing observations. Further, tracking LUC conversions rather than net changes in land cover can help facilitate better understanding of trends and fate of LUCC and its implications.

As pointed out by Pitman et al. (2009), implementing a common LUCC data set among different models is impossible. As a result, implementing the land cover maps and LUC conversion estimates presented here in different models may be subject to different approximations depending on the complexity and parameters associated with each model. However, we have chosen land cover classifications such that the data can be implemented in models without introducing much uncertainty. Preliminary results of regional and global carbon emissions for the last three decades, estimated by implementing these three sets of data in the ISAM, have already been used in the IPCC AR5. A detailed assessment of the range of biogeophysical and biogeochemical impacts produced by these three estimates is in progress using a coupled ISAM-CESM framework. We believe that the data sets presented here will be useful to modelers interested in studying the effects of historical LUCC on biogeophysics, biogeochemistry and hydrological cycle, as well as in general to the global change community interested in studying the impacts of historical LUCC. Digital versions of these data sets can be downloaded from the webpage (http://www.atmos.illinois.edu/~meiyapp2/datasets.htm).

Acknowledgements

We thank Navin Ramankutty for sharing the new historical cropland and pastureland data sets and answering several questions relating to the data sets. We thank Kees Klein Goldewijk for providing access to an older version of HYDE data. We would like to thank Richard

Houghton for sharing his latest regional cropland and pastureland data. We thank Chad Monfreda for providing the C_3/C_4 harvested cropland fraction map. We also owe thanks to Martin Jung and Markus Reichstein for providing FLUXNET-MTE data sets. We would also like to acknowledge ORNL DAAC for their extensive collection of satellite-based land cover data, which were of immense use to us. This work was supported by National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change Program (No. NNX08AK75G).

2.7 Tables

Table 2.1 Land cover classifications used in this study

No.	Land cover type	Symbol
1*	Tropical evergreen broadleaf forest	TrpEBF
2*	Tropical deciduous broadleaf forest	TrpDBF
3*	Temperate evergreen broadleaf forest	TmpEB
4*	Temperate evergreen needleleaf forest	TmpEN
5*	Temperate deciduous broadleaf forest	TmpDB
6*	Boreal evergreen needleleaf forest	BorENF
7*	Boreal deciduous needleleaf forest	BorDNF
8*	Savanna	Savanna
9*	C_3 grassland/steppe	C_3grass
10*	C_4 grassland/steppe	C_4grass
11*	Dense shrubland	Densesh
12*	Open shrubland	Openshr
13*	Tundra	Tundra
14*	Desert	Desert
15*	Polar desert/rock/ice	PdRI
16	Secondary tropical evergreen broadleaf	SecTrpE
17	Secondary tropical deciduous broadleaf	SecTrp
18	Secondary temperate evergreen broadleaf	SecTmp
19	Secondary temperate evergreen needleleaf	SecTmp
20	Secondary temperate deciduous broadleaf	SecTmp
21	Secondary boreal evergreen needleleaf	SecBorE
22	Secondary boreal deciduous needleleaf	SecBor
23*	Water/Rivers	Water
24	C_3 cropland	C_3crop
25	C_4 cropland	C_4crop
26	C_3 pastureland	C_3past
27	C_4 pastureland	C_4past
28	Urban land	Urban

Note: * Natural land cover classes used in this study. Except for water/rivers (No. 23), all other natural land cover classes were directly derived from the potential vegetation map of Ramankutty and Foley (1999). Note that C_3 and C_4 grasslands (Nos. 9 and 10) are considered to be a single land cover class in the potential vegetation map and during the initial stages of calculation. Partitioning to C_3 and C_4 types is carried out in the last step (step 5; Sect. 2.3).

Table 2.2 Regional areas of cropland and pastureland averaged for the period 2001–2005 estimated directly from RF (Updated estimates based on Ramankutty and Foley, 1999), HYDE (Klein Goldewijk et al., 2011) and HH (Houghton, 2008) data sets across 9 regions covering the world. The 9 regions are based on Houghton et al. (1983). Units are in million km^2. All values are rounded to one decimal place.

Regions	Cropland				Pastureland			
	RF	HYDE	HH	Range	RF	HYDE	HH	Range
North America	2.1	2.3	1.9	1.9 – 2.3	2.4	2.5	0.0	0.0 – 2.5
Latin America	1.6	1.5	1.4	1.4 – 1.6	4.8	5.4	2.8	2.8 – 5.4
Europe	1.2	1.2	0.1	0.1 – 1.2	0.6	0.7	0.0	0.0 – 0.7
North Africa and Middle East	0.8	0.9	0.3	0.3 – 0.9	1.8	3.0	0.0	0.0 – 3.0
Tropical Africa	2.0	2.0	1.9	1.9 – 2.0	7.0	8.0	0.0	0.0 – 8.0
Former USSR	2.0	2.2	0.4	0.4 – 2.2	3.3	3.6	0.0	0.0 – 3.6
China	1.3	1.6	0.7	0.7 – 1.6	3.5	5.2	0.0	0.0 – 5.2
South & South-East Asia	3.0	2.9	1.5	1.5 – 3.0	0.3	0.4	0.0	0.0 – 0.4
Pacific Developed Region	0.4	0.6	0.2	0.2 – 0.6	2.6	4.1	0.0	0.0 – 4.1
World	14.3	15.3	7.6	7.6 – 15.3	26.3	33.0	2.8	2.8 – 33.0

Table 2.3 Global area of various land cover types for 4 time slices based on ISAM-RF, ISAM-HYDE, and ISAM-HH estimates. 'Primary forest' includes TrpEBF, TrpDBF, TmpEBF, TmpENF, TmpDBF, BorENF, and BorDNF. 'Secondary forest' includes SecTrpEBF, SecTrpDBF, SecTmpEBF, SecTmpENF, SecTmpDBF, SecBorENF, and SecBorDNF. Shrublands are a combination of Denseshrub and Openshrub. 'Others' category includes Tundra, Desert, and PdRI. The estimates of cropland and pastureland area are slightly lower than the original estimates (Table 2.1) due to a difference in land mask used and other minor adjustments made in step 3 (Sect. 2.3) for consistency purposes (Unit: million km^2)

Land cover	1765	1900			2000			2005		
Type	RF/HYDE /HH	RF	HYDE	HH	RF	HYDE	HH	RF	HYDE	HH
Primary Forest	45.4	34.9	34.8	33.5	22.1	22.5	20.8	21.7	22.2	20.
Secondary Forest	0.0	2.9	2.9	3.1	7.9	7.0	7.5	8.3	7.2	7.8
C_3 Cropland	2.9	5.9	6.2	4.2	10.0	11.4	5.5	10.0	11.6	5.6
C_4 Cropland	0.6	1.7	1.8	1.2	2.9	3.4	1.5	2.9	3.4	1.5
C_3 Pastureland	3.0	9.1	9.1	3.3	18.0	24.4	4.2	18.0	24.6	4.3
C_4 Pastureland	1.2	3.0	3.6	1.7	5.9	7.7	2.6	5.5	7.3	2.6
C_3 Grasslands	14.6	15.6	15.4	20.2	16.5	13.8	26.0	17.2	14.1	26.
C_4 Grasslands	4.9	4.1	3.7	5.8	2.7	1.8	4.5	2.7	1.7	4.2
Savannas	14.2	13.0	12.5	14.2	9.1	7.2	14.2	8.8	7.1	14.
Shrublands	16.9	14.1	14.6	16.8	10.1	8.0	16.8	10.1	8.0	16.
Others	26.1	25.7	25.4	26.1	24.4	22.5	26.1	24.4	22.5	26.
Urban Land	0.0	< 0.1	< 0.1	< 0.1	0.4	0.4	0.4	0.5	0.5	0.5

Table 2.4 Area of forest cleared and forest regrown during the period 1765–2005 across 9 regions covering the world, based on ISAM-RF, ISAM-HYDE, and ISAM-HH estimates. Total deforested and forest regrowth estimates are based on four land-use activities only. However, changes in forest area effected due to calibration with satellite data (step 4; Sect. 2.3) are reflected in year 2005 forest estimates (Unit: million km^2)

Regions	Forest area in 1765	Total deforested area			Total forest regrowth			Estimated forest area in 2005
		ISAM-RF	ISAM-HYDE	ISAM-HH	ISAM-RF	ISAM-HYDE	ISAM-HH	
North America	9.6	3.3	3.5	3.3	2.4	2.0	2.2	5.8–6.2
Latin America	10.5	3.1	2.4	4.5	1.0	0.6	1.2	8.4–8.8
Europe	2.5	2.0	1.6	1.3	1.5	1.0	1. 1	1.2–1.4
North Africa and Middle East	0.2	0.1	0.1	0.1	< 0.1	< 0.1	< 0.1	~0.1
Tropical Africa	5.3	1.2	1.2	0.9	0.4	0.3	0.5	2.7–3.0
Former USSR	8.1	1.4	1.8	0.9	0.8	1.1	0.7	5.9–6.0
China	2.3	1.1	1.1	0.7	0.8	0.3	0.7	1.1–1.4
South & South-East Asia	5.8	2.0	2.1	2.4	0.7	0.4	1.2	2.0–3.1
Pacific Developed Region	1.2	0.4	0.4	0.4	0.3	0.2	0.3	~1.1
World	45.5	14.7	14.4	14.5	8.0	6.0	8.0	28.3–30.0

Table 2.5 Comparison of regional forest area estimated in this study with other published studies for the year 1990. The results from this study are provided as a range of forest area estimated from ISAM-RF, ISAM-HYDE, and ISAM-HH. An additional 'test case' was performed (following UMD land classification scheme) to facilitate direct comparisons with FAO estimates (Unit: million km^2)

Regions	Yang et al. (2010)	Klein Goldewijk (2001)	Hurtt et al. (2006)	IPCC AR5[a)]	This study	Test case	
						FAO[b)]	This study (UMD scheme)
North America	9.5	8.7	9.3	9.3	5.8-	5.1	4.1-4.5
Latin America	9.0	9.2	9.0	8.6	7.4-	10.2	9.8-10.1
Europe	2.1	2.2	1.6	1.5	1.3-	1.7	1.5
North Africa	0.1	< 0.1	< 0.1	< 0.1	<	0.1	0.4
Tropical Africa	4.3	3.3	4.4	4.0	2.8-	6.9	7.0-9.8
Former USSR	11.0	11.9	9.7	10.0	5.9-	8.1	6.3- 6.5
China	1.0	1.3	2.5	2.0	1.2-	1.7	1.8- 2.0
South & South-East Asia	4.1	3.3	3.3	3.4	3.1-3.2	3.6	3.3- 3.4
Pacific	1.2	1.4	1.1	1.1	1.1	2.2	2.4- 3.7
World	42.3	41.5	40.9	39.9	29.0	39.6	37.2-41.3

Note: a) Based on Hurtt et al. (2011), b) from Global Forest Resource Assessment (FRA) 2010

2.8 Figures

Figure 2.1 Global distribution of forest area during 2005 based on (a) 500 m resolution MODIS Land Cover Data Set (Friedl et al., 2010) following IGBP land classification scheme aggregated to 0.5°×0.5° resolution and (b) estimates by Hurtt et al. (2011). (Unit: % per grid cell area)

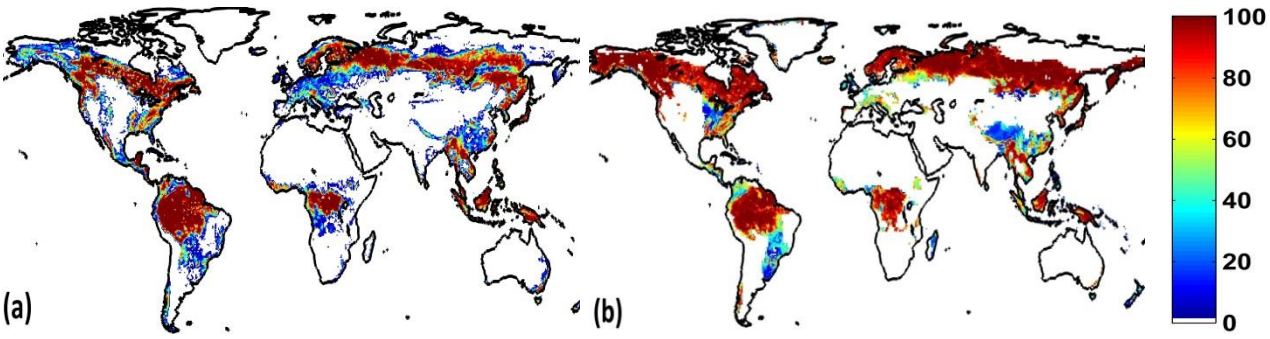


Figure 2.2 Schematic showing the process involved in step 3 to estimate LUCC and LUC conversions. Step 4 involves modification of priority factors estimated from step 3 using forest area estimated from MODIS land cover data (Friedl et al., 2010). '*i*' denotes year, which increases from 1765 to 2005/2007/2010 (ISAM-HH/ISAM-RF/ISAM-HYDE) in annual time steps. The priority factors shown here are just an example, and they vary for each land cover type from year to year between each grid cell.

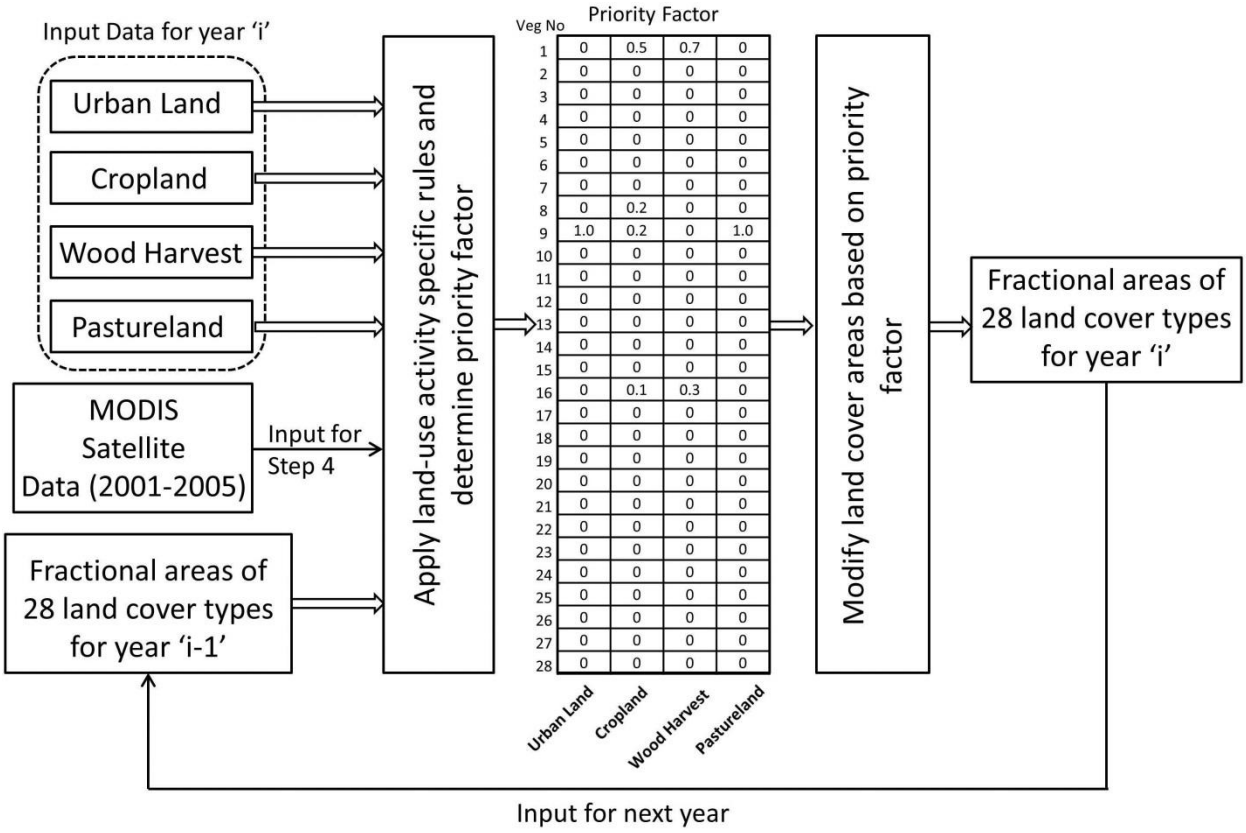


Figure 2.3 Estimated global forest area for the year 2005 based on ISAM-RF, (a) Without calibration (b) after calibration using MODIS land cover data (Friedl et al., 2010). (Unit: % per grid cell area)

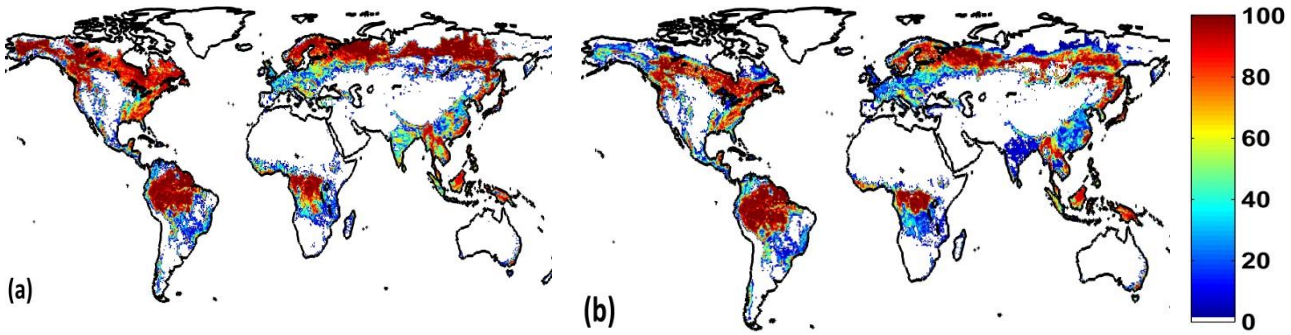


Figure 2.4 Estimated (a) primary and (b) secondary forest area for the year 2005 based on ISAM-RF. (Unit: % per grid cell area)

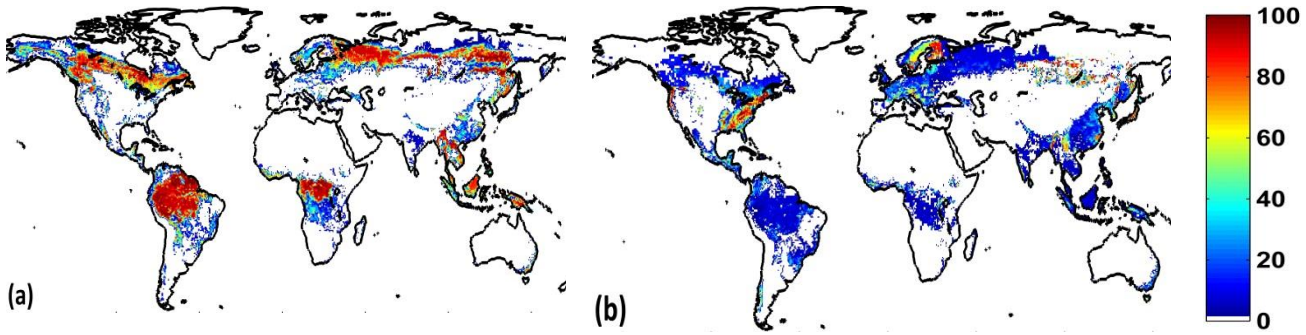
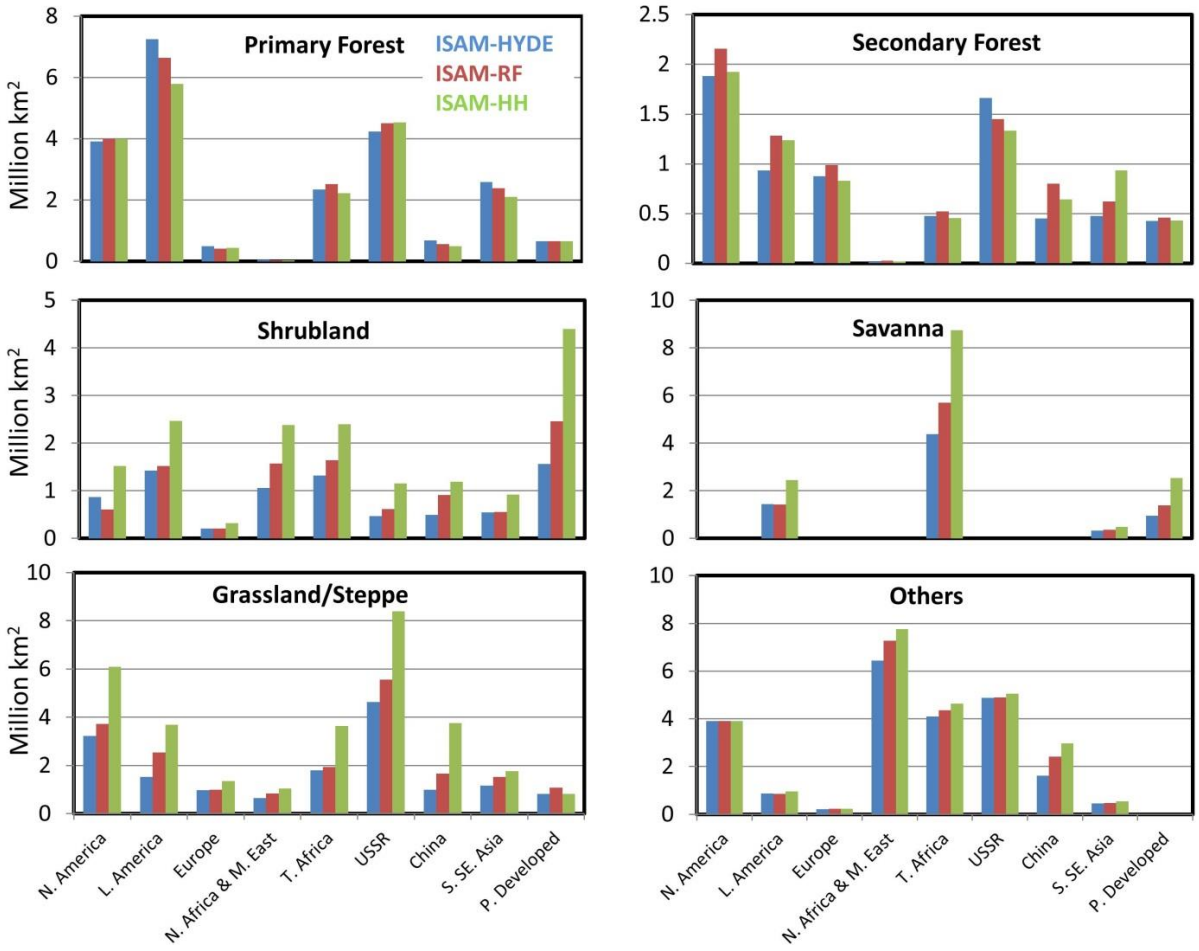


Figure 2.5 Regional comparisons of various natural land cover types during 2005 based on ISAM-RF, ISAM-HYDE, and ISAM-RF. 'Primary forest' includes TrpEBF, TrpDBF, TmpEBF, TmpENF, TmpDBF, BorENF, and BorDNF. 'Secondary forest' includes SecTrpEBF, SecTrpDBF, SecTmpEBF, SecTmpENF, SecTmpDBF, SecBorENF, and SecBorDNF. Shrublands are a combination of Denseshrub and Openshrub. Grasslands are a combination of C_3grass and C_4grass. 'Others' category includes Tundra, Desert and PdRI. (Unit: million km^2)



2.9 References

Arora V K, Boer G J (2006), Simulating competition and coexistence between plant functional types in a dynamic vegetation model. Earth Interactions, 10(10): 30

Bala G, Caldeira K, Wickett M, Phillips T J, Lobell D B, Delire C, Mirin A (2007). Combined climate and carbon-cycle effects of large-scale deforestation. Proc Natl Acad Sci USA, 104(16): 6550–6555 doi:10.1073/pnas.0608998104 PMID:17420463

Bonan G B (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science, 320(5882): 1444–1449 doi:10.1126/science.1155121 PMID:18556546

Bonan G B, Pollard D, Thompson S L (1992). Effects of boreal forest vegetation on global climate. Nature, 359(6397): 716–718 doi:10.1038/359716a0

Brovkin V, Claussen M, Driesschaert E, Fichefet T, Kicklighter D, Loutre M F, Matthews H D, Ramankutty N, Schaeffer M, Sokolov A (2006). Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity. Clim Dyn, 26(6): 587–600 doi:10.1007/s00382-005-0092-6

Canadell J G (2002). Land use effects on terrestrial carbon sources and sinks. Science in China(Series C), 45(suppl.): 1–9 PMID:18763058

Canadell J G, Le Quéré C, Raupach M R, Field C B, Buitenhuis E T, Ciais P, Conway T J, Gillett N P, Houghton R A, Marland G (2007). Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. Proc Natl Acad Sci USA, 104(47): 18866–18870 doi:10.1073/pnas.0702737104 PMID:17962418

Collatz G J, Berry J A, Clark J S (1998). Effects of climate and atmospheric CO_2 partial pressure on the global distribution of C_4 grasses: Present, past, and future. Oecologia, 114(4): 441–454 doi:10.1007/s004420050468

Ellis E C, Klein Goldewijk K, Siebert S, Lightman D, Ramankutty N (2010). Anthropogenic transformation of the biomes, 1700 to 2000. Glob Ecol Biogeogr, 19: 589–606 doi:10.1111/j.1466-8238.2010.00540.x

FAO (2001), Global Forest Resources Assessment 2000, Main Report, FAO Forestry Paper 140, Rome, Italy. (Available at http://www.fao.org/forestry/fra/2000/report/en/)

FAO (2006), Global Forest Resources Assessment 2005, Main Report, FAO Forestry paper 147, Rome, Italy. (Available at http://www.fao.org/forestry/fra/fra2005/en/)

FAO (2008), FAOSTAT. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy. (Available at: http://www.fao.org)

FAO (2010), Global Forest Resources Assessment 2010, Main Report, FAO Forestry paper 163, Rome, Italy. (Available at www.fao.org/forestry/fra/fra2010/en/)

Feddema J J, Oleson K W, Bonan G B, Mearns L O, Buja L E, Meehl G A, Washington W M (2005). The importance of land-cover change in simulating future climates. Science, 310(5754): 1674–1678 doi:10.1126/science.1118160 PMID:16339443

Findell K L, Pitman A J, England M H, Pegion P J (2009). Regional and global impacts of land cover change and sea surface temperature anomalies. J Clim, 22(12): 3248–3269 doi:10.1175/2008JCLI2580.1

Foley J A, Costa M H, Delire C, Ramankutty N, Snyder P (2003). Green surprise? How terrestrial ecosystems could affect Earth's climate. Front Ecol Environ, 1(1): 38–44

Foley J A, Defries R, Asner G P, Barford C, Bonan G, Carpenter S R, Chapin F S, Coe M T, Daily G C, Gibbs H K, Helkowski J H, Holloway T, Howard E A, Kucharik C J, Monfreda C, Patz J A, Prentice I C, Ramankutty N, Snyder P K (2005). Global consequences of land use. Science, 309(5734): 570–574 doi:10.1126/science.1111772 PMID:16040698

Foley J A, Ramankutty N, Brauman K A, Cassidy E S, Gerber J S, Johnston M, Mueller N D, O'Connell C, Ray D K, West P C, Balzer C, Bennett E M, Carpenter S R, Hill J, Monfreda C, Polasky S, Rockström J, Sheehan J, Siebert S, Tilman D, Zaks D P (2011). Solutions for a cultivated planet. Nature, 478(7369): 337–342 doi:10.1038/nature10452 PMID:21993620

Friedl M A, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010). MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens Environ, 114(1): 168–182 doi:10.1016/j.rse.2009.08.016

Giglio L, Randerson J T, van der Werf G R, Kasibhatla P S, Collatz G J, Morton D C, DeFries R S (2010). Assessing variability and long-term trends in burned area by merging multiple satellite fire products. Biogeosciences, 7(3): 1171–1186 doi:10.5194/bg-7-1171-2010

Hansen M C, DeFries R S, Townshend J G R, Sohlberg R (2000). Global land cover classification at 1km spatial resolution using a classification tree approach. Int J Remote Sens, 21(6–7): 1331–1364 doi:10.1080/014311600210209

Houghton R A (1999). The Annual Net Flux of Carbon to the Atmosphere from Changes in Land Use 1850–1990. Tellus B Chem Phys Meterol, 51(2): 298–313 doi:10.1034/j.1600-0889.1999.00013.x

Houghton R A (2005). Aboveground forest biomass and the global carbon balance. Glob Change Biol, 11(6): 945–958 doi:10.1111/j.1365-2486.2005.00955.x

Houghton R A (2008). Carbon flux to the atmosphere from land-use changes: 1850–2005. InTRENDS: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, TN, USA

Houghton R A, Hobbie J E, Melillo J M, Moore B, Peterson B J, Shaver G R, Woodwell G M (1983). Changes in the carbon content of terrestrial biota and soils between 1860 and 1980: a net release of CO_2 to the atmosphere. Ecol Monogr, 53(3): 235–262 doi:10.2307/1942531

Houghton R A, van der Werf G R, DeFries R S, Hansen M C, House J I, Le Quéré C, Pongratz J, Ramankutty N (2012). Chapter G2 carbon emissions from land use and land-cover change. Biogeosciences Discuss, 9(1): 835–878 doi:10.5194/bgd-9-835-2012

Hurtt G C, Chini L P, Frolking S, Betts R, Feddema J, Fischer G, Hibbard K, Houghton R, Janetos A, Jones C, Kindermann G, Kinoshita T, Klein Goldewijk K, Riahi K, Shevliakova E, Smith S, Stehfest E, Thomson A, Thorton P, van Vuuren D, Wang Y (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Clim Change, 109(1–2): 117–161 doi:10.1007/s10584-011-0153-2

Hurtt G C, Frolking S, Fearon M G, Moore B III, Shevliakova E, Malyshev S, Pacala S W, Houghton R A (2006). The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood harvest activity, and resulting secondary lands. Glob Change Biol, 12(7): 1208–1229 doi:10.1111/j.1365-2486.2006.01150.x

Jain A K, Yang X (2005). Modeling the effects of two different land cover change data sets on the carbon stocks of plants and soils in concert with co_2 and climate change. Global Biogeochem Cycles, 19(2): GB2015 doi:10.1029/2004GB002349

Jain A K, Yang X, Kheshgi H, McGuire A D, Post W P, Kicklighter D (2009). Nitrogen attenuation of terrestrial carbon cycle response to global environmental factors. Global Biogeochem Cycles, 23(4): GB4028 doi:10.1029/2009GB003519

Jung M, Henkel K, Herold M, Churkina G (2006). Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens Environ, 101(4): 534–553 doi:10.1016/j.rse.2006.01.020

Klein Goldewijk K (2001). Estimating global land use change over the past 300 years: The HYDE database. Global Biogeochem Cycles, 15(2): 417–433 doi:10.1029/1999GB001232

Klein Goldewijk K, Beusen A, de Vos M, van Drecht G (2011). The HYDE 3.1 spatially explicit database of human induced land use change over the past 12000 years. Glob Ecol Biogeogr, 20(1): 73–86 doi:10.1111/j.1466-8238.2010.00587.x

Klein Goldewijk K, Beusen A, Janssen P (2010). Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3.1. Holocene, 20(4): 565–573 doi:10.1177/0959683609356587

Klein Goldewijk K, Ramankutty N (2004). Land cover change over the last three centuries due to human activities: The availability of new global data sets. GeoJournal, 61(4): 335–344 doi:10.1007/s10708-004-5050-z

Klein Goldewijk K, van Drecht G, Bouwman A F (2006). Contemporary global cropland and grassland distributions on a 5 × 5 minute resolution. Journal of Land Use Science, 2(3): 167–190 doi:10.1080/17474230701622940

Lambin E F, Geist H J (2003). Regional differences in tropical deforestation. Environment, 45(6): 22–36

Lambin E F, Geist H J, Lepers E (2003). Dynamics of land use and land cover change in tropical regions. Annu Rev Environ Resour, 28(1): 205–241 doi:10.1146/annurev.energy.28.050302.105459

Lawrence P J, Feddema J J, Bonan G B, Meehl G A, O'Neill B C, Levis S, Lawrence D M, Oleson K W, Kluzek E, Lindsay K, Thornton P E (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. J Clim: 120208141321006 doi:10.1175/JCLI-D-11-00256.1

Leite C C, Costa M H, de Lima C A, Ribeiro C A A S, Sediyama G C (2011). Historical reconstruction of land use in the Brazilian Amazon (1940 – 1995). Journal of Land Use Science, 6(1): 33–52 doi:10.1080/1747423X.2010.501157

Li B B, Fang X Q, Ye Y, Zhang X (2010). Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data – a case study in Northeast China. Science China Earth Sciences, 53(11): 1689–1699 doi:10.1007/s11430-010-4053-5

Liu M, Tian H (2010). China's land cover and land use change from 1700 to 2005: estimations from high-resolution satellite data and historical archives. Global Biogeochem Cycles, 24(3): GB3003 doi:10.1029/2009GB003687

Loveland T R, Belward A S (1997). The IGBP-DIS global 1 km land cover data set, DISCover first results. Int J Remote Sens, 18(15): 3289–3295 doi:10.1080/014311697217099

McGuire A D, Sitch S, Clein J S, Dargaville R, Esser G, Foley J, Heimann M, Joos F, Kaplan J, Kicklighter D W, Meier R A, Melillo J M, Moore B III, Prentice I C, Ramankutty N, Reichenau T, Schloss A, Tian H, Williams L J, Wittenberg U (2001). Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO_2, climate and land use effects with four process-based ecosystem models. Global Biogeochem Cycles, 15(1): 183–206 doi:10.1029/2000GB001298

Meinshausen M, Smith S J, Calvin K V, Daniel J S, Kainuma M L T, Lamarque J F, Matsumoto K, Montzka S A, Raper S C B, Riahi K, Thomson A M, Velders G J M, van Vuuren D (2011). The RCP greenhouse gas concentrations and their extension from 1765 to 2300. Clim Change, 109(1–2): 213–241 doi:10.1007/s10584-011-0156-z

Mitchell T D, Jones P D (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol, 25(6): 693–712 doi:10.1002/joc.1181

Monfreda C, Ramankutty N, Foley J A (2008). Farming the Planet. Part 2: The geographic distribution of crop areas and yields in the year 2000. Global Biogeochem Cycles, 22: GB1022 doi:10.1029/2007GB002947

Moss R H, Edmonds J A, Hibbard K A, Manning M R, Rose S K, van Vuuren D P, Carter T R, Emori S, Kainuma M, Kram T, Meehl G A, Mitchell J F B, Nakicenovic N, Riahi K, Smith S J, Stouffer R J, Thomson A M, Weyant J P, Wilbanks T J (2010). The next generation of scenarios for climate change research and assessment. Nature, 463(7282): 747–756 doi:10.1038/nature08823 PMID:20148028

Olofsson J, Hickler T (2008). Effects of human land-use on the global carbon cycle during the 5 last 6000 years. Vegetation History and Archaeobotany, 17(5): 605–615 doi:10.1007/s00334-007-0126-6

Parry M L, Canziani O F, Palutikof J P, van der Linden P J, Hanson C E (2007). Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 211–272

Pielke R A, Marland G, Betts R A, Chase T N, Eastman J L, Niles J O, Niyogi D D S, Running S W (2002). The influence of landuse change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. Philos Trans R Soc Lond A, 360(1797): 1705–1719 doi:10.1098/rsta.2002.1027

Pielke R A, Pitman A, Niyogi D, Mahmood R, McAlpine C, Hossain F, Klein Goldewijk K, Nair U, Betts R, Fall S, Reichstein M, Kabat P, de Noblet N (2011). Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdisciplinary Reviews: Clim Change, 2(6): 828–850 doi:10.1002/wcc.144

Pitman A J, Avila F B, Abramowitz G, Wang Y P, Phipps S J, Noblet-Ducoudré N (2011). Importance of background climate in determining impact of land-cover change on regional climate. Nature Climate Change, 1(9): 472–475 doi:10.1038/nclimate1294

Pitman A J, de Noblet-Ducoudré N, Cruz F T, Davin E, Bonan G B, Brovkin V, Claussen M, Delire C, Gayler V, van den Hurk B J J M, Lawrence P J, van der Molen M K, Müller C, Reick C H, Seneviratne S I, Strengers B J, Voldoire A (2009). Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study., Geophysical Research Letters, 36, L14814, doi:10.1029/ 2009GL039076

Pongratz J, Reick C, Raddatz T, Claussen M (2008). A reconstruction of global agricultural areas and land cover for the Last Millennium. Global Biogeochem Cycles, 22(3): GB3018 doi:10.1029/2007GB003153

Pongratz J, Reick C H, Raddatz T, Claussen M (2009). Effects of anthropogenic land cover change on the carbon cycle of the last millennium. Global Biogeochem Cycles, 23(4): GB4001 doi:10.1029/2009GB003488

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

Ramankutty N, Foley J (1999). Estimating historical changes in global land cover: croplands from 1700 to 1992. Global Biogeochem Cycles, 13(4): 997–1028 doi:10.1029/1999GB900046

Reick C, Raddatz T, Pongratz J, Claussen M (2010). Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO_2. Tellus B Chem Phys Meterol, 62(5): 329–336 doi:10.1111/j.1600-0889.2010.00479.x

Shevliakova E, Pacala S W, Malyshev S, Hurtt G C, Milly P C D, Caspersen J P, Sentman L T, Fisk J P, Wirth C, Crevoisier C (2009). Carbon cycling under 300 years of land use change: importance of the secondary vegetation sink. Global Biogeochem Cycles, 23(2): GB2022 doi:10.1029/2007GB003176

Still C J, Berry J A, Collatz G J, DeFries R S (2003). Global distribution of C_3 and C_4 vegetation: carbon cycle implications. Global Biogeochem Cycles, 17(1): 1006–1020 doi:10.1029/2001GB001807

van der Werf G R, Randerson J T, Giglio L, Collatz G J, Mu M, Kasibhatla P S, Morton D C, DeFries R S, Jin Y, van Leeuwen T T (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmos Chem Phys, 10(23): 11707–11735 doi:10.5194/acp-10-11707-2010

Yang X, Richardson T K, Jain A K (2010). Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake. Biogeosciences Discuss, 7(2): 2739–2765 doi:10.5194/bgd-7-2739-2010

Yang X, Wittig V, Jain A K, Post W (2009). Integration of nitrogen dynamics into a global terrestrial ecosystem model. Global Biogeochem Cycles, 23: GB4028 doi:10.1029/2009GB003474

Ye Y, Fang X Q (2011). Spatial pattern of land cover changes across Northeast China over the past 300 years. J Hist Geogr, 37(4): 408–417 doi:10.1016/j.jhg.2011.08.018.

CHAPTER 3

CO_2 emissions from land use change affected more by nitrogen cycle, than by the choice of land cover data

3.1 Abstract

The high uncertainty in land-based CO_2 fluxes estimates is thought to be mainly due to uncertainty in quantifying historical changes between forests, croplands and grassland, but also due to different processes included in calculation methods. Inclusion of a nitrogen cycle in models is fairly recent and strongly affects carbon fluxes. In this study, for the first time, we use a model with C and N dynamics with three distinct historical reconstructions of land-use and land-use change (LULUC) to quantify LULUC emissions and uncertainty that includes the integrated effects of not only climate and CO_2, but also nitrogen. The modeled global average emissions including N dynamics for the 1980's, 1990's and 2000 to 2005 were 1.8±0.2, 1.7±0.2, and 1.4±0.2 GtC/yr respectively (mean and range across LULUC data sets). The tropical emissions were 0.8±0.2, 0.8±0.2 and 0.7±0.3GtC/yr, and the non-tropics were 1.1±0.5, 0.9±0.2 and 0.7±0.1 GtC/yr. Between the 1980s and the 2000s, the HYDE data set indicated a decrease in emissions in the tropics (30%) and non-tropics (50%); RF showed little change in the tropics and a 34% decline in the non-tropics; Houghton showed little change in either region. Compared to previous studies that did not include N dynamics, modeled net LULUC emissions were higher, particularly in the non-tropics. In the model, N limitation reduces regrowth rates of vegetation in temperate areas resulting in higher net emissions. Our results indicate that exclusion of N dynamics leads to an underestimation of LULUC emissions by around 70% in the non-tropics, 10% in the tropics and 40% globally in the 1990's. The differences due to inclusion/exclusion of the N cycle of 0.1 GtC/yr in the tropics, 0.6 GtC/yr in the non-tropics and 0.7 GtC/yr globally (mean across land cover data sets) in the 1990's were greater than differences due to the land cover data in the non-tropics and globally (0.2 GtC/yr). While land cover information is improving with satellite and inventory data, this study indicates the importance of accounting for different processes, in particular the N cycle.

3.2 Introduction

Land-use and land-use change (LULUC) refers to carbon (C) fluxes from the land due to human activity: that resulting from the use or management of land within one type of land cover (e.g., forest management for wood harvest) and changes in land-cover type (e.g. deforestation, afforestation, conversion of grasslands to pastureland). In total, LULUC was responsible for ~11% of all anthropogenic CO_2 emissions (7.8±0.4 GtC/yr fossil fuel; 1.0±0.5 GtC/yr LULUC) in the decade 2000 to 2009 (le Quéré et al., 2012).

The land and the ocean each take up about 30% of all anthropogenic C emissions (Denman et al., 2007; Le Quéré et al., 2011). The land takes up C from the atmosphere due to natural processes, affected by environmental change such as CO_2 and N fertilization effects, and climate change (e.g. longer growing seasons in northern extra-tropical forests) (Denman et al., 2007). The atmospheric measurements of [CO_2] combined with O_2: N ratios suggest that the land is currently acting as a net sink of CO_2 despite large-scale tropical deforestation (Denman et al., 2007; Raupach, 2011). Both the IPCC (Denman et al., 2007) and the Global Carbon Project (Le Quéré et al., 2012) calculate land sink due to the natural response of ecosystems to environmental change as the residual from other better-constrained flux terms and LULUC emissions calculated by models (2.5±0.8 GtC/yr, le Quéré et al., 2012). Thus this term is often known as the "residual terrestrial flux". Uncertainties in LULUC emissions propagate into uncertainties in the residual terrestrial uptake calculations, making these two terms the most uncertain in the C budget (Denman et al., 2007; Le Quéré et al., 2012).

Estimates of the flux of C from LULUC vary widely between different model estimates (Houghton et al., 2012). According to the most recent IPCC assessment (Denman et al., 2007), C emissions due to LULUC for the 1990's had a range of 0.5 - 2.7 GtC/yr, with a median value of 1.6 GtC/yr based on two results: the Houghton (2003) book-keeping model and FAO (2005) data, and the tropical satellite study of DeFries et al. (2002) also using the Houghton book-keeping model. With improvements in data on land cover change and biomass, and better understanding, information and modeling of different land processes, the mean estimate has been revised downwards and the range across results is reduced despite the much larger number of modeled estimates now published. A recent inter-comparison study of many published estimates reported a mean, standard deviation and range across 13 process-based vegetation models and book-keeping models of 1.1 ± 0.2 GtC/yr (full range 0.75 - 1.50 GtC/yr) for the 1990's

(Houghton et al., 2012). The authors of the inter-comparison used the limited amount of literature assessing uncertainty in LULUC emission estimates, along with expert judgment to suggest an uncertainty of ± 0.5 GtC/yr.

It is widely acknowledged that a key uncertainty in LULUC emissions stems from uncertainties in estimating historical changes in areal coverage between forests, croplands and grassland, though the uncertainties have significantly narrowed with time mainly due to improved data from satellites and inventories (Houghton, 2010; Hurtt et al., 2011; Klein Goldewijk and Ramankutty, 2004; Lepers et al., 2005; Ramankutty et al., 2007; Verburg et al., 2011). Further uncertainty stems from incomplete understanding of all the processes affecting the net flux of C from LULUC, different approaches adopted to calculate emissions, and data related uncertainties. Several previous inter-comparison studies (e.g. Houghton et al., 2012; Ito et al., 2008; Ramankutty et al., 2007) have evaluated the overall range of uncertainty associated with estimates of net flux of C resulting from LULUC. However, complex linkages between the various contributing factors have made it difficult to quantify and attribute the resulting uncertainties to each of its sources.

In an earlier study, Jain and Yang (2005) quantified the uncertainties resulting from using two different but commonly used land-use change data sets (RF - Ramankutty and Foley, 1999; and Houghton and Hackler, 2001) to drive the C cycle component of a land-surface model, the Integrated Science Assessment Model (ISAM) for the time period 1765 – 1990. Differences in the rates of changes in cropland area between the two data sets contributed significantly to uncertainty in estimated C fluxes, and argued that further refinement of land use data sets using ground and satellite-based measurements was required. The Jain and Yang (2005) study was useful in explaining and quantifying the uncertainty due to LULUC on C flux as a part of wider studies on estimating LULUC related uncertainties (Piao et al., 2008; Ramankutty et al., 2007; Ricciuto et al., 2008).

In recent years, several LULUC data sets have been updated. Improvements have primarily taken place on three aspects: Using historical inventory data with higher level of spatial detail; integrating multiple and advanced high-resolution satellite estimates; an improved methodology to downscale inventory data to grid cell level. Three of the most commonly used data sets were harmonized using a globally consistent methodology by Meiyappan and Jain

(2012): (1) and (3) The HYDE spatially explicit data set (Klein Goldewijk et al., 2010, 2011) which is the basis of the Hurtt et al. (2011) data set supplied for Earth System Models being used in the upcoming IPCC Fifth Assessment Report; (2) The spatially explicit RF data set (Ramankutty and Foley, 1999), updated to include pasture conversions (Ramankutty et al., 2008); and (3) The Houghton data set (Houghton and Hackler, 2001) updated with FAO (2005) forest area data (Houghton, 2008, the version that was used by Meiyappan and Jain, 2012) and more recently with FAO (2010) data which substantially revised down deforestation rates for the 1990's

The effects of inclusion of different processes in calculating LULUC fluxes have been explored with various process-based global vegetation models. Several studies have shown that emissions from LULUC activities are different when considering the fertilization effects of changing [CO_2] on ecosystem C balance (Churkina et al., 2008; Pongratz et al., 2009, Arora and Boer, 2010). Most process models now include the effects of climate and CO_2 on vegetation, but few include the effects of nitrogen (N).

N is a limiting nutrient for plant growth in mid- and high-latitude regions (Vitousek and Howarth, 1991). In tropical regions, N is not considered a limiting nutrient, because the warmer and wetter tropical climate enhances N mineralization in soils (Vitousek and Howarth, 1991) and biological N fixation is high (Yang et al., 2009). The N cycle is rapidly changing due to human activity (Canfield et al., 2010; Galloway et al., 2004, 2008). Enhanced N in the atmosphere can act as a pollutant or have a fertilization effect on plants (Reay et al., 2008). Climate, CO_2 and N all interact to alter plant growth (Jain et al., 2009) and decomposition, thus affecting both the C lost when vegetation is removed, and the rate of C accumulation in regrowing vegetation and soils (Mathers et al., 2006).

A recent modeling study by Zaehle et al. (2011) indicates that anthropogenic N inputs account for about a fifth of the C sequestered by terrestrial ecosystem between 1996 and 2005. Churkina et al. (2008) estimated a C uptake of 0.75 - 2.21 GtC/yr during the 1990's by re-growing forest in response to enhanced N deposition. The wide ranges in their study arise from assumptions made about proportions and age of re-growing forests. However neither study included the effects of LULUC.

Yang et al. (2010) modeled for the first time the effect of including a fully coupled N cycle (in ISAM) on global LULUC. ISAM results indicated that the contribution of N deposition to C uptake was about 0.13 GtC/yr in regrowing secondary forests, and 0.31 GtC/yr in all ecosystem types. Consideration of full N dynamics limited C uptake due to N limitation in regrowing forests in northern temperate regions in particular. The study was very sensitive to land transitions in tropical regions. While N is not a limiting nutrient in primary tropical forests, the results suggested strong N limitation in the secondary forests of tropical regions, because land use change activities (harvesting, burning) remove large amounts of N from the system. N removal due to LULUC constrained the fertilizing effects of N deposition and atmospheric CO_2 in some regions, but less in others depending on climatic conditions emphasizing the need to consider the interactive effects of all three drivers (climate, CO_2, N) on net LULUC flux.

In this paper, we build upon our previous studies to provide revised estimates of C emissions from historical LULUC looking for the first time at the effects of N under different LULUC scenarios. This study presents several crucial updates on multiple fronts, in particular: (1) We use a fully coupled Carbon-Nitrogen (C-N) cycle component of the ISAM (Yang et al., 2009), very few of the current generation of global vegetation models include a N cycle component, and only ISAM has been applied specifically to estimate LULUC emissions (2) The study incorporates the impact of N limitation and N deposition on the C sink associated with secondary forest regrowth including the effects of wood harvest activities (Yang et al., 2010), (3) The estimates have been extended until the year 2010 where possible, and (4) We use three historical reconstruction of LULUC (Meiyappan and Jain, 2012; data available from http://www.atmos.illinois.edu/~meiyapp2/datasets.htm) based on new and updated data sets (Klein Goldewijk et al., 2011; updated estimates based on Ramankutty and Foley (1999) and Ramankutty et al. (2008); and, Houghton, 2008). In addition, all the three reconstructed data sets include the effects of urban land expansion (Klein Goldewijk et al., 2010) and wood harvest (Hurtt et al., 2011).

3.3 Materials and methods

Overview of the ISAM C-N model

The C-N cycle component of the Integrated Science Assessment Model (ISAM) is used to assess the C emissions from LULUC. The structure, parameterization, and performance of ISAM has been previously discussed in detail (Jain and Yang, 2005; Jain et al., 2009; Yang et al., 2009). Here, we provide an overview. The model calculates C and N fluxes between vegetation and the atmosphere, above and below ground litter, and soil organic matter compartments of the terrestrial biosphere at 0.5^{o}x0.5^{o} spatial resolution. The modeled C cycle accounts for important feedback processes, including impact of increasing atmospheric [CO_2] on NPP; impact of temperature and precipitation changes on photosynthesis, autotrophic and heterotrophic respiration; and the effect of N deposition on C uptake by plants. The modeled N cycle accounts for major processes as described in Yang et al. (2009). In addition, the model accounts for both symbiotic and non-symbiotic biological N fixation. The performance of ISAM and its N cycle has been extensively calibrated and evaluated using field measurements (Jain et al., 2005; Yang et al., 2009).

Each 0.5^{o}x0.5^{o} grid cell contains at least one of the eighteen land-cover types (Yang et al., 2010), of which ten are forest land-cover types and the other three cropland, pastureland and urban land. ISAM accounts for five climatic types of primary forest (tropical evergreen, tropical deciduous, temperate evergreen, temperate deciduous and boreal) and their corresponding "secondary forests". The model accounts separately for forest regrowth following agricultural abandonment and wood harvest, and this is what we refer to as "secondary forest" (Yang et al., 2010).

The land conversions in the model are carried out based on the method described in Meiyappan and Jain (2012). We start with a map of potential natural vegetation at 0.5^{o} x 0.5^{o} resolution, which is indicative of the land cover that would have existed if human activities were absent. We then advance in time (starting from 1765 to 2010), by superimposing the year-to-year cropland, pastureland, wood harvest and urban land maps in the same order of preference. We define rules, specific to each land disturbance activity (cropland, pastureland, wood harvest and urban land), for replacing natural vegetation. In general, following cropland and pastureland expansion, the natural vegetations present in a grid cell are removed proportional to its area and demand for cropland/pastureland. Upon abandonment (reduction in cropland/pastureland area between two consecutive years), the land recovers back to the dominant potential natural

vegetation in the grid cell. Wood is preferentially harvested from primary forest, and secondary (regrowing) forest is used when the extent of primary forest is less than the demand. Urban land expansion usually occurs at the expense of cropland abandonment and in other cases from natural vegetations. The resulting land cover maps for the period 2000 - 2005 are compared with remote sensing based land cover maps (500m resolution MODIS data - Friedl et al., 2010) spanning the same period. Discrepancies in forest area between satellite data and model estimates are used to accordingly adjust the land-disturbance activity specific rules to increase (or decrease) the proportions at which forest was cleared (or regrown) historically following expansion (or abandonment) of agricultural activity, such that rerunning the model with adjusted rules results in land cover maps whose forest distribution closely agrees with remote sensing observations for recent years. Thus, the three reconstructions start with a common potential natural vegetation map and end with a map whose forest distribution are consistent with satellite estimates, but the pathway they follow between the starting and ending point is constrained by the land-use data sets used.

Emissions of C due to LULUC are calculated as described in Jain and Yang (2005). In brief, upon removal of natural vegetation from a grid cell, a specified fraction of vegetation biomass is transferred to litter reservoirs, effectively representing plant material left on the ground following deforestation activities (Yang et al., 2009). The remaining vegetation materials are either burned to clear the land for agriculture, which releases C and N (in gaseous and/or mineral form) contained in the burned plant material; or is transferred as C and N to wood and/or fuel product reservoirs and subsequently released at three different rates depending on the assigned product categories.

LULUC data

The three historical land-cover data reconstructions (ISAM-HYDE, ISAM-RF and ISAM-HH) were based on cropland and pastureland area change in the three updated historical land use change data sets: (1) HYDE 3.1 (Historical Database of the Global Environment) (Klein Goldewijk et al., 2011), (2) RF (Ramankutty and Foley,1999) including new pastureland estimates and updated cropland estimates based on and Ramankutty et al. (2008), and (3) Houghton and Hackler (2001) deforestation estimates updated in Houghton (2008) with revised deforestation rates from FAO (2005) respectively. The HYDE and RF data sets are both based on

FAOSTAT agricultural statistics including data on change in agricultural land area (FAO, 2009) which is available from 1960, making assumptions on the change in other land cover (e.g. forest) to meet agricultural demand. The Houghton (2008) data set is based primarily on FAO Forest Resource Assessment area change and biomass data (FAO, 2005) making assumptions about change in other land cover (e.g., croplands, pasture) to account for forest area change, supported by FAOSTAT data. A variety of other historical information is used to estimate land use transitions prior to the availability of FAO data in each data set. A common spatially explicit data set for wood harvest based on FAO data (Hurtt et al., 2011) and urban land extent (Klein Goldewijk et al., 2010) was applied to all three reconstructions. ISAM-HYDE, ISAM-RF and ISAM-HH estimates start from the year 1765 and extend until 2010, 2007 and 2005 respectively. All three reconstructions start with a common land-cover map during 1765 and follow different pathways as determined by the land-use data sets to attain forest area distributions close to satellite estimates of forests for recent years. The sum of non-forested land-cover types (herbaceous vegetation, cropland, pastureland and urban land) matches satellite estimates. However, there are discrepancies between the land-use data sets and satellite estimates in the extent of individual herbaceous land-cover types.

Model Simulations Performed

The ISAM was initialized with an atmospheric [CO_2] of 278 ppmv, representative of approximate conditions in the starting year (1765 AD) of the model simulation, to allow vegetation and soil C pools to reach an initial steady state. During the time period of 1765 - 2010, net C exchanges between atmosphere and terrestrial ecosystems are calculated based on observed changes in climate (updated estimates based on Mitchell and Jones, 2005), atmospheric [CO_2] (Meinshausen et al., 2010), wet and dry atmospheric N deposition rates (Galloway et al., 2000), and three distinct historical reconstructions of LULUC as harmonized in Meiyappan and Jain (2012).

Two separate model runs are carried out to calculate the contribution of LULUC to the terrestrial C fluxes (Table 3.1). In the first model run (A1), atmospheric [CO_2], climate and N deposition rates are varied with time based on prescribed values and the LULUC is assumed to be zero over time. In the second model run (A2), atmospheric [CO_2], climate, N deposition rates and LULUC are varied with time. The second model run (A2) was performed for each of the

three historical LULUC reconstruction used in this study. The emissions due to LULUC are estimated by subtracting C fluxes calculated in first model run (A1) from the second model run (A2). With this approach we captured the interactive effects of CO_2, Climate and N limitation on LULUC emissions.

We carried out two additional model runs (B1 and B2) to study the impact of excluding the interactive effects of N limitation on LULUC emissions (Table 3.1). Both experiments B1 and B2 are similar to A1 and A2 respectively, but they did not include the effects of N limitation. Land is always assumed to have sufficient N for plant growth. Subtracting carbon fluxes calculated in experiment B1 from that of B2 provides an estimate of LULUC emissions that only includes the interactive effects of CO_2 and climate. This (B2-B1) model experiment is analogous to the majority of other model approaches to calculating the LULUC flux in models that include only climate and CO_2 effects (e.g. McGuire et al, 2001; Pongratz et al., 2009; Piao et al., 2009; van Minnen et al., 2009, Arora et al. 2010 non-interactive runs; Stocker et al. 2011). The difference between the two sets of experiments (A2 – A1) and (B2 – B1) is an indicator of the effect of additionally considering N cycle effects and its interactions with CO_2 and climate on LULUC fluxes. We did not look at the effects of N on LULUC alone (ie excluding climate and CO_2 effects) as the paper attempts the best quantification of LULUC including all possible drivers and processes, and to assess the possible uncertainty in LULUC estimates by failing to account for N effects.

1) *LULUC flux including N effect = A2 (Δ climate, CO_2, N, LULUC) – A1 (Δ climate, CO_2, N)*

2) *LULUC flux excluding N effect = B2 (Δ climate, CO_2, LULUC) – B1 (Δ climate, CO_2)*

3) *Effect of N on LULUC flux = (B2 – B1) – (A2 – A1)*

3.4 Results

Global net LULUC emissions based on different land cover reconstructions

Large inter-annual variations in global net C emissions from LULUC are observed in the model runs based on each of the three data sets, for the period 1900 - 2010 (Figure 3.1). These variations are mainly induced by the effects of inter-annual variations in climate on LULUC fluxes. In particular, soil respiration, decomposition of slash and litter, and NPP in

growing vegetation are affected by changes in temperature and precipitation in both the runs subject to LULUC (A2 and B2) and those not (A1 and B1) (e.g. McGuire et al., 2001; Jain and Yang, 2005). Since the natural vegetation responds to climate drivers the same way in B1 and B2, the flux shown here (B2-B1) here reflects the combined effect of LULUC and climate variability (in addition to CO_2 and N) on the land affected by LULUC only.

From 1900 to 2005, the global cumulative net emissions from LULUC were 178, 160 and 163 GtC for ISAM-HYDE, ISAM-RF and ISAM-HH respectively. The ISAM-HYDE estimated global total C emissions for the time period 1900 - 2010 were 180 GtC. *(All data in this section are from model runs including the N dynamics unless otherwise stated)*. All three estimated emission trajectories show substantially different trends over the period 1900 to 1960, although all have a mean value of ~1.5 GtC/yr (Figure 3.1). The net emissions based on all three data sets peaked in the 1950's, with ISAM-HH reaching its peak slightly later than the other two data sets. This result from rapid deforestation due to expansion of agriculture in the tropics around the early 1950's followed by a rapid reduction in the rates of deforestation around the late 1950's and early 1960's, with less of a reduction based on ISAM-HH data. Emissions estimates based on ISAM-HH data are very different from those based on ISAM-RF and ISAM-HYDE in the 1960's. Emissions over the last three decades then follow similar trends based on all three data sets; an increase from 1970 to 1990 and a decline since 1990.

The mean decadal net emissions based on ISAM-HYDE data are higher during the 1980's and lower during the 1990's and 2000's compared to other two data sets, which show similar emissions during the 1980's and 1990's (Table 3.2). Thus the decline in emissions from the 1980's to the 2000's is much more pronounced in ISAM-HYDE. The reasons can be found looking at the rate of conversion of land types in the underlying harmonized data sets (Figure 3.2). ISAM-HYDE shows a sharp decrease in the expansion rates of both cropland and pastureland between 1980 and 2005 (Figure 3.2a-d), and a sharp decrease in deforested area (Figure 3.2e) which is offset to a lesser extent each decade by a declining expansion of the area of secondary forest regrowth (Figure 3.2f) (partly reforestation on abandoned agricultural land and partly conversion of "natural" forests to secondary regrowth forests after wood harvest). In contrast, ISAM-RF and ISAM-HH data show an increase in conversion to cropland (Figure 3.2a, b) and a decrease in conversion of forests to pastures (Figure 3.2c). Both ISAM-RF and ISAM-

HH show an increase in the expansion rate of secondary forest regrowth from the 1980's to the 2000's partly offsetting the loss of primary forest area (Figure 3.2e, f).

Emissions based on ISAM-HH data become higher than the other two estimates in 2000 to 2005 (Figure 3.1 and Table 3.2) because the conversion of forests to croplands and pastures (Figure 3.2a, c), and hence the overall area of deforestation (Figure 3.2e) are higher.

Regional differences in LULUC emissions

There are substantial differences in regional estimates of LULUC emissions between the model results based on the different data sets (Table 3.2). Except for Tropical America, Eurasia and China, there is no consistent trend exhibited among the three estimates. These three regions show a generally decreasing trend between the 1980's to the end of the data set for all three data-sets, with the decline being much more pronounced in ISAM-HYDE than in ISAM-RF and ISAM-HH.

Land-use change emissions based on ISAM-HYDE have decreased substantially over the last three decades for the tropics (30% decline) and non-tropics (50%). In contrast, the estimated emissions based on ISAM-RF show very little change in the tropics and a smaller decrease in the non-tropics (30%) between 2000 to 2005 compared to both the 1980's and 1990's, which were very similar. ISAM-HH shows very little change in the tropics, and a small increase from the 1980's to the 1990's then a similar decline again to the 2000's (2000 to 2005 average) in the non-tropics.

Over the last three decades, net emission estimates based on ISAM-HH data are higher for tropical regions and lower for non-tropical regions compared to net emission estimates based on other two data sets (Table 3.2). This is because ISAM-HH data shows much higher deforestation rates for agricultural land in tropical regions (especially in Tropical America) (Figure 3.2a, c, e). In non-tropical regions, the ISAM-HH data set (based on forest statistics) has lower conversion of forests to croplands than the other two datasets, and assumes no clearing of forests for pastureland (forest clearing would have been assumed converted to cropland or secondary forests). The other two data sets (based on agricultural statistics), derived based on a rule-based approach to clear vegetation, have a fraction of pastureland expansion at the expense of forests (Meiyappan and Jain, 2012) (Figure 3.2c). Houghton (2008) (which forms the basis for

ISAM-HH) assumes that the expansion of pasture area in North America, China and Pacific Developed regions occurred in the 1950's, and therefore has negligible impact on C emissions for recent years. On the other hand, ISAM-HYDE and ISAM-RF indicate that in the non-tropics forest area was converted to pastures over the last three decades (Figure 3.2c).

In the non-tropics, forest regrowth area is generally higher in ISAM-HYDE and ISAM-RF than in ISAM-HH across all three periods (Figure 3.2f). Forest regrowth would be expected to have increased the C stocks in secondary forest ecosystems (Jain et al., 2009; Reay et al., 2008; Shevliakova et al., 2009; Yang et al., 2010; Churkina et al., 2008) partially offsetting the higher emissions from forest to pasture/cropland conversion we see in ISAM-HYDE and ISAM-RF than in ISAM-HH in the non-tropics. However, the net non-tropical emissions of ISAM-HYDE and ISAM-RF remain higher than ISAM-HH. Part of the reason for this is that the regrowth is limited in the model due to N availability, and therefore the CO_2 fertilization effect is constrained.

Effects of including the N cycle

Including the N cycle in the model resulted in higher net emissions compared to the model runs without the interactive N cycle (Table 3.3, numbers in brackets are runs without the N-cycle). These results indicate that failing to account for the effects of N dynamics may lead to an underestimate in LULUC emissions by around 40% globally across all three data sets. The effects were more pronounced in non-tropical regions, where simulations without the N cycle were lower by 61 to 76% across all three data sets, while in the tropics emissions were lower only by 7 to 9%.

3.5 Discussion

Comparison with Other Studies

Our mean estimate of global net LULUC emission with N dynamics and wood harvest of 1.68 GtC/yr (range across results 1.48 to 1.83 GtC/yr) for the 1990's is the highest compared to the other published estimates as shown in Table 3.3 (excluding Denman et al., 2007 which is a synthesis based on old estimates). Breaking it down regionally, where other published estimates were available for comparison, our net emissions are similar in the tropics (mean 0.78 GtC/yr)

but much higher in the non-tropics (0.90 GtC/yr). While other published results find that tropical emissions are higher than non-tropical emissions, our estimates based on two data sets (ISAM-RF and ISAM-HYDE) with N dynamics show the opposite trend, i.e. higher LULUC net emissions for non-tropics than tropics. Our modeling results indicate that without considering the N dynamics effect, the estimated non-tropical LULUC emissions for ISAM-RF, ISAM-HYDE and ISAM-HH cases are underestimated by 0.66, 0.58 and 0.53 GtC/yr respectively for the 1990's, emphasizing the importance of including N dynamics in estimating LULUC emissions. The range of non-tropical emission estimates when N dynamics is excluded in our study (0.17 - 0.43 GtC/yr) are not only well within the range of values of other published studies, but also lower than estimates for the tropics.

N is usually not considered as a limiting nutrient in the tropical regions, because warmer and wetter tropical climate enhance N mineralization in soils, and biological N fixation is high. Therefore, it is not surprising that ISAM estimated tropical emission with (0.56 – 1.13 GtC/yr) and without (0.51 - 1.04 GtC/yr) N dynamics are approximately the same as eachother (Table 3.3), and are well within other model estimated range of values (0.0 - 1.44 Gt/C).

It is interesting to note that Houghton's own book-keeping model estimates (Houghton, 2010) are the highest for the tropics and the lowest for the non-tropics as compared to other model estimates (Table 3.3). This is, unsurprisingly, similar to the results we found using the ISAM-HH data set compared to the other data sets within our modeling study, as it is driven by the underlying data assumptions in the Houghton data set based on FAO FRA forest data (FAO, 2005). The FAO data indicate a net loss of total forest area in the tropics, and vice-versa in the non-tropics (Houghton, 2010) for the last three decades. In contrast, other data sets (HYDE or RF) used by other modeling studies indicate a decrease in forest area for both tropics and non-tropics (This cannot be directly interpreted from the data in Figure 3.2 as some of the area of primary deforestation goes to secondary forests after harvesting and some does not, likewise only a portion of secondary forest regrowth happens on deforested land, some happens on agricultural land, so the numbers cannot be directly summed to get net change in forest area). Note that the latest FAO FRA (FAO, 2010) substantially revised down deforestation rates in the tropics.

The land cover data may not be the full reasons for discrepancies. Houghton (2010) is even higher than our ISAM-HH results in the tropics and even lower in the non-tropics. Thus, it

might also be partly due to the differences in the modeling framework used by Houghton (2010) and other studies shown in Table 3.3. Houghton (2010) estimates are based on book-keeping model that tracks areas of land conversions and calculates subsequent changes in C pools using standard growth and decay curves derived from actual field inventory data from the literature that are unchanging over the calculation period (representing either recent or historic climate and environmental conditions) and averaged over a large region or vegetation type. Most other modeling studies, with the exceptions of satellite based tropical region estimates of DeFries et al. (2002) and Achard et al. (2004), model soil and vegetation processes and how they are affected by climate, atmospheric CO_2, and, in this study, N drivers that vary spatially and possibly temporally. A sensitivity analysis based on process-based model and book-keeping model approaches suggests that book-keeping model estimated LULUC emissions were about 40% higher than the process based modeling approach, due primarily to higher soil carbon emissions assumed to be 25% soil carbon loss following land use change (Reick et al., 2010).

Most process based studies, including this study, use historical transient CO_2 and climate as an external driving force and run the model with and without land use and derive the LULUC emissions as the difference (e.g. McGuire et al., 2001; Pongratz et al., 2009 "LULUC+CO_2"; van Minnen et al., 2009; Piao et al., 2009; Stocker et al., 2011). Shevliakova et al. (2009) ran with present climate and CO_2 in the both the with- and without-LULUC simulations.

The LULUC past emissions not only affect the "managed" vegetation that is subject to LULUC, but also the "natural" or "primary" vegetation. This has been referred to as the "feedback flux' (Strassman et al., 2008) or the "coupling flux" (Pongratz et al., 2009). The feedback flux on natural vegetation is typically to be considered part of the "residual terrestrial flux" as it is an indirect effect of human activity and not considered as part of net LULUC emissions. In the case above where LULUC emission are derived by the difference between the no-LULUC case and with-LULUC cases, the effects of past LULUC emissions on the natural vegetation are factored out, only the past LULUC effects on the vegetation that is subject to LULUC is included. However some coupled climate-carbon cycle model studies such as Arora and Boer (2010) include the effects of LULUC emissions on natural vegetation which is why their flux of 0.25 - 0.84 GtC/yr in the 1990's based on different data sets are much lower than other estimates, including our own. When they apply the same approach as we use here, their

estimated emissions based on RF data increase from 0.71 GtC/yr (their Fig 3.10a, thin orange line) to 1.06 GtC/yr (their Fig 3.10a, thick orange line) (pers. comm. data supplied by Arora for analysis). The interactive effects of LULUC on atmospheric [CO_2] merit further investigation, but are beyond the scope of this study.

Our modeled LULUC emissions for the 2000's vary between 1.2 - 1.7 GtC/yr (Table 3.2), consistent with, but at the high end of most recent estimated range across a number of published studies of 0.4 - 1.8 GtC/yr (Houghton et al., 2012).

Uncertainty in LULUC Emissions Estimates

Our modeled estimates give an indication of uncertainty in LULUC emissions due to the choice of data set. Estimated ranges across the three data sets for 1980's, 1990's and 2000's respectively were ±0.26 GtC/yr, ±0.18 GtC/yr and ±0.21 GtC/yr. The estimated uncertainty due to data set variability is much lower than other uncertainty estimates (see below) partly reflecting more accurate and revised land-use data sets applied in a globally consistent methodology to produce historical LULUC estimates (Meiyappan and Jain, 2012) but also as it does not account for uncertainty in other data, the model approach or implementation.

Our results further indicate a large uncertainty due to the missing process of the N cycle in other estimates. Failure to account for the N cycle may underestimate net C flux due to LULUC by 0.1 GtC/yr in the tropics, 0.6 GtC/yr in the non-tropics and 0.7 GtC/yr globally (mean across land cover data sets).

A recent meta-analysis study by a range of experts for the Global Carbon Project (Houghton et al., 2012), estimates the total errors resulting from data related uncertainty and incomplete understanding of all the process to be in the order of about ±0.5 GtC/yr based on expert judgment, drawing on the range across many published model studies, and studies that specifically looked at uncertainty due to data or modeling approaches. Previous publications for the Houghton book-keeping model approach gave an uncertainty estimate of ±0.7 GtC/yr (Houghton, 2010), that have since been revised down to ±0.5 GtC/yr (Houghton pers comm). The most recent IPCC estimated uncertainty of ±1.1 GtC/yr for 1990's (Denman et al. 2007) can now be considered too high. The higher end based on Houghton (2003) was revised downwards due to the reduction in the deforestation estimates for tropical regions in subsequent FAO FRA

(FAO 2005, 2010) brought about by integration of satellite-based estimates (e.g. Nepstad et al., 2009; Hansen et al., 2009). The lower end of the range was based on DeFries et al. (2002) is an underestimate, as it is based on satellite measurements for three tropical regions, and does not account for legacy emissions deforestation rates prior to the period of analysis (1980's and 1990's) (Ramankutty et al., 2007).

Differences in Land Processes Included

In our study, secondary forest regrowth only occurs as a result of wood harvest and agricultural abandonment on land that was originally covered by forests (i.e. a reduction of agricultural area in a grid cell will regrow forest). In some countries or regions, for example North America Europe, Japan, China and India (Kenji, 2000; Merker et al., 2004; FAO, 2005; FAO, 2010), there are active programs of afforestation and reforestation. These may not be captured by the data sets of change in agricultural and pasture areas, particularly if the forests are established on previously grassland areas, or if they shift agriculture to grassland areas so the agricultural area does not decline. Hence, our study may be underestimating the forest area in some regions and hence the C uptake by the afforested land.

This study does not include the effects of fire suppression and woody encroachment, which are suggested to contribute to regional C sink (e.g. in the USA, see Pacala et al., 2001). This is because the effects of these processes have not yet been well defined due to lack of comprehensive data (Denman et al., 2007).

C emissions due to the common practice of shifting cultivation in the tropics (clearing forest often by fire for agriculture then abandoning to regrowth after a number of years) are estimated to have a significant impact on historical LULUC emissions (Hurtt et al., 2006, 2011). This creates a mosaic of cropped fields often with trees and fallows intermixed with secondary and mature forests and cause some loss of ecosystem C (Houghton and Hackler, 2006). We did not specifically model the effects of shifting cultivation due to huge uncertainties in magnitude and spatial distribution, and as some of these effects would be captured in the data sets of changing forest or agricultural area we already used (Hurtt et al., 2006, 2011).

Natural disturbances such as fire, pests, disease, drought, wind, snow, ice, and floods affect 104 Mha of forest on average each year (FAO, 2006), with local- to national-scale

ecological significance (e.g. Giglio et al., 2010; van der Werf et al., 2010; also see Lambin et al., 2003 and Foley et al., 2003). Our study has not considered emissions due natural disturbances because it is not human induced LULUC, and in any case it is typically assumed that disturbance is followed by regrowth and the net effects are minimal (unless the land is subsequently converted to agricultural land).

A key missing processes is the decomposition of soil C following drainage of tropical peatlands (Ballhorn et al., 2009). According to Hooijer et al. (2010), draining and burning of peatlands in Southeast Asia are thought to add another 0.3 GtC yr^{-1} to land-use emissions.

Summary and implications of results for climate modeling and climate policy

Emissions of CO_2 from LULUC constitute a significant portion of global emissions, and therefore strongly affect global climate. Modeling them correctly has implications for global climate policy. The estimated cumulative LULUC emissions over the period 1900 - 2010 based on ISAM-HYDE data are ~180 GtC, which are ~33% of total C emissions (345 GtC from burning fossil fuels - Friedlingstein et al., 2010). The contribution of LULUC to global anthropogenic C emissions (land-use plus fossil fuel) in 1990's and 2000's were ~18 - 22% and 14 - 17% respectively (using fossil fuel emissions as in Le Quéré et al., 2012) for our modeled results across three underlying data sets and including the N cycle.

Our estimated net global emissions from LULUC (mean and range) across three data sets are 1.88 (1.7 to 2.21) GtC/yr for the 1980's, 1.66 (1.48 to 1.83) GtC/yr for the 1990's, and 1.44 (1.22 to 1.65) for the 2000's (Table 3.2). Our estimates are higher than other published estimates that range from 0.80 to 1.5 GtC/yr for the 1990's (Table 3.3: Achard et al., 2004; Arora and Boer, 2010; DeFries et al., 2002; Houghton, 2010; Piao et al., 2009; Pongratz et al., 2009; Stocker et al., 2011; Strassmann et al., 2008; Shevliakova et al., 2009; Van Minnen et al., 2009; Yang et al., 2010; Kato et al., 2012) and 1.1 GtC/yr for the 2000's (Houghton et al., 2012, Friedlingstein et al., 2010). If LULUC emissions are higher than assessed, it means fossil fuel emissions would have to be even lower to meet the same mitigation target.

Our results are higher than other published estimates because they include the effects of N limitation on regrowth of forests following wood harvest and agricultural abandonment. This effect is particularly noticeable in the cooler non-tropics where N removal through harvest or

burning is not compensated by N deposition or N mineralization. The estimated LULUC emissions for the tropics are 0.79±0.25 for the 1980's, 0.78±0.29 for the 1990's and 0.71±0.33 GtC/yr for the 2000's, and for the non-tropics regions are 1.08±0.52, 0.90±0.19 and 0.69±0.12 GtC/yr for the three decades (Table 3.2). Not only are our results much higher in the non-tropics than other results (Table 3.3), but for two of the data sets they are higher in the non-tropics than in the tropics. This is because the estimated non-tropical LULUC emissions with N dynamics considered are 0.53 - 0.66 GtC/yr higher than without N dynamics for the 1990's in the non-tropics and 0.62 - 0.72 GtC/yr higher globally. Without considering the N cycle, our model results of 0.85 - 1.2 GtC/yr globally, 0.51 - 1.04 GtC/yr in the tropics and 0.17 - 0.43 GtC/yr in the non-tropics in the 1990's across the three data sets are similar to other published studies (Table 3.3). Our model results indicate that failing to account for the N cycle underestimates by about 40% globally (0.66 GtC/yr), 10% in the tropics (0.07 GtC/yr) and 70% in the non-tropics (0.59 GtC/yr).

Many inventory studies in both managed and natural forests find higher sinks than in the past and attribute this to the effects of changing climate and [CO_2] (Luyssaert et al., 2008; Lewis et al., 2009; Phillips et al., 2008; Pan et al., 2011). Our results are not in conflict with this. Climate and CO_2 still enhance uptake in northern re-growth forests, but the effects are limited when N removal due to LULUC is considered. Since the total net flux of CO_2 between the land and atmosphere is known from atmospheric measurements, higher emissions from land under LULUC in fact imply a greater sink in land not experiencing LULUC and are therefore consistent with inventories finding greater sinks in unmanaged forests. The total net flux the atmosphere "sees" from the land is the same; in that sense our results do not imply different climate impacts. But our results do have implications for modeling of anthropogenic versus natural land fluxes (both natural and anthropogenic sources and sinks are underestimated without the N cycle), and thus for climate policy around estimating human-induced emissions and mitigation potential on the land.

We evaluate the uncertainties in LULUC emissions estimates resulting from uncertainties in determining land-cover change using three historical LULUC reconstructions based on our best estaitmes of LULUC that include not only climate and CO_2 but also N. Over the period 1900 - 1970, our model results for the global LULUC emissions based on three different LULUC

reconstructions exhibit substantially different trends (Figure 3.1). The global total emissions are very similar thereafter, with emissions increasing until about 1990 and then declining. Uncertainty in LULUC emissions due to the underlying data set constitutes about ±0.2 GtC/yr over the period 1980 to 2009.

While the three LULUC estimates show reasonably good agreement at the global scale, there are significant disagreements between them at the regional scale (Table 3.2). Regional discrepancies in location of CO_2 emissions are irrelevant to the global climate impacts of CO_2 as it is well mixed gas in the atmosphere. However they indicate a much larger uncertainty still exists in underlying land cover data than is implied by looking at global decadal averages and this uncertainty may affect the overall amount of global LULUC emissions and thus climate. The regional differences also have implications for national-level greenhouse gas reporting and accounting under the UNFCCC and Kyoto Protocol, and for assessing future LULUC mitigation potential. Therefore, the results presented here suggest that the uncertainty in regional LULUC data need to be reduced in order to improve climate change projections.

Regional differences in forest cover will affect regional climate through biophysical properties such as albedo, surface roughness, heat transfer and water recycling: for example afforestation in mid to high-latitudes reduces albedo and has a warming affect that runs counter to the cooling effect of CO_2 uptake (e.g. Brovkin et al., 2006; Findell et al., 2007; Kvalevag et al., 2010; Pitman et al., 2009; Pongratz et al., 2010). However, assessing the implications of regional data differences on biophysical climate effects is beyond the scope of this study.

Ongoing improvements in satellite data and interpretation for measuring not only changes in land cover, but also land management (e.g shifting cultivation selective logging) and biomass density will be critical in reducing uncertainties. Reconciling and improving data sets produced from different sources (e.g. FAO forest assessment and FAO agricultural assessments), to provide more information about land-use transitions is also expected to further reduce uncertainties.

Acknowledgements

This work was supported in part by National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change Program (No. NNX08AK75G) and the Office of

Science (BER), U.S. Department of Energy (DOE-DE-SC0006706). JH was funded by a Leverhulme Research Fellowship, UK.

3.6 Tables

Table 3.1 Design of the Simulation Experiments. Tick mark (✓) indicates the environmental factor was varied with time. Cross mark (×) indicates the environmental factor was held constant at initial value. Inclusion of N deposition is irrelevant when N dynamics is inactive in the model.

Experiment	**CO_2**	**Climate**	**N Deposition**	**LCLUC**	**N Dynamics**
A1	✓	✓	✓	×	Active
A2	✓	✓	✓	✓	Active
B1	✓	✓	-	×	Inactive
B2	✓	✓	-	✓	Inactive

Table 3.2 Regional breakdown of decadal mean net LULUC emissions (GtC/yr) for the 1980's, 1990's and 2000's based on ISAM-HYDE, ISAM-RF, and ISAM-HH data sets.

Region/Global	1980's				1990's				2000's			
	ISAM-HYDE	ISAM-RF	ISAM-HH	Mean & Range	ISAM-HYDE	ISAM-RF	ISAM-HH	Mean & Range	ISAM-HYDE [1]	ISAM-RF[2]	ISAM-HH[3]	Mean & Range
Tropical America	0.26	0.33	0.59	*0.39±0.17*	0.20	0.34	0.64	*0.39±0.22*	0.14	0.24	0.46	*0.28±0.16*
Tropical Africa	0.01	-0.03	0.11	*0.04±0.07*	0.04	-0.03	0.11	*0.04±0.07*	0.03	-0.04	0.09	*0.03±0.06*
Tropical Asia	0.34	0.35	0.40	*0.37±0.03*	0.31	0.34	0.38	*0.34±0.03*	0.25	0.43	0.53	*0.41±0.14*
Tropics Total	***0.61***	***0.65***	***1.11***	***0.79±0.25***	***0.56***	***0.65***	***1.13***	***0.78±0.29***	***0.43***	***0.63***	***1.08***	***0.71±0.33***
North America	0.30	0.28	0.19	*0.25±0.06*	0.27	0.28	0.21	*0.25±0.03*	0.25	0.23	0.28	*0.25±0.03*
Eurasia	0.71	0.60	0.29	*0.53±0.21*	0.47	0.62	0.34	*0.48±0.14*	0.39	0.46	0.22	*0.36±0.12*
China	0.59	0.15	0.08	*0.27±0.26*	0.19	0.14	0.07	*0.13±0.06*	0.12	0.09	0.06	*0.09±0.03*
Oceania	0.00	0.03	0.02	*0.02±0.01*	0.00	0.05	0.08	*0.04±0.04*	0.02	-0.08	0.01	*-0.02±0.5*
Non-Tropics Total	***1.61***	***1.06***	***0.56***	***1.08±0.52***	***0.92***	***1.09***	***0.70***	***0.90±0.19***	***0.80***	***0.69***	***0.57***	***0.69±0.12***
Global	***2.21***	***1.70***	***1.72***	***1.88±0.26***	***1.48***	***1.74***	***1.83***	***1.68±0.18***	***1.22***	***1.33***	***1.65***	***1.40±0.21***

[1]Average for the period 2000-2009
[2]Average for the period 2000-2007
[3]Average for the period 2000-2005

Table 3.3 Comparison of ISAM estimated LULUC emissions for 1990's with other model and data studies. The decade 1990 - 1999 was chosen for comparison, as most of the estimates in literature covered this time period. Estimates that do not account for N dynamics are provided in brackets.

Study	LULUC Data	Tropics	Non-Tropics	Global
This Study:	ISAM-RF	0.65 (0.59)	1.09 (0.43)	1.74 (1.02)
	ISAM-HYDE	0.56 (0.51)	0.92 (0.34)	1.48 (0.85)
	ISAM-HH	1.13 (1.04)	0.70 (0.17)	1.83 (1.21)
	Range	0.56 – 1.13 (0.51 – 1.04)	0.70 – 1.09 (0.17 – 0.43)	1.48 – 1.83 (0.85 – 1.21)
Other studies:				
Strassmann et al. (2008)	HYDE	(1.02)		(1.08)
van Minnen et al. (2009)	HYDE	(0.70)	(0.60)	(1.3)
Arora and Boer (2010)[1]	RF			1.06
Piao et al. (2011)	HYDE	(0.74)	(0.48)	(1.22)
Yang et al. (2010)	HYDE/RF			1.44 (1.03)
Houghton (2010)	Houghton	(1.44)	(0.06)	(1.50)
Pongratz et al. (2009)[2]	Pongratz			(1.30)
Shevliakova et al. (2009)	RF+HYDE pastures			(1.31)
Shevliakova et al. (2009)	HYDE			(1.07)
Kato et al. (2012)	Hurtt (HYDE)			(1.00 -1.28)
Stocker et al. (2011)				(0.93)
DeFries et al. (2002)[3]	AVHRR	(0.50-1.50)		
Achard et al. (2004)[4]	Landsat	0.60 -1.10		
Denman et al. (2007) range[5]				(0.50 – 2.70)
Houghton et al. (2012) range[6]				(0.75 – 1.50)
Other Studies Range[7]		(0.50-1.44)	(0.06 -0.48)	(0.80 – 1.50)

[1]This result is based on the data underlying the thick orange line figure 3.10a of Arora and Boer (2010), data supplied by Arora (pers comm). Their study represents the approach most similar to ours for calculating the LULUC flux (see text for details).

[2] Underlying data set described in Pongratz et al. (2008) is based on RF cropland and RF pasture with rates of pasture changes from HYDE. Pastureland was preferential allocated on natural grassland.

Table 3.3 (Cont.)

[3]Calculated using the Houghton and Hackler (2003) book-keeping in combination of AVHHR satellite data for LULUC.
[4]Calculated using the biomass and biomass change of tropical forest estimates of FAO (FAO, 1997) and Landsat data for Land-Cover Change. These estimates may have implicitly accounted for the N dynamics effect.
[5]Denman et al. (2007) is not an estimate in itself, but is a synthesis range across two estimates including uncertainty, DeFries et al. (2002) and Houghton (2003), that has since been updated and revised downwards (Houghton, 2010).
[6]Houghton et al. (2012) give the mean and standard deviation across thirteen different model estimates of LULUC as 1.12 ± 0.25 GtC/yr, full range as 0.75 - 1.50 GtC/yr. Their estimate of uncertainty in mean LULUC emissions is about ± 0.5 GtC/yr,
[7]The range values given here are based on the published studies included in this table and do not account for the ranges in Denman et al. (2007) and Houghton et al. (2012) as these are themselves ranges across other published estimates. The estimates of Denman et al. (2007) are now out of date for the reasons discussed in footnote 5.

3.7 Figures

Figure 3.1 ISAM estimated global land-use emissions for the period 1900-2010 (GtC/yr) based on ISAM-HYDE, ISAM-RF and ISAM-HH data sets. Estimates based on ISAM-RF and ISAM-HH estimates extend until year 2007 and 2005 respectively.

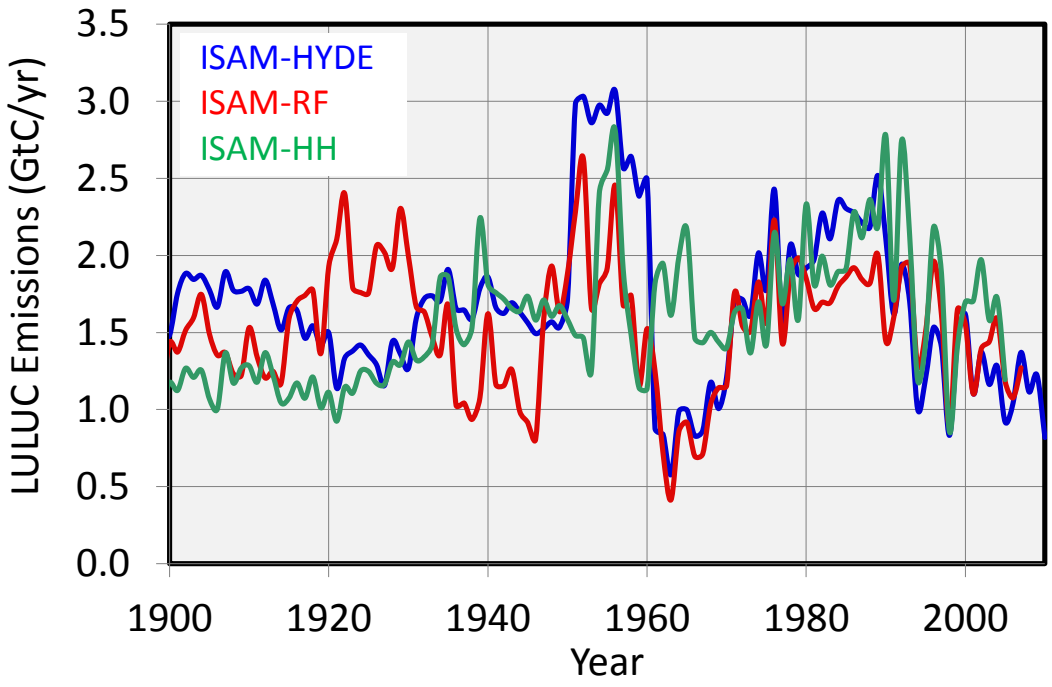


Figure 3.2 Average rate of conversions in the tropics and non-tropics from (a) forest to crop, (b) herbaceous to crop, (c) forest to pastureland, (d) herbaceous to pasturelands, and (e) deforested (includes forest area loss due to wood harvest) and (f) reforested areas due to expansion and abandonment of cropland, pastureland and wood harvest decadally.

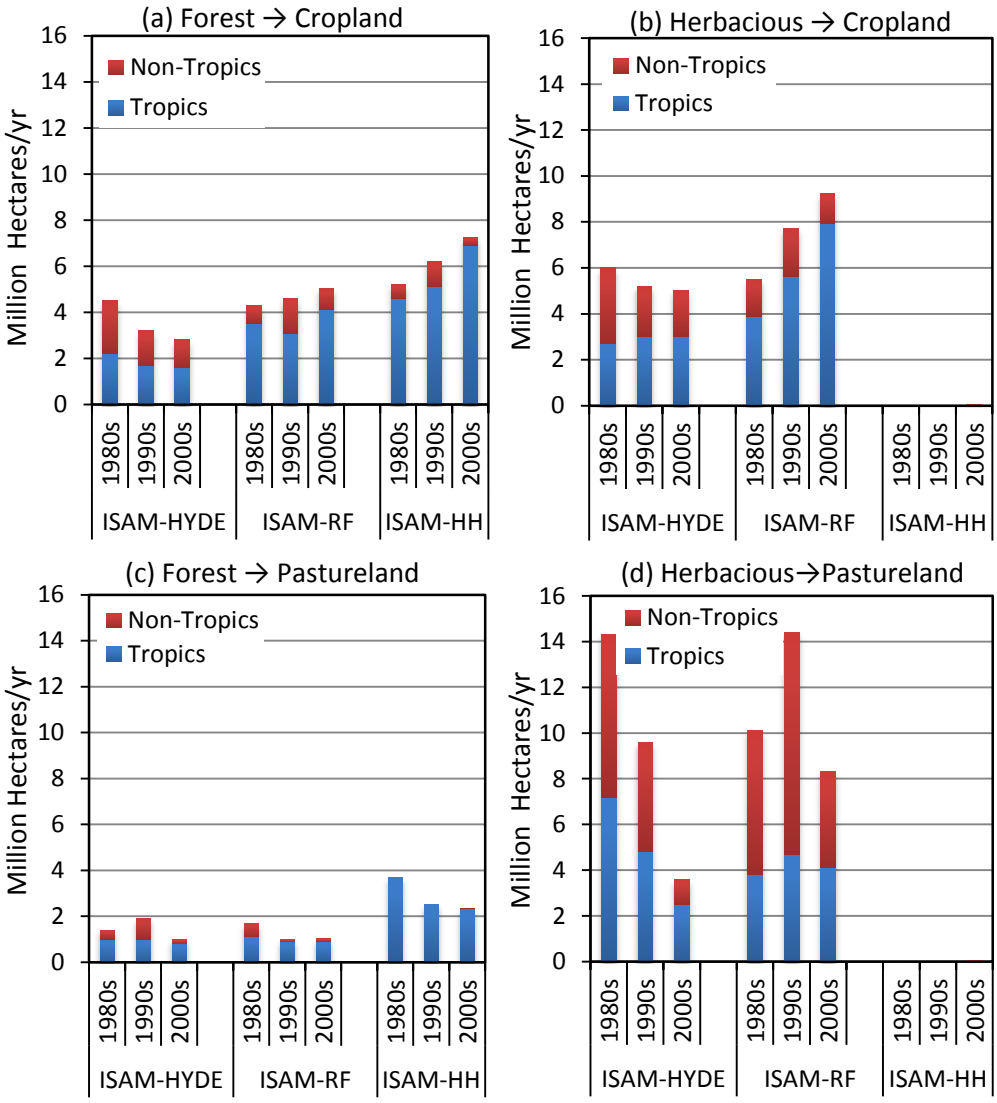
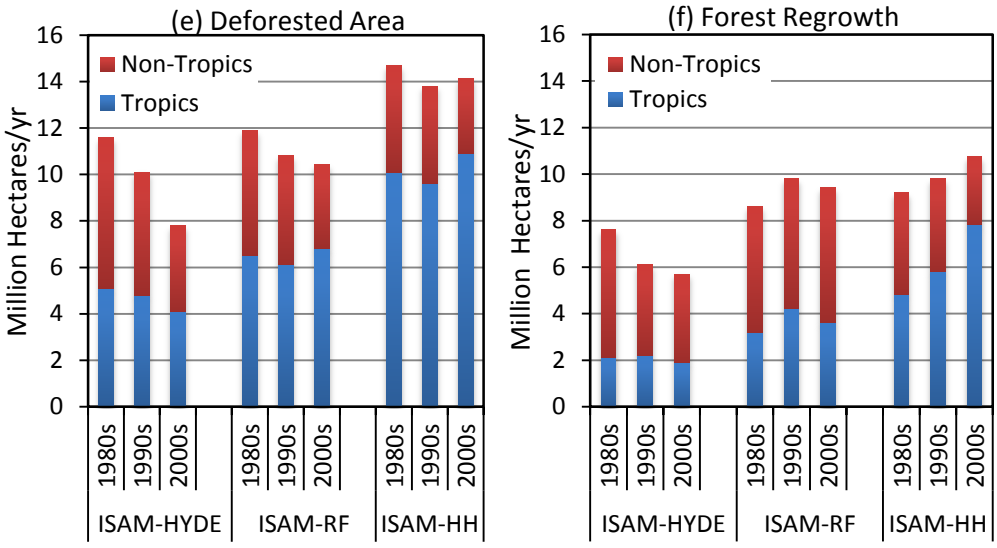


Figure 3.2 (cont.)



3.8 References

Achard, F, Eva HD, Mayaux P, Stibig H-J., and Belward A (2004) Improved estimates of net carbon emissions from land cover change in the tropics for the 1990s. *Global Biogeochem. Cy., 18*, GB2008, doi:10.1029/2003GB002142.

Arora, V and Boer GJ (2010) Uncertainties in the 20th century carbon budget associated with land use change. *Glob. Change Biol., 16*, 3327–3348, doi:10.1111/j.1365-2486.2010.02202.x.

Ballhorn U, Siegert F, Mason M, and Limin S (2009). Derivation of burn scar depths and estimation of carbon emissions with LIDAR in Indonesian peatlands, *Proceedings of the National Academy of Sciences, 106*(50), 21213-21218.

BrovkinV, Claussen M, Driesschaert E et al. (2006). Biogeophysical effects of historical land cover changes simulated by six earth system models of intermediate complexity. *Climate Dynamics, 26*(6), 587-600.

Canadell JG., Le Quéré C, Raupach MR et al. (2007) Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proc. Natl. Acad. Sci., 104*, 18866–18870.

Canfield, D. E., Glazer, A. N., & Falkowski, P. G. (2010). The evolution and future of Earth's nitrogen cycle. *Science, 330*(6001), 192-196.

Churkina, G., Trusilova, K., Vetter, M., & Dentener, F. (2007). Contributions of nitrogen deposition and forest regrowth to terrestrial carbon uptake. *Carbon Balance and Management, 2*(5).

DeFries RS, Houghton RA, Hansen MC, Field CB, Skole D, and Townshend J (2002), Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 90s. *P. Natl. Acad. Sci. USA, 99*, 14256-14261.

Denman KL, Brasseur G, Chidthaisong A, et al. (2007) Coupling between changes in the climate system and biogeochemistry. In: *Climate Change (2007), The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the International Panel on Climate Change* (eds Solomon S, Quin D, Manning M, et al.), pp 499–587, Cambridge Univ. Press, Cambridge, U. K.

FAO (1997) Estimating Biomass and Biomass Change of Tropical Forests. A primer. FAO Forestry Paper 134, Rome.

FAO (2001) Global Forest Resources Assessment 2000, Main report. FAO Forestry Paper 140, Rome, Italy.

FAO (2006) Global Forest Resources Assessment 2005. FAO forestry Paper 147, Rome.

FAO (2009 http://faostat.fao.org/site/377/default.aspx#ancor (11/09)

FAO (2010) Global Forest Resources Assessment 2010. FAO forestry Paper 13, Rome.

Feddema JJ, Oleson KW, Bonan GB, Mearns LO, Buja LE, Meehl GA, et al. (2005) The importance of land-cover change in simulating future climates. *Science, 310*(5754), 1674.

Findell KL, Shevliakova E, Milly P, and Stouffer RJ (2007) Modeled impact of anthropogenic land cover change on climate. *Journal of Climate, 20*(14), 3621-3634.

Friedl M A, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010). MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens Environ, 114(1): 168–182

Friedlingstein P, Dufresne J-L, Cox PM, and Rayner P (2003) How positive is the feedback between climate change and the carbon cycle? *Tellus,* 55B**,** 692–700.

Friedlingstein P, Houghton RA, Marland G et al. (2010) Update on CO_2 emissions. *Nature Geoscience*., ***3,*** 811-812, doi 10-1038/ngeo1022.

Foley JA, Costa, MH, Delire C, Ramankutty N, Snyder P (2003) Green surprise? How terrestrial ecosystems could affect Earth's climate. *Front Ecol Environ, 1(1),* 38–44

Giglio L, Randerson JT, van der Werf GR, Kasibhatla PS, Collatz GJ, Morton DC, DeFries RS (2010) Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences, 7*(3): 1171–1186

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., & Cosby, B. J. (2003). The nitrogen cascade. *BioScience, 53*(4), 341-356.

Galloway J, Dentener F, Capone D et al. (2004) Nitrogen cycles: past, present, and future. *Biogeochemistry, 70*(2), 153-226.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., ... & Sutton, M. A. (2008). Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science, 320*(5878), 889-892.

Hansen MC, Stehman SV, Potapov PV, Arunarwati B, Stolle F, and Pittman K (2009) Quantifying changes in the rates of forest clearing in Indonesia from 1990 to 2005 using remotely sensed data sets. *Environ. Res. Lett., 4,* 034001, doi:10.1088/1748-9326/4/3/034001.

Hooijer A, Page S, Canadell J, Silvius M, Kwadijk J, Wosten H, and Jauhiainen J (2010) Current and future CO2 emissions from drained peatlands in Southeast Asia, *Biogeosciences*, *7*(5), 1505-1514.

Houghton, RA, and Hackler JL (2001) Carbon flux to the atmosphere from land-use changes: 1850 to 1990. *ORNL/CDIAC-131, NDP-050/ R, 1*.

Houghton, RA (2003) Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850-2000. *Tellus B, 55*,378-390.

Houghton, RA (2005) Above ground forest biomass and the global carbon balance. *Global Change Biology, 11*, 945-958.

Houghton, RA (2008) Carbon Flux to the Atmosphere from Land-Use Changes: 1850-2005. In *TRENDS: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Houghton, RA (2010) How well do we know the flux of CO2 from land-use change? *Tellus Series B-Chemical And Physical Meteorology*, **62 (5)**, 337-351.

Houghton RA, House JI, Pongratz J, van der Werf, GR, DeFries RS, Hansen MC, Le Quéré C, and Ramankutty N (2012) Carbon emissions from land use and land-cover change. *Biogeosciences., 9*, 5125-5142.

Hurtt, GC, Chini, LP, Frolking S et al. (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim Change*, *109*(1–2), 117–161.

Hurtt G, Frolking S, Fearon M, Moore B, Shevliakova E, Malyshev S, Pacala S, and Houghton R (2006) The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. *Global Change Biol., 12*(7), 1208-1229.

Ito A, Penner JE, Prather MJ et al. (2008) Can we reconcile differences in estimates of carbon fluxes from land-use change and forestry for the 1990s?. *Atmos. Chem. Phys., 8*, 3291–3310.

Jain, AK, and Yang, X (2005) Modeling the effects of two different land-cover change data sets on the carbon stocks of plants and soils in concert with CO_2 and climate change. *Global Biogeochem.Cycles, 19*(2), 1-20.

Jain AK, Yang X, Kheshgi H, McGuire AD, Post W, Kicklighter D (2009) Nitrogen attenuation of terrestrial carbon cycle response to global environmental factors. *Global Biogeochem.Cycles, 23*, GB4028, doi:10.1029/2009GB003519.

Kato, E., Kinoshita, T., Ito, A., Kawamiya, M., & Yamagata, Y. (2011). Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model, Journal of Land Use Science, doi:10.1080/1747423X.2011.628705.

Kenji, K (2000) Recycling of Forests: Overseas Forest Plantation Projects of Oji Paper Co., Ltd. *Japan TAPPI Journal, 54* (1); 45-48.

Klein Goldewijk K, Beusen A, van Drecht G, De Vos M (2011) The HYDE 3.1 spatially explicit database of human-induced land-use change over the past 12,000 years. *Glob Ecol Biogeogr, 20(1):* 73–86.

Klein Goldewijk K, Beusen A, Janssen P (2010) Long term dynamic modeling of global population and built-up area in a spatially explicit way, HYDE 3.1. *Holocene, 20(4),* 565–573.

Klein Goldewijk K and Ramankutty, N (2004) Land use changes during the past 300 years. In *Land Use, Land Cover and Soil Sciences*, In *Encyclopedia of Life Support Systems* (EOLSS) (ed Verheye WH). Developed under the Auspices of the UNESCO, EOLSS Publishers, Oxford, UK, [http://www.eolss.net].

Kvalevåg MM, Myhre G, Bonan G and Levis S (2010) Anthropogenic land cover changes in a GCM with surface albedo changes based on MODIS data. *Int. J. Climatol.* **30**, 2105–2117.

Lambin EF, Geist HJ, Lepers E (2003) Dynamics of land use and land cover change in tropical regions. *Annu Rev Environ Resour, 28(1),* 205–241.

Le Quéré, C., Raupach, MR, Canadell JG et al. (2009) Trends in the sources and sinks of carbon dioxide. *Nat. GeoSci., 2*, 831–836.

Le Quéré C., Andres R.J., Boden T., et al. (2012) The global CO_2 budget 1959 – 2011, *Earth System Science Data Discussions, 5, 1107-1157.*

Leite, CC, Costa, MH, de Lima CA, Ribeiro CAAS, and Sediyama GC (2011) Historical reconsruction of land use in the Brazilian Amazon (1940 – 1995). *Journal of Land Use Science, 6*, 33-52.

Leite, CC, Costa, MH, Soares-Filho BS and de Barros Viana Hissa L (2012) Historical land use change and associated carbon emissions in Brazil from 1940 to 1995. *Global Biogeochem. Cycles*, *26*, GB2011, doi:10.1029/2011GB004133.

Lewis SL, Lopez-Gonalez G, Sonké B, Affum-Baffo K, Baker TR (2009) Increasing carbon storage in intact African tropical forests. *Nature*, 477, 1003-1006.

Li BB, Fang XQ, Ye Y, and Zhang X (2010) Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data – A case study in Northeast China. *Science China Earth Sciences,* DOI: 10.1007/s11430-010-4053-5.

Liu M and Tian, H (2010) China's land cover and land use change from 1700 to 2005: Estimations from high-resolution satellite data and historical archives. *Global Biogeochem. Cycles*, *24*, GB3003, doi:10.1029/2009GB003687.

Lepers E, Lambin EF, Janetos AC, DeFries R, Achard F, Ramankutty N, and Scholes RJ (2005) A synthesis of information on rapid land-cover change for the period 1981-2000, *Bioscience*, *55* (2), 115-124.

Luyssaert S, Schulze ED, et al. (2008) Old-growth forests as global carbon sinks. *Nature* **455**(7210): 213-215

Mathers NJ, Harms B, and Dalal RC (2006) Impacts of land-use change on nitrogen status and mineralization in the Mulga Lands of Southern Queensland, *Austral Ecol*., 31(6), 708–718.

Meinshausen M, Smith SJ, Calvin KV et al. (2011) The RCP Greenhouse Gas Concentrations and their Extension from 1765 to 2300. *Climatic Change*, DOI: 10.1007/s10584-011-0156-z.

Meiyappan P and Jain, AK (2012) Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. *Front. Earth Sci., 6(2)*: 122–139, DOI 10.1007/s11707-012-0314-2.

McGuire AD, Stitch, S, Clein, JS et al. (2001) Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO_2. climate and land use effects with four process-based ecosystem models, *Global Biogeochemical Cycles, 15,* 183-206.

Merker S, Yustian I and Muhlenberg M (2004) Losing Ground but Still Doing Well-Tarsius dianae in Human-Altered Rainforests of Central Sulawesi, Indonesia. In: *Land Use, Nature Conservation and the Stability of Rainforest margins in Southeast Asia* (eds. Gerold, G, Fremerey, M and Guhardja, E), Springer, 299-311.

Mitchell J and Jones, PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatology, 25,* 693-712, Doi: 10.1002/joc.1181.

Nepstad D, Soares-Filho B, Merry F et al. (2009) The end of deforestation in the Brazilian Amazon. *Science, 326,* 1350–51.

Pacala SW, Hurtt GC, Moorcroft PR, Caspersen JP (2001) Carbon storage in the US caused by land use change. In *The Present and Future of Modeling Global Environmental Change.* Terra Scientific Publishing. Toyko, Japan, pp. 145-172.

Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, Phillips OL, Shvidenko A, Lewis SL, Canadell JG, Ciais P, Jackson RB, Pacala SW, McGuire AD, Piao S, Rautiainen A, Sitch S and Hayes D (2011) A large and persistent carbon sink in the world's forests, *Science*, 333, 988-993.

Phillips OL, Lewis SL, Baker TR, Chao K-J and Higuchi N (2008) The changing Amazon forest. *Philosophical Transactions of the Royal Society, Series B,* 363, 1819-1828.

Piao, S., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luyssaert, S., ... & Vesala, T. (2008). Net carbon dioxide losses of northern ecosystems in response to autumn warming. *Nature, 451*(7174), 49-52.

Piao S, Ciais P, Friedlingstein P, de Noblet-Ducoudre´ N, Cadule P, Viovy N, and Wang T (2009) Spatiotemporal patterns of terrestrial carbon cycle during the 20th century. *Global Biogeochem. Cycles, 23,* GB4026, doi:10.1029/2008GB003339.

Piao, S., Ciais, P., Lomas, M., Beer, C., Liu, H., Fang, J., ... & Woodward, I. (2011). Contribution of climate change and rising CO2 to terrestrial carbon balance in East Asia: A multi-model analysis. *Global and Planetary Change*., 75, 133-142.

Pitman A, de Noblet-Ducoudré N, Cruz F, Davin E, Bonan G, Brovkin V, et al. (2009) Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophysical Research Letters, 36*(14), L14814.

Pongratz, J., Reick, C., Raddatz, T., & Claussen, M. (2008). A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochemical Cycles, 22*(3), GB3018.

Pongratz J, Reick CH, Raddatz T, and Claussen M (2009) Effects of anthropogenic land cover change the carbon cycle of the last millennium. *Global Biogeochem. Cy.*, *23*, GB4001, doi:10.1029/2009GB003488.

Pongratz J, Reick CH, Raddatz T, and Claussen M (2010) Biogeophysical versus biogeochemical climate response to historical anthropogenic land cover change. *Geophysical Research Letters*, 37, L08702, doi:10.1029/2010GL043010.

Ramankutty, N, and Foley, JA (1999) Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochem.Cycles*, *13*(4), 997-1027.

Ramankutty N, Gibbs HK, F Achard, DeFries R, Foley JA, and Houghton RA (2007) Challenges to estimating carbon emissions from tropical deforestation. *Glob. Change Biol.*, *13*, 51-66; doi: 10.1111/j.1365-2486.2006.01272.x.

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

Raupach M (2011) Carbon cycle: Pinning down the land carbon sink. *Nature Climate Change, 1*, 148-149, doi: 10.1038/nclimate1123.

Reay DS, Dentener F, Smith P, Grace J, and Feely RA (2008) Global nitrogen deposition and carbon sinks. *Nature Geoscience*, *1*, 430-437.

Reick C, Raddatz T, Pongratz J, and Claussen M (2010) Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO_2. *Tellus B*, doi:10.1111/j.1600-0889.2010.00479.x.

Ricciuto DM, Davis KJ, and Keller K (2008) A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty. Global *Biogeochem. Cycles*, *22*, GB2030, doi:10.1029/2006GB002908.

Shevliakova E, Pacala SW, Malyshev S et al. (2009) Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. *Global Biogeochem.Cycles, 23*, GB2022, doi: 10.1029/2007GB003176.

Sitch, S., V. Brovkin, W. von Bloh, D. van Vuuren, B. Eickhout, and A. Ganopolski (2005), Impacts of future land cover changes on atmospheric CO2 and climate, Global Biogeochem. Cycles, 19, GB2013, doi:10.1029/2004GB002311.

Stocker BD, K Strassmann, and Joos F (2011) Sensitivity of Holocene atmospheric CO_2 and the modern carbon budget to early human land use: analyses with a process-based model. *Biogeosciences, 8*, 69–88, doi:10.5194/bg-8-69-2011.

Strassmann KM, Joos F and Fischer G (2008) Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO_2 increases and future commitments due to losses of terrestrial sink capacity. *Tellus B, 60*, 583–603.

Tian H, Melillo J, Lu C et al. (2011) China's terrestrial carbon balance: Contributions from multiple global change factors, *Global Biogeochem. Cycles, 25*, GB1007, doi:10.1029/2010GB003838.

van der Werf GR, Randerson JT, Giglio L et al. (2010) Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos Chem Phys, 10(23)*, 11707–11735.

Van Minnen JG., Klein Goldewijk K, Stehfest E, Eickhout B van Drecht G and Leemans R (2009) The importance of three centuries of land-use change for the global and regional terrestrial carbon cycle. *Climatic Change, 97*, 123–144.

Vitousek, P. M., & Howarth, R. W. (1991). Nitrogen limitation on land and in the sea: how can it occur?. *Biogeochemistry*, *13*(2), 87-115.

Verburg PH, Newmann K, and Nol L (2011) Challenges in using land use and land cover data for global change studies *Global Change Biology*, *17*, 974-989.

Yang X, Wittig V, Jain A, and Post W (2009) Integration of Nitrogen Dynamics into a Global Terrestrial Ecosystem Model *Global Biogeochem. Cycles, 23*, GB4028, doi:10.1029/2009GB003519.

Yang X, Richardson TK, and Jain AK (2010) Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake *Biogeosciences*, *7*, 3041–3050.

Ye Y, Fang XQ (2011) Spatial pattern of land cover changes across Northeast China over the past 300 years. *J. Hist Geogr, 37(4)*, 408–417.

Zaehle, S., Ciais, P., Friend, A. D., & Prieur, V. (2011). Carbon benefits of anthropogenic reactive nitrogen offset by nitrous oxide emissions. *Nature Geoscience, 4*(9), 601-605.

CHAPTER 4

Increased influence of nitrogen limitation on CO_2 emissions from future land use and land-use change

4.1 Abstract

In the latest projections of future greenhouse gas emissions for the Intergovernmental Panel on Climate Change (IPCC), few Earth System Models included the effect of nitrogen limitation, a key process limiting forest regrowth. Few included forest management (wood harvest). We estimate the impacts of nitrogen limitation on the CO_2 emissions from land use and land-use change (LULUC), including wood harvest, for the period 1900-2100. We use a land-surface model that includes a fully coupled carbon and nitrogen cycle, and accounts for forest regrowth processes following agricultural abandonment and wood harvest. Future projections are based on the four Representation Concentration Pathways used in the IPCC Fifth Assessment Report, and we account for uncertainty in future climate for each scenario based on ensembles of climate model outputs. Results show that excluding nitrogen limitation will underestimate global LULUC emissions by 34-52 PgC (20-30%) during the 20th century (range across three different historical LULUC reconstructions) and by 128-187 PgC (90-150%) during the 21st century (range across the four IPCC scenarios). The full range for estimated LULUC emissions during the 21st century including climate model uncertainty is 91 to 227 PgC (with nitrogen limitation included). The underestimation increases with time because: (1) Projected annual wood harvest rates from forests summed over the 21st century are 380-1080% higher compared to those of the 20th century, resulting in more regrowing secondary forests, (2) Nitrogen limitation reduces the CO_2 fertilization effect on net primary production of regrowing secondary forests following wood harvest and agricultural abandonment, and (3) Nitrogen limitation effect is aggravated by the gradual loss of soil nitrogen from LULUC disturbance. Our study implies that: (1) Nitrogen limitation of CO_2 uptake is substantial and sensitive to nitrogen inputs, (2) If LULUC emissions are larger than previously estimated in studies without nitrogen limitation, then meeting the same climate mitigation target would require an equivalent additional reduction of fossil fuel emissions, (3) The effectiveness of land-based mitigation strategies will critically depend on the interactions between nutrient limitations and secondary forests resulting from LULUC, and (4) It

is important for terrestrial biosphere models to consider nitrogen constraint in estimates of the strength of future land carbon uptake.

4.2 Introduction

The term "land-use change" typically refers to conversion of one land cover type to another, such as clearing forest to grow crops. In contrast, "land use" refers to management without changing the land cover, such as wood harvest and agricultural management (e.g. cropping practices, irrigation). CO_2 emissions from land use and land-use change (LULUC) represents the 'net effect' of CO_2 sources (emissions from deforestation, logging and other direct human activities), and CO_2 sinks (as vegetation regrows following land disturbance). LULUC emissions are estimated as 0.9±0.5 PgC/yr (1 PgC = 10^{15} gC) to the atmosphere, for the decade 2004-2013 (Le Quéré et al., 2015). On a relative scale, LULUC emissions are ~10% of the CO_2 emissions from fossil fuel combustion and cement manufacture (8.9±0.4 PgC/yr), for the same decade (Boden et al., 2013).

To balance the carbon budget, total anthropogenic CO_2 emissions (fossil fuels + LULUC) should equal the sum of CO_2 accumulated in the atmosphere, the oceanic sink, and the remaining CO_2 exchanged between the atmosphere and terrestrial biosphere. We have a good quantitative understanding and constrained estimates of fossil fuel emissions, atmospheric growth rates, and the oceanic sink. From these better-constrained fluxes, and modeled estimates of LULUC emissions, the remaining terrestrial biosphere flux is inferred as a residual sink of 2.9±0.8 PgC/yr for the decade 2004-2013, thereby offsetting roughly one-quarter of the total anthropogenic carbon emissions (Le Quéré et al., 2015). The uncertainty in estimating the residual terrestrial sink is mainly attributable to uncertainties in estimating LULUC emissions (Ballantyne et al., 2015; Houghton et al., 2012). The residual terrestrial sink indicates an increased net carbon accumulation by the terrestrial ecosystems which are sensitive to changing environmental controls (e.g. climate, CO_2 fertilization, nitrogen deposition) (Ballantyne et al., 2015; Ciais et al., 2013; Le Quéré et al., 2015; Schimel et al., 2015; Shevliakova et al., 2013). Thus, the terrestrial biosphere provides a subsidy to human activities by net absorption of atmospheric CO_2, slowing down the rate of climate change significantly. An understanding of how this subsidy may change in the future, in response to changing environmental controls, is essential to understanding the magnitude of the climate change problem. Constraining the future residual terrestrial sink hinges

on our estimates of sources and sinks from LULUC with narrow enough uncertainty bounds. The uncertainties arise not only due to the range of possibilities on how the future world might evolve with respect to LULUC and its environmental controls, but also in our understanding of various processes that affect the LULUC fluxes.

In a recent article, we studied the role of LULUC emissions on the carbon budget for the period 1765-2010. The study used a terrestrial ecosystem component of a land surface model, the Integrated Science Assessment Model (ISAM) that includes a fully coupled carbon-nitrogen cycle and detailed representation of secondary forest dynamics to account for forest regrowth processes following agricultural abandonment and wood harvest. We showed that failing to account for nitrogen dynamics, a key process limiting forest regrowth, underestimated LULUC emissions by ~70% in the non-tropics, ~10% in the tropics, and ~40% globally during 1990s compared to simulations that included the nitrogen dynamics. The study conveyed two key messages: (1) nitrogen limitation will significantly reduce the effect of carbon sinks on regrowing secondary forests (see Pongratz, 2013), and (2) historically, more secondary forests have resulted from wood harvest than from agricultural abandonment, underscoring the importance of forest management in estimating LULUC emissions (also see Yang et al. (2010); Ciais et al. (2013)).

The 21st century scenarios based on the Coupled Model Intercomparison Project phase 5 (CMIP5) project a 380-1080% rise in global forest wood harvest rates (area harvested each year summed over the century) compared to those of the 20th century (Table 4.1), due to rapid increase in demand for bioenergy and wood products (Hurtt et al., 2011). Therefore, the effect of nitrogen limitation on the rates of carbon sink on regrowing secondary forests could be much greater in the future compared to the historical period, having significant implications for the effectiveness of land-based mitigation policies. Accounting for both nitrogen limitation and forest management are beyond the current capabilities of many global climate models involved in CMIP5 (Ciais et al., 2013), thus giving them the tendency to be too optimistic in simulating future carbon sinks (Walker et al., 2015; Wårlind et al., 2014; Wieder et al., 2015a). Accordingly, the overall aim of this study is to understand how future LULUC emissions are influenced by the interactions among LULUC, nitrogen limitation, and anthropogenic environmental changes (CO_2 fertilization, climate change, and nitrogen deposition that reduce

the nitrogen limitation effect). We place specific emphasis on land management. The overall aim can be split into three parts.

First, we study the magnitude of LULUC emissions (with nitrogen limitation effect) attributable to "land use" (management) as compared to "land-use change", and how the magnitude is influenced by anthropogenic environmental changes. To quantify this effect, first we study the relative contribution of the direct effects of human LULUC activities versus the indirect effects of anthropogenic activity via environmental changes (climate, CO_2, and nitrogen deposition) to total LULUC emissions (see methods; for significance, see Houghton (2013a)). We then breakdown these contributions into its two component activities: "land use" and "land-use change". The land use activities considered include wood harvested from forests and subsequent regrowth processes. The land-use change activities include clearing of natural ecosystems for expansion of cropland and pastureland and forest regrowth following agricultural abandonment.

Second, we study the impact of nitrogen limitation on the 21st century LULUC fluxes. For this purpose, all LULUC fluxes estimated for the above objective are also simulated without the effect of nitrogen limitation. We quantify the impacts of nitrogen limitation by comparing the LULUC fluxes simulated between with and without nitrogen limitation case. We assess the future impacts of nitrogen limitation on LULUC fluxes relative to that of 20th century.

Third, we carry out a comprehensive assessment of the uncertainties in estimates of future LULUC emissions due to the different mitigation scenarios of the IPCC Fifth Assessment Report (IPCC, 2013), and uncertainties in climate projections underlying each scenario. This is important given that significant uncertainties in simulating future terrestrial carbon fluxes among Earth System Models are attributable to differences in simulated climate (Ciais et al., 2013). The uncertainties in projected climate (see Figure S1, S2) reflect uncertainties in emissions scenarios, model initializations, and gaps in process understanding (Knutti and Sedláček, 2012). Using one terrestrial ecosystem model, but consistently driven by different climate model projections enables us to study how much of total uncertainty in future LULUC emissions are attributable to differences in climate projections alone. Our future climate uncertainty analysis builds on historical uncertainties in quantifying the spatial and temporal patterns of historical LULUC (methods).

4.3 Materials and methods

We use a data-modeling approach to study the three objectives discussed above. This section briefly describes: the land surface model used to simulate LULUC fluxes, model forcing data, and model simulations performed.

Model details

We use a terrestrial ecosystem component of a land surface model, Integrated Science Assessment Model (ISAM) to assess the impacts of LULUC on terrestrial carbon fluxes. The terrestrial component of ISAM simulates carbon and nitrogen fluxes between the vegetation and the atmosphere (net land-to-atmosphere flux), above and below ground litter, and soil organic matter at 0.5^{o}x0.5^{o} spatial resolution (Jain and Yang, 2005). ISAM includes detailed representation of nitrogen dynamics (Yang et al., 2009) and secondary forest dynamics (Yang et al., 2010). The carbon cycle feedbacks modeled includes the influence of: (1) increasing atmospheric (CO_2) on Net Primary Productivity (NPP), (2) temperature and precipitation changes on photosynthesis, autotrophic and heterotrophic respiration, and (3) nitrogen deposition on carbon uptake by plants. The modeled nitrogen cycle accounts for major processes such as denitrification, mineralization, immobilization, nitrification, leaching, symbiotic, and non-symbiotic biological nitrogen fixation. Our water cycle is based on the LINKAGES model (Hanson et al., 2004; Pastor and Post, 1985). The model operates at two time steps. The vegetation carbon including NPP, litter production, and nitrogen demand by plants are calculated annually. Decomposition of soil and litter, and nitrogen cycle are calculated weekly. The structure, parameterization, and evaluation of nitrogen cycle in ISAM are detailed in Yang et al. (2009). Jain et al. (2009) show that the model can simulate the response on historical terrestrial carbon fluxes due to nitrogen limitation, LULUC, and changes in (CO_2), climate change, and nitrogen deposition. ISAM and its extended versions have continually been evaluated and improved using both field observations and model inter-comparison activities (El-Masri et al., 2013; Huntzinger et al., 2012; Ito et al., 2008; Tian et al., 2015; Walker et al., 2014). Results from ISAM have been a part of the global carbon budget (Le Quéré et al., 2015), and several IPCC Assessment Reports, including the most recent Fifth Assessment Report (Ciais et al., 2013).

Each 0.5^{o}x0.5^{o} lat/lon grid cell in ISAM is occupied by one or more of the 18 land cover types (Yang et al., 2010) that include 5 types of primary forests classified by ecozone, and their corresponding "secondary forests", 5 types of non-forested vegetation (e.g. grassland, savanna, shrubland), bareland, cropland, and pastureland. The model separately accounts for forest regrowth following agricultural abandonment and wood harvest, here termed as "secondary forest".

Net land-to-atmosphere carbon flux calculation in ISAM

Here, we provide an overview of the calculation of the Net Ecosystem Exchange (NEE), the carbon exchanged between the ecosystem and the atmosphere. In the "simulations performed" section, we describe how NEE calculated with different experimental setup is combined to estimate LULUC fluxes. Our terminology of carbon fluxes follows Chapin et al. (2006).

Following LULUC, the vegetation biomass is released as carbon to the atmosphere as three components in the ISAM.

(1) Loss of soil organic carbon due to oxidation of organic matter when native soils are cleared for agriculture. We assume 25% of soil organic carbon stored in the top meter of the soil is lost to the atmosphere upon clearing (E_s), with most loss occurring within the first year of clearing soils. The 25% loss is the average estimate across observational studies (Table 3 of Houghton and Goodale, 2004), and consistent with the assumption of Houghton's bookkeeping model (Houghton, 2010). We also test the sensitivity of our results to this model parameter.

(2) Part of biomass is shed as litter that enters the soils and decays on-site. As a result of decomposition, there is a heterotrophic respiration (HR) that includes losses by herbivory and the decomposition of organic debris by soil biota.

(3) Part of vegetation biomass enters the wood and fuel product pools and decays at rates dependent on the product pool type following McGuire et al. (2001). E_p is the emissions from product pools that collectively represents 1 year (agriculture and agriculture products), 10 year (paper and paper products), and 100 year (lumber and long-lived products) product pools. The fraction of vegetation biomass that goes into the three product pools depends on the LULUC activity and region, following Houghton and Hackler (2001). The three decay pools represent the

woody material removed from the site. In reality, the harvest from agriculture and forestry may be transported to locations far off from the harvested grid cell and allowed to decay. However, due to lack of such global datasets, we assume the product pool decays at the grid cells of origin.

We calculate the NEE carbon flux following LULUC as

$$NEE = HR - NPP + E_p + E_s \quad \text{Eq. (1)}$$

In Eq. (1), positive values for NEE represent flux to the atmosphere. NPP is the carbon accumulated in vegetation (carbon fixed through photosynthesis minus autotrophic respiration). In the case of LULUC, NPP accounts for the carbon accumulated from forest regrowth following agricultural abandonment and wood harvest. The model down regulates NPP depending on the magnitude of simulated nitrogen limitation (nitrogen demand minus supply) (see Yang et al. (2009) for equations). All the right hand side terms of Eq. (1) are influenced by both LULUC, natural (nitrogen limitation) and anthropogenic environmental changes (CO_2 fertilization, climate change, and nitrogen deposition that partly offsets nitrogen limitation effects).

To highlight, NEE following LULUC (Eq. 1) includes three components: emissions following disturbance, legacy fluxes (delayed carbon fluxes from soil and product pool decays), and carbon fluxes induced by anthropogenic environmental changes. Legacy fluxes include both source (decay) and sink terms (regrowth of secondary forests following agricultural abandonment and wood harvest in previous years) and they cause an imbalance between NPP and HR (Pongratz et al., 2014).

Model forcing data

Overview

The basis of our study is driving data and climate model output from the Coupled Model Intercomparison Project Phase 5 (CMIP5) for future scenarios of land-use change and fossil fuel emissions. The CMIP5 is coordinated by the World Climate Research Programme in support of IPCC Fifth Assessment Report (Taylor et al., 2012). The IPCC CMIP5 analysis features four Representation Concentration Pathways (RCPs) for the future (>2005 AD) derived from Integrated Assessment Models (IAMs); each describing a possible pathway of future greenhouse gas concentration depending on human behavior including energy use, land use and mitigation

policy. No single pathway is more likely than another. The products for each RCP include gridded estimates of LULUC, and atmospheric emissions and concentrations of GHGs for the future that are harmonized to transition smoothly from historical estimates/observations. Coordinated experiments, carried out by more than 20 modeling group from around the world used these data products for conducting a range of climate modeling experiments that included projecting future climate change. The RCPs have been extensively described in literature (Moss et al., 2010; van Vuuren et al., 2011). We provide a summary of the RCPs with emphasis on its LULUC characteristics in supplementary text S1.

Atmospheric forcing data

ISAM requires forcing data on climate, atmospheric (CO_2), and nitrogen deposition. Data for atmospheric (CO_2) is as per CMIP5 experiments (Meinshausen et al., 2011) (see Figure S3 for how the atmospheric (CO_2) varies with time for the RCPs). Gridded estimates of airborne nitrogen deposition are from Lamarque et al. (2011) (see supplementary text S2 and Figure S4 for further details). We account for changes in two climate variables: temperature and precipitation. We do not explicitly simulate the effects of radiation on carbon fluxes. Climate data for the historical period is from CRU TS3.21 (Harris et al., 2014). To account for the climate uncertainties for the RCPs (2006-2100), we use monthly climate projections from a suite of 43 climate models from the CMIP5 multi-model ensemble database (Table S1). All climate data are at 0.5°x0.5° lat/lon and interpolated to weekly time steps within the model. Additional details on climate data processing are available in supplementary text S3.

LULUC data

We prescribed LULUC data from the land-use harmonization (LUH) database used for CMIP5 (Hurtt et al. (2011); http://luh.umd.edu/). The data covers the period 1500-2100 annually and at 0.5°x0.5° lat/lon resolution. The historical LULUC in the Hurtt data is based on the HYDE 3.1 reconstruction for cropland and pastureland transitions (Klein Goldewijk et al., 2011), and wood harvest from Food and Agriculture Organization (FAO). Specifically, we include three types of wood harvest from LUH database that are provided as fractional area of each grid cell: wood harvest from primary forested land (variable code in LUH: gfsh3), mature secondary forested land (gfsh1), and young secondary forested land (gfsh2). Hurtt et al. (2011) estimated

wood harvest area by combining two other estimates: (1) biomass extracted from wood harvest, and (2) model-based estimates of historical above-ground carbon stocks and forest extent.

The future aggregate (for larger world regions) land demands (cropland, pastureland, and wood harvest) in the Hurtt data are based on the four RCPs derived from IAMs. The "harmonization" in the Hurtt data downscales the aggregate regional land demands to individual grid cells within the region, while simultaneously ensuring that the downscaled maps are spatially consistent with the historical reconstruction.

We use a map of potential natural vegetation and a rule-based approach (Meiyappan and Jain, 2012) to transform the prescribed LULUC information into estimates of annual land cover areas (and underlying land conversions) for each grid cell, consistent with the land cover types of ISAM, similar to the approach taken in other land models (Lawrence et al., 2012; Pitman et al., 2009). The rules are specific to each LULUC activity, and are broadly consistent with our understanding of historical LULUC dynamics. The rules are generalizations of regional case studies of LULUC drivers, and have been calibrated using remote-sensing observations. Supplementary text S4 gives further details on the LULUC implementation in the model. The LULUC characteristics for the study period (1900-2100), historically and for each future RCP, are shown in Figure 4.1 and Table 4.1, and summarized in the supplementary text S1.

Simulations Performed

We initialized ISAM with an atmospheric (CO_2) of 278 ppmv, representative of approximate conditions in the starting year (1765 AD – pre-industrial conditions) of the model simulation, to allow vegetation and soil carbon pools to reach an initial steady state.

The basic approach to calculate LULUC emissions (E_{LULUC} in Eq. 2) is by comparing the NEE (Eq. 1) calculated between two simulations, one with LULUC (NEE_LULUC in Eq. 3) and the other without LULUC (NEE_noLULUC in Eq. 4).

$$E_{LULUC} = NEE\_LULUC - NEE\_noLULUC \quad \text{Eq. (2)}$$

where,

$$NEE\_LULUC = HR\_LULUC - NPP\_LULUC + E_p\ \_LULUC + E_s\ \_LULUC \quad \text{Eq. (3)}$$

$$NEE\_noLULUC = HR\_noLULUC - NPP\_noLULUC \quad \text{Eq. (4)}$$

In Eq. (4), NEE_noLULUC represents the effects of environmental changes on potential natural vegetation. The E_p and E_s terms do not appear in Eq. (4), because they are zero when there is no LULUC. The LULUC emissions (Eq. 2) are influenced directly by humans through LULUC (Eq. (3); hereafter referred as "direct emissions" from LULUC), and indirectly by human-induced environmental changes on lands undergoing LULUC through Eq. (3, 4) ("indirect emissions" from LULUC). In this paper, the total of direct and indirect emissions is referred to as "total emissions" from LULUC. The definition has been widely adopted for over a decade (McGuire et al., 2001; Pongratz et al., 2014). All the three emissions (i.e. direct, indirect, and total emissions) are net fluxes and they include both source and sink terms.

We carried out a series of with and without LULUC simulations (Table 4.2) for the time period 1765-2100. Table 4.3 summarizes how the results from the simulations listed in Table 4.2 were combined to estimate Eq. (2) that represents different LULUC flux components, and is further explained in text S5.

In principle, the effects of nitrogen limitation on terrestrial ecosystems are a natural response of the system to human-induced environmental change and hence can be counted as indirect effects of human activity along with climate and (CO_2) change. However, in our simulations, we kept nitrogen limitation separate from other environmental effects because our study aims to understand the interactive effects of including nitrogen cycle on LULUC fluxes.

Accounting for uncertainties in future climate projections and LULUC reconstructions

We carried out the simulations (Table 4.2) and associated calculations (Table 4.3) separately for each RCP using corresponding forcing data for LULUC and environmental drivers. Specifically, to account for uncertainties in climate projections within each RCP, we repeated simulations Ref_1, Ref_2, A1, A2, B1 and B2 by varying the climate data listed in Table S1, but keeping the data for other drivers (CO_2, nitrogen deposition, and LULUC) same across the simulations. We carried out simulations C1, C2, D1 and D2 once for each RCP, as they are independent of climate change. The climate-induced uncertainty in simulating total LULUC emissions is purely from indirect effects of human activity on emissions mediated through environmental change, because by design, direct effects of human LULUC activities on

emissions are independent of environmental changes. The simulations cover the period 1765-2100, and we present results for 1900-2100.

There are significant uncertainties in quantifying historical LULUC, resulting from differences in inventory datasets (Meiyappan and Jain, 2012) and reconstruction methodologies (Klein Goldewijk and Verburg, 2013). The HYDE reconstruction for cropland and pastureland used in CMIP5 is just one realization of what could have happened in the past. In an earlier study (Jain et al., 2013) we forced ISAM with three LULUC reconstructions to estimate an array (uncertainty range) of "total LULUC emissions" for the 20th century (including cropland and pastureland transitions from HYDE, Ramankutty (2012), and FAO (2006); all using common data for wood harvest based on Hurtt et al. (2011)). Here, we used those estimates for comparison with our future estimates. New to this study is separately calculating direct and indirect LULUC emissions and separating emissions by LULUC activity (i.e. land-use change and wood harvest) for the three historical LULUC reconstructions (from simulations analogous to Tables 4.2 & 4.3).

4.4 Results

Historical simulations: Overview

We first present comparison of two key modeled estimates from our historical simulations with observationally derived global estimates: (1) model simulated above-ground vegetation (tree foliage + woody biomass) carbon in forests vs. FAO-based gridded estimates, and (2) our model simulated NPP vs. NPP modeled from MODIS derived radiation absorption by plants. These comparisons are broadly intended to evaluate how well the historical simulations can reproduce the current conditions. While evaluating the model performance over the historical period is no guarantee of good performance over the 21st century, it does add confidence in the model's suitability for assessing impacts of the interactions between LULUC and environmental change on terrestrial carbon fluxes. While comparison of two model simulated variables is not indicative of overall model performance, the two variables compared here are critical to modeling LULUC emissions. For example, our modeled emissions from wood harvest depend on how well we simulate above-ground vegetation carbon in forests. Similarly, as our NPP is regulated by modeled nitrogen demand and supply specific to land cover type (Yang

et al., 2009), an overall agreement in NPP compared to an independent estimate adds confidence in our modeled nitrogen cycle, and its applicability to scientific questions addressed in this paper. Following this comparison, we present LULUC emission estimates for the 20th century. We highlight model uncertainty in the discussion section.

Model evaluation

Globally, our model simulated above-ground vegetation carbon in forests of 268 PgC (year 2000) compares to 234 PgC estimated from FAO-based gridded statistics (Kindermann et al., 2008; note the study does not provide uncertainty estimates). A zonal (Figure 4.2) and spatial comparison (Figure S5) indicates that our simulated above-ground carbon in forests is higher in tropics and northern non-tropics, and lower in southern non-tropics. The reasons underlying the systematic latitudinal bias between the two estimates could stem from both data and model sources (e.g. bias from methods used to fill missing country data in FAO and gridding procedure; bias in our model forcing data and errors in model parameter and structure), and includes differences in definition of forest (FAO forest definition of percent cover >10% and height > 5m (Annex 2 of FAO (2005, 2010)) vs. our model definition of percent cover >60% and height >2m based on the IGBP land classification scheme (Loveland and Belward, 1997)).

Next, we compared our model simulated NPP across six land cover types averaged globally over a 5 year period (2001-2005) with that from modeled NPP from MODIS derived radiation absorption by plants (Zhao and Running, 2010; Zhao et al., 2005; Running et al., 2004). Note that we are comparing two modeled estimates with inherent errors and uncertainties (Cleveland et al., 2015). Nonetheless, results show that our model simulated NPP across all land cover types fall within the standard deviation range from radiation-based estimates (Figure 4.3).

Carbon fluxes from LULUC during the 20th century (with nitrogen limitation effect)

Globally, the total LULUC emissions averaged across the three LULUC reconstructions were 163 PgC (range: 156-174 PgC) cumulated over the 20th century (Table 4.4) (all numbers discussed include nitrogen limitation effect, unless explicitly noted). The total LULUC emissions are about 58-65% of fossil fuel emissions, and 37-40% of total carbon emissions over the 20th century (266 PgC from fossil fuel combustion and cement production - Boden et al. (2013)). Most of the historical total LULUC emissions were direct emissions (Table 4.4; Figures 4.4a,

4.5a). The indirect emissions averaged across the three reconstructions were close to zero (-22 to 21 PgC), because of partly offsetting environmental effects. For example, enhanced carbon sinks in regrowing forests under increasing (CO_2), also leads to higher emissions when harvested.

Regionally, the non-tropics accounted for about two-thirds (52-71%) of cumulative 20th century total LULUC emissions (Table 4.4; Figure 4.4). The total LULUC emissions from non-tropics are greater than the tropics mainly because: (1) nitrogen limitation in the non-tropics reduced the carbon uptake rates on regrowing secondary forests (Figures 4.4k, 4.5g; compare with and without nitrogen limitation cases), and (2) historically, two-thirds of global secondary forest area following wood harvest is from the non-tropics (Table 4.1).

Splitting the total LULUC emissions based on LULUC activity, 60-65% of the total global LULUC emissions are from land-use change, and the rest 35-40% is from wood harvest (55-72 PgC) (Figure 4.4a). Breaking down regionally, wood harvest accounted for 11-22% of total LULUC emissions in the tropics (Figure 4.4f), and 50-57% in the non-tropics (Figure 4.4k).

Future simulations: Overview

First, using RCP8.5 as example, we describe how the key mechanisms in our model impact nitrogen limitation over time (the mechanisms qualitatively apply to all RCPs). Next, we present the overall effects of these mechanisms on the 21st century LULUC fluxes compared to that of 20th century. For this purpose, we use the mean emissions value of the three LULUC reconstructions for the 20th century. Third, we isolate the effect of nitrogen cycle on LULUC emissions by comparing results between with and without nitrogen cases. Finally, we quantify the uncertainties in LULUC emissions resulting from uncertainty in projections of climate change.

Model response to nitrogen limitation

Nitrogen limitation exists if there is not enough mineral nitrogen available for plant growth and litter decomposition. The difference between nitrogen demand and supply is the magnitude of nitrogen limitation. Results (Figure 4.6) show that the total carbon uptake (NPP) by secondary forests increases with time (in both with and without nitrogen cases) due to both CO_2 fertilization effect on regrowing forests, and increase in the area of regrowing forests (following

wood harvest and agricultural abandonment). However, the carbon uptake in secondary forests is significantly lower when nitrogen limitation is included especially in the non-tropics. The tropics have relatively less nitrogen limitation, because warmer and wetter climate enhances nitrogen mineralization in soils. Furthermore, the difference in carbon uptake rates between with and without nitrogen limitation simulations increases over time reflecting the progressively increasing nitrogen limitation effect on CO_2 fertilization (note that the area of secondary forests is the same in both with and without nitrogen simulations). Our modeled response is consistent with ground-based studies that generally indicate that younger regrowing secondary forests require more nitrogen to support new production under increasing (CO_2) (Davidson et al., 2004; Finzi et al., 2006, 2007; Herbert et al., 2003; Hungate et al., 2003; Lebauer and Treseder, 2008; Luo et al., 2004, 2006; Oren et al., 2001; Murty et al., 2002; Reich et al., 2006).

Next, we describe how key nitrogen variables in our model vary with increasing nitrogen limitation. First, biological nitrogen fixation in both primary and secondary forests increases over time, with the increase being greater in secondary forests (Figure 4.7a, b). We model nitrogen fixation as a function of evapotranspiration (see discussion section for limitations of this approach). Therefore, tropical forests fix more nitrogen than non-tropical forests in our model, consistent with spatial observations (Cleveland et al., 1999). Second, increasing nitrogen demands from CO_2 fertilization causes both primary and secondary forests to uptake more nitrogen per unit area with time (Figure 4.7e, f), thus reducing ecosystem nitrogen losses (Figure 4.6). Third, with increasing nitrogen limitation, the nitrogen-use efficiency (NPP per unit nitrogen uptake; qualitatively similar to Carbon: Nitrogen ratio of vegetation) increases with time, especially in the secondary forests of the nitrogen limited non-tropics (Figure 4.6b-h).

Increasing anthropogenic nitrogen deposition (external forcing to our model) provides an additional nitrogen input to terrestrial ecosystems (Figure 4.7i). However, its effect on enhancing regrowth sinks (or reducing LULUC emissions) depends on how much of the nitrogen deposition occurs on regrowing forests. There are three major sources of nitrogen losses attributable to LULUC (inferred by comparing "with" and "without" nitrogen limitation simulations). (1) Anaerobic respiration by denitrifying bacteria (soil decomposition) increases with time due to increases in litterfall (leaf litter + dead wood) from LULUC (Figure 4.7j). (2) Leaching as soil nitrate dissolves in rainwater and excess water percolates through soil (Figure 4.7k). Leaching is

higher in the tropics because of more rainfall, and relatively more soil nitrogen compared to the non-tropics. Both denitrification and leaching depends on our simulated soil nitrate content and soil moisture. (3) Removal of nitrogen from soils and vegetation from LULUC disturbance including slash burning and decay from product pools (Figure 4.7l) as also documented in earlier studies (Davidson et al., 2007; Herbert et al., 2003; Mathers et al., 2006; Schipper et al., 2007).

In summary, our model simulations suggest that large areas of secondary forests will not respond to CO_2 fertilization as strongly as they would when adequate nitrogen was available to meet the plant demands. In the with-nitrogen cycle simulation, our model responds to increasing nitrogen limitation by increasing nitrogen fixation, reducing nitrogen losses, and increasing nitrogen-use efficiency. LULUC activities result in a gradual loss of nitrogen from the system, thus increasing the nitrogen limitation. In the following section, we explore the overall effects of these mechanisms on the simulated future LULUC emissions.

Carbon fluxes from LULUC during the 21st century (with nitrogen limitation effect)

Both globally and regionally, the total LULUC emissions estimated across the four RCPs are smaller or comparable to 20th century mean estimates (Table 4.4). Globally, the direct LULUC emissions due to human activity estimated across the RCPs are a smaller source to the atmosphere by 40-80% compared to 20th century estimates (Table 4.4; Figures 4.4, 4.5 and S6). In contrast, the indirect LULUC emissions due to human environmental change are a much larger source to the atmosphere for the RCPs (55 to 77 PgC from Table 4.4) compared to 20th century (-20 to 20 PgC), making the total LULUC emissions much larger than when considering direct emissions alone. In this section, we further explore direct LULUC emissions. Interactions between nitrogen limitation and other environmental factors explain indirect LULUC emissions. We discuss indirect emissions in the next section.

In general, across all the RCPs, the net deforestation rates are significantly lower compared to the 20th century (Table 4.1; Figure 4.1) resulting in smaller direct emissions (Figure 4.4). Further, large amount of cropland and pastureland expansion that occurred in the 20th century, are being abandoned in the future (RCP 4.5 and RCP6.0) due to land protection policies (Figure 4.1; Table 4.1). Specifically, forest expansion in RCP4.5 is due to carbon taxation policies that encourage protection of forests (text S1). The higher emissions in the 20th century

from land clearing are partly offset in the future under RCP4.5 and RCP6.0 because of carbon accumulation in forests regrowing on abandoned land. This resulted in negative direct emissions from land-use change (sinks) for RCP4.5 globally (Figure 4.4c), and for both RCP4.5 and RCP6.0 in the tropics (Figure 4.4h, i).

In contrast to land-use change, direct emissions from wood harvest are equal to or larger than the 20th century estimates across all the RCPs (excluding one outlier RCP6.0 elaborated in discussion section), especially in the non-tropics (Figure 4.4). This is because the RCPs project a 380-1080% global rise in wood harvest rates compared to 20th century due to rapid increase in demand for bioenergy and wood products (Table 4.1). The higher wood harvest results in more regrowing forests that become increasingly nitrogen limited due to the mechanisms explained in previous section (excluding (CO_2) down-regulation that is an indirect effect). Thus emissions become much greater compared to slower and smaller sinks in regrowth plus temporary sinks in product pools. As a result, the wood harvest contribution to direct LULUC emissions increase in the future, especially in the already nitrogen limited non-tropics (Table 4.4; Figure 4.4). The higher rates of wood harvest also result in higher direct (and total) LULUC emissions in the non-tropics than in the tropics during the 21st century (Figure 4.5; Table 4.4).

Interestingly, despite large net forest regrowth in the non-tropics under RCP4.5 (Table 4.1), its direct LULUC emissions (Table 4.4) are higher than or comparable to RCP2.6 and RCP8.5, both of which show a loss of forest area (Table 4.1). This is because the direct emissions from wood harvest are higher in RCP4.5 (Figure 4.4m) where afforestation provides additional forest biomass to meet the prescribed wood harvest demands.

For both the tropics and non-tropics, the uncertainties in estimating direct emissions for the 20th century based on the three LULUC reconstructions (range from Table 4.4) is ~50% greater than the scenario-based uncertainty for the 21st century (maximum difference across the four RCPs from Table 4.4), indicating that historical LULUC reconstructions are more uncertain regionally than the likely future LULUC outcomes.

Impacts of including nitrogen cycle on LULUC fluxes

All the aforementioned estimates include the effect of nitrogen limitation. To understand the impact of nitrogen cycle, we simulated LULUC fluxes without nitrogen limitation effect i.e.

by assuming sufficient nitrogen is available for plant growth and litter decomposition. We quantify the impacts of nitrogen limitation by comparing the LULUC fluxes estimated between with and without nitrogen limitation case.

There are two key results. First, inclusion of nitrogen limitation increases the total LULUC emissions by 128-187 PgC globally for the 21st century, roughly 3-5 times larger compared to the increase for 20th century (Table 4.4; Figure 4.5). This increase is predominantly attributable to two of the component fluxes (Figure 4.4): direct emissions from wood harvest in the non-tropics, and indirect emissions from LULUC in both the tropics and non-tropics.

As described before, when we consider nitrogen dynamics, most of the regrowing forests become increasingly nitrogen limited. This restricts the rate of regrowth after harvest resulting in larger total emissions from wood harvest under nitrogen limited conditions, particularly in the non-tropics. When nitrogen dynamics were not considered, direct emissions from wood harvest were smaller, and total LULCC emissions were a sink under two mitigation scenarios (RCP4.5 and RCP6.0; see text S6 for elaboration) globally, and for all scenarios in the non-tropics (Table 4.4).

Most of the difference in indirect emissions between with and without nitrogen cases can be explained by the interactions between the nitrogen cycle and carbon cycle impacts on areas undergoing land-use change (Figure 4.4b-e). Without nitrogen limitation, the higher emissions from deforestation in a CO_2-fertilized world (due to higher carbon stocks) are partly compensated by stronger regrowth sinks (from CO_2-fertilization) on forests regrowing on abandoned land. However, when we include nitrogen limitation, the sinks on forests regrowing on abandoned land are weakened due to CO_2 down-regulation effect, especially in the nitrogen limited non-tropics where the net indirect flux shifts from a sink to a source (Figure 4.4l-0). This effect is also reflected in Figure 4.5, where the difference in indirect emissions between with and without nitrogen limitation case increases with time for the non-tropics. Under increasing (CO_2), plants need more nitrogen to support production. The insufficient availability of nitrogen limits the CO_2 fertilization effect on plant growth in our model (Figure 4.6; also see Norby et al., 2010; Wieder et al., 2015a). In the tropics, inclusion of nitrogen shifts the indirect flux due to land-use change from a source to a slightly bigger source for RCP2.6, RCP4.5 and RCP6.0 (Figure 4.4g-i).

While the above mechanisms apply for land experiencing wood harvest, the indirect wood harvest emissions are small whether nitrogen limitation is considered or not, because weaker sinks from CO_2 fertilization also result in lower emissions in the subsequent harvest cycle (except when wood harvest expands to new areas). This is despite our modeling assumption that wood is preferentially harvested from primary forests or mature secondary forests across most regions (consistent with the assumption made in producing the wood harvest data - Hurtt et al. 2011). In our model, regrowing secondary forests requires roughly 30 (tropics) to 40 (non-tropics) years to attain 80% maturity, and about 90 (tropics) to 150 (non-tropics) years to attain full maturity (Figure 4.8; Text S7). The high wood harvest rates projected for the future (Figure 4.1; Table 4.1) result in harvesting young regrowing secondary forests (as primary or mature secondary forests reduce following LULUC) that have not accumulated sufficient biomass (especially in grid cells with high wood harvest rates; see Figure S7).

Another key result of our simulations is that the total LULUC emissions from the non-tropics are greater than the tropics for 1900-2100 when nitrogen limitation is considered (Figures 4.4, 5, Table 4.4). In the simulations without nitrogen limitation, the LULUC emissions for the tropics were greater than the non-tropics, after 1940s (Figure 4.5). This result is consistent with majority of modeling studies that only include the interactive effects of CO_2 and climate in their calculations of total LULUC emissions (Jain et al., 2013).

It is worth noting that there are other important interactions that determine indirect emissions in our modeled results (Figure 4.4). For example, converting forests to agriculture increases indirect emissions in the methodological set up of the model experiments, i.e. comparing a hypothetical no-LULUC case with a representative with-LULUC case. This is because the hypothetical forest that exists in the no-LULCC case has greater sinks from CO_2 fertilization than the sinks in non-forests in the with-LULCC case. This capacity for an additional sink is lost due to deforestation; its magnitude depends on both deforested area and the strength of CO_2 fertilization (Pongratz et al., 2009; Strassman et al., 2008). Indirect emissions are increasingly affected by climate change in the future, for example, a warmer climate projected for the future (Figures S1, S2) increases indirect emissions due to enhanced soil decomposition (R_h) and forest decline in some regions. Concurrently, higher decomposition also releases plant

usable mineral nitrogen from soils that enhances carbon uptake in regrowing forests during initial stages (McGuire et al., 2007).

Climate induced uncertainties in simulating carbon fluxes from LULUC (with nitrogen limitation effects)

A key source of uncertainty in projecting future LULUC emissions is that due to the indirect human-induced effects via climate change. Here we evaluate the combined uncertainty from two climate variables: temperature and precipitation.

The range of uncertainties in the 21st century cumulative total LULUC emissions, across all the four RCPs driven by different CMIP5 climate model projections are: 91-227 PgC (globally), 21-96 PgC (tropics), and 51-126 PgC (non-tropics) (Table 4.5). Globally, for all RCPs the estimated range of total LULUC emissions due to climate uncertainty are roughly 50% of the mean value (Figure 4.5, Table 4.5), and are larger than the scenario-based difference of 54 PgC (estimated from Table 4.4 as the maximum difference in the mean estimates of total LULUC emissions across the RCPs). In some cases, the climate induced uncertainties in indirect LULUC emissions (Table 4.5; max-min range) are larger than its mean estimates (Table 4.4), making even the sign of indirect emissions uncertain (Figure 4.5h, i). The uncertainty tends to increase with time in the higher emission scenarios (Figure 4.5), reflecting the progressively increasing model spread in CMIP5 projected climate (Knutti and Sedláček, 2012). RCP6.0 has the smallest uncertainty range across all RCPs, partly because only 24 climate model projections were available for RCP6.0 when we carried out the simulations (Table S1).

Most of the uncertainty results from including wood harvest, because its spatial extent is much larger compared to land-use change (Figure 4.1; Table 4.1). The uncertainties in simulated indirect (and total) LULUC emissions are greater over the non-tropics than the tropics (Table 4.5) mainly because of large uncertainties in projected temperature over the temperate zones of the northern hemisphere (Figures S1, S2) where most of the non-tropical wood harvest occurs (Figure 4.9, S7).

4.5 Discussion

Comparison with previous studies

Previous studies that have examined the future LULUC fluxes using CMIP5 data differ from the LULUC flux estimates presented here on multiple aspects: LULUC activities included (e.g. wood harvest) and their implementation in the model; model processes considered (e.g. nitrogen, secondary forest dynamics); and the type of model itself. A direct one-to-one comparison is confounded by these multiple source of differences. Therefore, our approach is to compile the available estimates and identify the causes of difference from our study.

Wood harvest

Hurtt et al. (2011) provides wood harvest, as biomass extracted from each grid cell. They also provide "wood harvest area" in each grid cell, estimated as the sum of primary, mature secondary, and young secondary forest area required to meet the biomass demand from wood harvest. Therefore, the biomass extracted and the wood harvest area was meant to be consistent with each other. In this study, we implemented "wood harvest area" data in ISAM ("LULUC data" in Section 4.3) to calculate biomass extracted (Table 4.6) and LULUC emissions. The 20th century biomass harvested from forest simulated by the ISAM compares well with Hurtt data, because the available forest area (and average forest biomass per grid area) was adequate to meet the historical demands. However, for all four RCPs, our model estimated forest biomass from wood harvest is much lower compared to Hurtt data, due to two reasons.

First, the forest harvest rates for the RCPs (especially RCP6.0 with highest wood harvest area; Table 4.1) were higher than the contemporary (circa 2005) forest area (and biomass as evaluated in Figure 4.2) in ISAM (Figure 4.9). Specifically, RCP 6.0 shows high wood harvest rates for the Himalayas and China (Figure S7) that seem inconsistent with the contemporary forest area estimated from satellites (Figure 4.9; see Figure S8 for MODIS derived land cover map). Therefore, the modeled forest area could not fully meet the prescribed wood harvest demands. The forest definition in ISAM is consistent with the IGBP land-classification scheme, and its contemporary forest estimates have been calibrated using the most recent version of MODIS land-cover data (Meiyappan and Jain, 2012). In the year 2005, the MODIS estimated global forest area (~30 million km^2) was 25% less than the ~40 million km^2 estimated by Hurtt et al. (2011).

Second, following wood harvest, regrowing forests require about 90 (non-tropics) to 150 (tropics) years to attain full maturity (Figure 4.8; text S7). Therefore, lower contemporary forest areas in ISAM compared to Hurtt data implies that we had to use more secondary regrowing forests with lower biomass to meet the future harvest area demands, when enough primary forests or mature secondary forests were not available in the grid cell. Hypothetically, even if our harvest area had equaled Hurtt data, a higher fraction of the total harvested area in our model would be from secondary regrowing forests with lower biomass compared to Hurtt estimates. Most of the discrepancy in forest area stems from the non-tropics (Table 3 of Meiyappan and Jain (2012)), where additionally, the biomass harvested from forests are also higher in Hurtt data than in our model (Table 4.6).

A part of the discrepancy in forest area between our study and Hurtt et al. is attributable to difference in the definition of forest. Hurtt et al. count savannas as forest (using a different tree density threshold to identify forests), but in ISAM, savannas are classified as herbaceous (Meiyappan and Jain, 2012). This difference in definition implies that we did not use savannas for wood harvest, and our estimate of deforestation (Table 4.1 and Figure 4.1) does not include savannas converted to cropland and pastureland. Multiple definitions for savannas exist (Scholes and Archer, 1997). In our model, even counting savannas (as per MODIS-IGBP land cover) within forests will not make much of a difference outside the tropics (Figure 4.9). Clearly a lack of consensus on how different land-cover types are defined is a source of uncertainty in LULUC emission estimates.

Historical LULUC emissions

The historical 'total' LULUC emissions simulated by ISAM have been compared previously (see Jain et al., 2013; Ciais et al., 2013). For the historical period, we limit discussion to LULUC fluxes that require elaboration.

The direct LULUC emissions estimated without nitrogen limitation are most comparable to estimates from Houghton's bookkeeping model (Houghton, 2003, 2008), in terms of definition. Strikingly, our estimated non-tropical emissions for 20th century (63 PgC from Table 4.4 based on average of three reconstructions) are 57% higher than Houghton's estimate of 40 PgC (data from Figure 1b of Richter and Houghton (2011)). Most of the difference is explained

by the underlying LULUC datasets. Two of our three LULUC reconstructions (based on Ramankutty (2012) and HYDE agricultural datasets; see Meiyappan and Jain, 2012) show more net forest loss in the non-tropics compared to the third dataset (based on FRA forest data; FAO (2006)) also used in Houghton's model. Our estimates using FRA data for the non-tropics (41 PgC from Table 4.4 range) compares with Houghton's bookkeeping estimate.

Future LULUC emissions

A comparison of our future LULUC emissions with other published estimates is shown in Table 4.7. Some climate models participating in CMIP5 did not simulate LULUC emissions for the future, but instead were driven using CO_2 emissions from LULUC estimated by the IAMs which produced the RCPs. These estimates include both wood harvest and land-use change. The definition and methodology of calculating LULUC emissions differed among the IAMs (van Vuuren et al., 2011; Pongratz et al., 2014). Our total (and direct) LULUC emissions estimated without nitrogen limitation (including wood harvest) are either much smaller emissions or even sinks compared with the IAM estimates. Our total (and direct) LULUC emissions estimated with nitrogen limitation are larger than IAM estimates.

The LUCID-CMIP5 project, using five Earth System Models estimated the range of total LULUC emissions for two RCPs. None of the models account for nitrogen limitation, they vary significantly in their carbon-cycle representations, and only one model (MPI-ESM-LR) included wood harvest. The authors acknowledge that their LULUC emissions from MPI-ESM-LR are overestimated due to high initial carbon stocks. Excluding MPI-ESM-LR, the range is 24-70 PgC (RCP2.6) and 30-67 PgC (RCP8.5). For comparison, our estimated global total LULUC emissions excluding wood harvest and nitrogen limitation, but including climate uncertainties are (cumulated over the 21st century): 14-43 PgC (RCP2.6) and 5-44 PgC (RCP8.5) (Figure 4.4c, e).

Kato et al. (2013) using a terrestrial carbon cycle model, estimated total LULUC emissions for the 21st century. Their estimates do not account for nitrogen limitation, wood harvest, and changes in future climate (static climate corresponding to current conditions are used). Kato et al. estimates are larger compared to our estimates without wood harvest and nitrogen limitation (Table 4.7).

Using a coupled climate model, Lawrence et al. (2012) reported LULUC emissions from one single with-LULUC simulation, without a reference no-LULUC simulation. Therefore, their LULUC emission estimates include only instantaneous and legacy fluxes from LULUC. Fluxes from regrowth sinks and decomposition of on-site resides are not counted towards LULUC flux (Pongratz et al., 2014). Their model includes nitrogen cycle, the effect of wood harvested from both forest and non-forest trees, in addition to cropland and pastureland transitions. Their estimates are the larger compared to all published studies and our estimates that include nitrogen limitation and wood harvest across all the scenarios. A large part of their LULUC emissions results from including wood harvest (their Figure 4.8a, c).

Stocker et al. (2014) using a dynamic global vegetation model reported both total and direct LULUC emissions that include nitrogen cycle, wood harvest, and cropland and pastureland transitions. They accounted for carbon and nitrogen pools between primary and secondary land separately. However, they do not explicitly model secondary forest regrowth dynamics i.e. the process formulations and model parameters are identical between primary and secondary land. Notably, Stocker et al. used one climate model output (corresponding to model 32 in Table S1) to estimate indirect LULUC emissions (total minus direct emissions) which are smaller compared to our mean estimates (Table 4.4). However, our multi-model climate sensitivity analysis indicate that the choice of climate data used to force a model can result in substantially different indirect emissions (difference of up to 88PgC globally; Table 4.5).

Wang et al. (2015) using an Earth System Model reported total LULUC emissions for RCP4.5 and 8.5. In addition to a nitrogen cycle, they also included phosphorous limitation, and wood harvested from both forests and non-forests. Their estimates are lower compared to both our total (and direct) emissions and other studies that include nitrogen limitation and wood harvest.

To summarize, we find that global estimates of LULUC emissions cumulated for 21st century are highly uncertain varying by ~300 PgC (range: -36 to 266 PgC) across published studies (estimates including nitrogen cycle when available). RCP4.5 has the widest range of results varying by ~200 PgC, varying from sink to a source (Table 4.7). Three studies (ours, Stocker and Wang) that included nitrogen limitation, wood harvest, and regrowth sinks also show widest range of estimates for RCP4.5. There are multiple reasons that could explain these

differences, with one possible reason being difference in implementing afforestation data across models (Di Vittorio et al., 2014).

Model uncertainties

The multi-model comparison presented above characterizes uncertainties across different models. However, an important source of uncertainty is model parameterization i.e. a single model can produce different LULUC emission estimates by varying model parameters within their uncertainty range (Exbrayat et al., 2013). During model development (e.g. Yang et al., 2009), we have evaluated and calibrated key model parameters based on available observations. However, limited observations also make some of the model parameters highly uncertain. By perturbing two key model parameters as an example, we highlight the impacts of parameter uncertainties on our emission estimates.

First, we test the sensitivity of our assumption that 25% of organic carbon stored in the top meter of the soil is released to the atmosphere when native soils are cleared for cultivation (section 4.3). While numerable meta-analysis (Don et al., 2011; Guo and Gifford, 2002; Murty et al., 2002; Post and Kwon, 2000) broadly report 25-30% loss on an average across all ecosystems, soil types, management practices, and decomposition processes, the variability about the average is large (range: 15-50% from Table 3 of Houghton and Goodale (2004)). We estimated total LULUC emissions (with nitrogen limitation) assuming 22.5% and 32.5% loss roughly corresponding to the 25^{th} and 75^{th} percentile across the observational range. Results show that our global mean estimates (Table 4.4) could vary by a maximum of -6% (25^{th} %ile) to +18% (75^{th} %ile) across the RCPs (Table S2).

Second, we model biological nitrogen fixation as a function (linear regression) of evapotranspiration, specific to biome type (based on Schimel et al. (1996)). Nitrogen fixation is the largest source of nitrogen input to terrestrial ecosystems; however, its magnitude is also highly uncertain (overall range of 40-290 $TgNyr^{-1}$ with estimates being revised downwards; see Cleveland et al. (1999); Wang and Houlton (2009); Vitousek et al. (2013); Sullivan et al. (2014)). By perturbing the regression parameters across all biomes by ±50% (our maximum assumed standard error), the mean estimates for global total LULUC emissions across RCPs (Table 4.4) vary by -6.4% (+50% perturbation) to +7.3% (-50% perturbation) (Table S3).

The high uncertainty in nitrogen fixation not only reflects limited measurements, but also gaps in mechanistic understanding of nitrogen fixation (Thomas et al., 2015). Consequently, parameters are just one source of uncertainty in our model. Incomplete understanding on various processes including nitrogen fixation cause structural uncertainties in model. For example, the relationship between nitrogen fixation and evapotranspiration is not from mechanistic understanding, but broadly captures the spatial observation that higher rates of nitrogen fixation are from humid settings with relatively high evapotranspiration (Cleveland et al., 1999). Further, nitrogen fixation can occur via free-living bacteria or symbiotic relationships (Batterman et al., 2013; Houlton et al., 2008; Vitousek et al., 2013). Therefore, harvesting of nitrogen-fixing trees may have different consequences for regrowth patterns than evapotranspiration would imply. Nonetheless, most land models to date estimate nitrogen fixation solely as a function of evapotranspiration or NPP (e.g. Hayes et al., 2011; Oleson et al., 2013; Wania et al., 2012; Zaehle and Friend, 2010); while, both NPP and evapotranspiration based approach have shortcomings, the NPP based approach contradicts empirical knowledge (Wieder et al., 2015b). Few land models have moved towards a more mechanistic representation of nitrogen fixation that echoes empirical understanding (Gerber et al., 2010; Wang et al., 2010). Implementing new approaches in models requires substantial efforts on observational data synthesis to parameterize and evaluate model improvements (Wieder et al., 2015b).

In summary, both parameter and structural uncertainty across all land models including ours extend beyond those discussed above (e.g. Jones et al., 2013; Todd-Brown et al., 2014). Consequently, these modeling uncertainties impose limits on the accuracy of simulated terrestrial processes.

Caveats

In this study, secondary forests result only from agricultural abandonment and wood harvest. This is because we infer secondary forests from changes in cropland, pastureland and wood harvest areas (Hurtt et al., 2011; Meiyappan and Jain, 2012). Several countries across the world create more forests through massive reforestation and afforestation efforts that add to the land carbon sink (Fang et al., 2014; FAO, 2010). We account for carbon sinks from conversion of crops and pastures to secondary forests on lands that were historically forested (reforestation) and non-forested (afforestation). However, we do not account for afforestation on land that is not

cropland or pastureland as secondary-to-secondary land conversion information is unavailable (Hurtt et al., 2011). During 2000-05, Houghton (2013b) estimated that afforestation in the tropics had contributed to ~1% of the region's total gross sinks. Globally, the share might increase in the future, as countries increase their land carbon storage through management practices as modeled in IAM scenarios and even pledged under the United Nations Framework Convention on Climate Change (UNEP, 2013).

This study does not include wood harvested from non-forest tree types and savannas, which Hurtt et al. (2011) count as forest. This is because ISAM classifies these types as herbaceous (Yang et al., 2010). Herbaceous land-cover types have lesser capacity to store carbon than forests. Our analysis of Hurtt et al. (2011) data indicates that accounting for non-forest wood harvest would have increased our gross carbon source from wood harvest by 10-37% during the 21st century (Table 4.6). A part of this biomass harvested would be compensated through regrowth sinks thus making a minor difference to our estimated total LULUC emissions.

We infer land-use changes in the model using net changes in cropland and pastureland areas between consecutive years within each grid cell (Hurtt et al., 2011; Meiyappan and Jain, 2012). This is because existing land use reconstructions (including HYDE used in Hurtt data) draw upon (sub-) national land use statistics at annual time steps that are the net changes. In reality, it is the gross changes (all area gains and losses) that determine the LULUC fluxes. For example, land use statistics collected at administrative level (e.g. state or country level data typically used in historical reconstructions) can indicate zero change in cropland area between two years, but it does not imply that cropland area has remained unchanged in every grid within the administrative region. Similarly, within a grid cell, different sub-grid areas can undergo land cover change (gross changes) in rotation (e.g. crop to forest, forest to grass, and grass to crop), but at the grid cell level the net change in land cover areas could fully or partly cancel out (Fuchs et al., 2014). However, these sub-grid changes would still affect the carbon fluxes and land carbon storage over time. In such cases, we might be underestimating the total LULUC emissions. Currently, there is no consensus on how a given LULUC data be implemented within a model (Brovkin et al., 2013; Pitman et al., 2009; Wilkenskjeld et al., 2014). Our interpretation of net changes in land-use area within grid cells is consistent with the economic rationale in

spatial land-use allocation modules of IAMs that humans tend to lower the cost associated with relocating land areas (Meiyappan et al., 2014; Verburg and Overmars, 2009).

Several other LULUC activities such as shifting cultivation, agricultural management, fire management, land degradation, peatlands, erosion, and woody encroachment have not been included in this study. These factors together could be significant in the global carbon budget, but estimates for some of these factors are highly uncertain even for recent past (Houghton et al., 2012). From 2000-05, Houghton (2013b) estimated that direct emissions from shifting cultivation (0.082 PgC/yr) accounted for ~7% of total direct emissions in the tropics. Gross sources from shifting cultivation are much larger (~27% of the total gross sources), but regrowth sinks on fallows balances most of the gross sources.

We represent cropland as a 'generic' category in our model. Therefore, we do not explicitly simulate the management effects of bioenergy crops/plantations on LULUC emissions. The treatment of bioenergy across the four independent IAM groups that produced the four RCPs is different (test S1). For example, bioenergy is included in wood harvest in RCP8.5, whereas bioenergy is included in cropland in RCP2.6. The 'land-use change' effects of implementing bioenergy within croplands (as opposed to its land use/management effects) are however captured by Hurtt et al. (2011) data that drive our land-surface model, and hence by our LULUC emission estimates. For example, Hurtt estimates for RCP2.6 shows the largest increase in cropland area due to bioenergy (Figure 4.1), mostly at the expense of forests (Table 4.1).

The study does not account for two key model processes. First is the co-limitation of phosphorus with nitrogen, especially in the moist tropics (Vitousek et al., 2010). Only recently have models started to include phosphorus dynamics (Goll et al., 2012; Yang et al., 2013; Zhang et al., 2013), and only Zhang et al. (2013) represent LULUC. Second are the impacts of LULUC on climate realized through biogeophysical pathways (Mahmood et al., 2013). Brovkin et al. (2013), Kumar et al. (2013), and Lawrence et al. (2012) have examined the biogeophysical impacts of LULUC for the RCPs.

Summary and implications of results for climate modeling and climate policy

Our analysis offers insight into complex interactions among CO_2 emissions from LULUC, environmental changes, and nitrogen limitation effect on the regrowth sinks. Table 4.8

summarizes our model estimated uncertainty across different drivers. There are four key conclusions from our modeling study.

First, nitrogen limitation of CO_2 uptake is substantial and sensitive to nitrogen inputs. In our model, excluding nitrogen limitation underestimated global total LULUC emissions by 34-52 PgC (~21-29%) during the 20th century and by 128-187 PgC (90-150%) during the 21st century (Table 4.8). The difference increases with time because nitrogen limitation will progressively down-regulate the magnitude of CO_2 fertilization effect on regrowing forests, due to decreasing supply of plant-usable mineral nitrogen. Further, regrowing forests become increasingly nitrogen limited due to LULUC-related nitrogen losses from the system. Without large amounts of nitrogen input to the system, the regrowing forests are likely to be nitrogen limited. To meet the same mitigation target despite larger total LULUC emissions would require an equivalent greater reduction of fossil fuel emissions.

Second, including nitrogen limitation changes the region with the highest total LULUC emissions from the tropics to the non-tropics. The tropics had higher emissions in our simulations without nitrogen limitation, and also earlier studies that considered only the interactive effects of CO_2 and climate. Total LULUC emissions from the non-tropics are greater when the nitrogen cycle is included mainly because the carbon uptake capacity of secondary forests following LULUC is limited by nitrogen deficiency.

Third, historically, the indirect effects of anthropogenic activity through environmental changes in land experiencing LULUC (indirect emissions) are small compared to direct effects of anthropogenic LULUC activity (direct emissions). As a result, including or excluding indirect emissions had a minor influence on the estimated total LULUC emissions historically. In contrast, the indirect LULUC emissions for the 21st century are a much larger source to the atmosphere, in simulations with nitrogen limitation (Table 4.4). This is because of the gradual weakening of the photosynthetic response to elevated (CO_2) caused by nitrogen limitation.

In this study, we separately accounted for the effects of nitrogen limitation in both direct and indirect LULUC emissions. In principle, the nitrogen limitation effects are also an indirect effect of anthropogenic activity due to environmental change impacts on natural plant processes, hence can be fully counted within indirect emissions (i.e. exclude the effect of nitrogen limitation

from direct emissions, and add it to indirect emissions). Following such an accounting procedure will further increase the indirect LULUC emissions for the 21st century (123-162 PgC; calculated from Table 4.4 as the difference between total LULUC emissions estimated with nitrogen limitation and direct LULUC emissions estimated without nitrogen limitation), and will dominate over direct emissions (-39–31 PgC; Table 4.4 without nitrogen limitation case). By either method, our results indicate that treatment of environmental factors can substantially influence the estimated total LULUC emissions for the future (see Houghton (2013a) for an associated discussion).

Fourth, the choice of climate model projection used to force a land model can substantially impact the estimated indirect (and total) LULUC emissions (Table 4.8). The climate induced uncertainty ranges are larger than the mean estimates of global indirect LULUC emissions cumulated over the 21st century for three RCP scenarios. Further, the indirect LULUC emission estimated for the non-tropics are affected more by climate uncertainties than for the tropics, because larger areas under LULUC (especially wood harvest) in the non-tropics coincide spatially with regions where climate uncertainties are high.

While interpreting our results, the limitations highlighted earlier should be kept in mind. Notably, using one land-surface model is potentially a limiting factor because it does not represent a broad range of model physics response, especially given that there are significant uncertainties in modeling both nitrogen and carbon cycles (Houghton et al., 2012; Friedlingstein and Prentice, 2010), LULUC activities considered (Houghton et al., 2012), and even the method of implementing a given LULUC dataset across biosphere models (Brovkin et al., 2013; Pitman et al., 2009). Conversely, using a single land-surface model is more appropriate for our analysis because we can consistently isolate the effects on LULUC emissions due to different LULUC activities, LULUC flux definitions, historical LULUC forcings, and future climate forcings. The above effects cannot be consistently isolated using multi-model comparisons because model-based differences (e.g. different land cover representations) make attribution difficult.

In summary, Hurtt et al. (2011) show that excluding wood harvest alone can underestimate secondary land by 57% on an average for RCPs, and so the associated carbon source. Even if land management is represented, excluding nitrogen limitation will overestimate the carbon sinks on land recovering from LULUC, thereby underestimating total LULUC

emissions. It is the total LULUC emissions that the atmosphere sees which can be mitigated by reversing or avoiding any LULUC activity. Notwithstanding the aforementioned caveats, our study implies that the effectiveness of land-based mitigation strategies would critically depend on the interactions between nutrient limitations and secondary forests resulting from LULUC. Therefore, it is important for terrestrial biosphere models to consider nitrogen limitation in estimates of the strength of the future land carbon sink, especially on regrowing forests.

Acknowledgements

We thank Benjamin Houlton and the anonymous reviewers for helpful suggestions. We acknowledge the WCRP's Working Group on Coupled Modelling and the climate modeling groups for producing and making available their model output. The NASA Land Cover and Land Use Change Program (NNX14AD94G), the US National Science Foundation (NSF-AGS-12-43071), and the U.S. Department of Energy (DOE-DE-SC0006706) supported this work. JH was funded by a Leverhulme Early Career Research Fellowship, UK.

4.6 Tables

Table 4.1 Net change in forest area estimated by the Integrated Science Assessment Model (net forest loss including afforestation, and forest regrowth following cropland and pastureland abandonment; negative values indicate a net loss in forest area) and the annual forest harvested areas summed over a hundred year period (from Hurtt et al. 2011). The historical estimates are averages of the three LULUC reconstructions described in Jain et al. (2013). The data for the 21st century correspond to the four Representative Concentration Pathways (RCPs). All units in million km^2/century.

Region	Net change in forest area					Cumulative wood harvest area from forests (Hurtt et al., 2011)				
	20th century	21st century				20th century	21st century			
	Historical	RCP 2.6	RCP 4.5	RCP 6.0	RCP 8.5	Historical	RCP 2.6	RCP 4.5	RCP 6.0	RCP 8.5
Global	-4.6	-2.6	2.3	-0.5	-2.1	16	76	87	188	137
Tropics	-2.2	-1.1	0.9	-0.2	-1.0	6	45	60	98	72
Non-Tropics	-2.4	-1.5	1.4	-0.3	-1.1	10	31	27	90	65

Table 4.2 Design of the simulations. Tick mark (✓) indicates the variable was varied with time. Cross mark (×) indicates the variable was held static at initial (assumed zero for nitrogen deposition and LULUC) value. Inclusion of nitrogen deposition is irrelevant when nitrogen dynamics is inactive in the model, and is indicated by a hifen (-).

Simulation	CO_2	Climate	Nitrogen Deposition	Land-Use Change	Wood Harvest	Nitrogen Dynamics
Ref_1	✓	✓	✓	×	×	Active
A1	✓	✓	✓	✓	✓	Active
B1	✓	✓	✓	✓	×	Active
C1	×	×	×	✓	✓	Active
D1	×	×	×	✓	×	Active
Ref_2	✓	✓	-	×	×	Inactive
A2	✓	✓	-	✓	✓	Inactive
B2	✓	✓	-	✓	×	Inactive
C2	×	×	-	✓	✓	Inactive
D2	×	×	-	✓	×	Inactive

Table 4.3 Summary of how the different simulations mentioned in Table 4.2 were combined to estimate land use and land-use change (LULUC) fluxes with varying environmental factors, LULUC activities, and nitrogen dynamics. Total LULUC emissions are the sum of direct and indirect LULUC emissions. Land-use change is abbreviated as 'LUC' and wood harvest as 'WH'.

LULUC flux estimated	Effects Included			Calculation Method
	LULUC activities	Changing environmental factors	Nitrogen Dynamics	
Total emissions with nitrogen limitation	LUC+WH	CO_2 + nitrogen deposition + climate	Active	A1 – Ref1
	LUC			B1 – Ref1
	WH			(A1 – B1)
Total emissions without nitrogen limitation	LUC+WH	CO_2 + climate	Inactive[a]	A2 – Ref2
	LUC			B2 – Ref2
	WH			(A2 – B2)
Direct emissions with nitrogen limitation	LUC+WH	None	Active	C1
	LUC			D1
	WH			(C1 – D1)
Direct emissions without nitrogen limitation	LUC+WH	None	Inactive[a]	C2
	LUC			D2
	WH			(C2 – D2)
Indirect emissions with nitrogen limitation	LUC+WH	CO_2 + nitrogen deposition + climate	Active	(A1 – Ref1) – C1
	LUC			(B1 – Ref1) – D1
	WH			(A1 – B1) – (C1 – D1)
Indirect emissions without nitrogen limitation	LUC+WH	CO_2 + climate	Inactive[a]	(A2 – Ref2) – C2
	LUC			(B2 – Ref2) – D2
	WH			(A2 – B2) – (C2 – D2)

[a] Inclusion of nitrogen deposition is irrelevant for "without" nitrogen limitation case.

Table 4.4 Direct, indirect, and total emissions from land use and land-use change (LULUC). The estimates for the 20th century are based on the three LULUC reconstructions. The estimates for the 21st century are based on the four Representative Concentration Pathways (RCPs). Two sets of estimates are shown, one with and the other without the effect of nitrogen limitation. For each RCP, an array of LULUC fluxes were estimated, using outputs from a suite of climate model projections from the CMIP5 multi-model ensemble database (Table S1). The estimates shown for the RCPs are the mean across the array of estimates. Positive values indicate a land to atmosphere flux. Units are in PgC/century.

	With Nitrogen Limitation Effect					**Without Nitrogen Limitation Effect**				
	20th century	**21st century**				**20th century**	**21st century**			
Land-use affected ecosystem exchange	**Mean (& range)**	**RCP 2.6**	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**	**Mean (& range)**	**RCP 2.6**	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**
Global										
Direct LULUC emissions	+167 (135 to 186)	+81	+68	+35	+96	+123 (93 to 142)	+14	-39	-31	+31
Indirect LULUC emissions	-4 (-22 to 21)	+56	+55	+77	+71	0 (-18 to 29)	-5	-25	+6	-13
Total LULUC emissions	+163 (156 to 174)	+137	+123	+112	+167	+123 (122 to 124)	+9	-64	-25	+18
Tropics										
Direct LULUC emissions	+61 (43 to 85)	+22	-2	+5	+25	+60 (43 to 84)	+15	-17	-3	+33
Indirect LULUC emissions	-1 (-9 to 8)	+29	+31	+37	+40	0 (-8 to 18)	+8	0	+8	+11
Total LULUC emissions	+60 (51 to 76)	+51	+29	+42	+65	+60 (35 to 76)	+23	-17	+5	+44
Non-Tropics										
Direct LULUC emissions	+106 (80 to 143)	+59	+70	+30	+71	+63 (41 to 89)	-1	-22	-28	-2
Indirect LULUC emissions	-3 (-20 to 25)	+27	+24	+40	+31	0 (-10 to 11)	-13	-25	-2	-24
Total LULUC emissions	+103 (82 to 123)	+86	+94	+70	+102	+63 (48 to 87)	-14	-47	-30	-26

Table 4.5 Climate projections induced uncertainties in simulating total (direct + indirect) LULUC emissions for the 21st century. The numbers shown for total LULUC emissions are the maximum range of estimates obtained by forcing the Integrated Science Assessment Model (ISAM) with multiple climate model outputs (Table S1). The "(Max – Min) value" is calculated as the difference between the maximum and minimum value from the estimated range. The mean estimates are provided in Table 4.4. Positive values for emissions indicate a land to atmosphere flux. The estimates provided here include the effect of nitrogen limitation. Units are in PgC/century.

Region	**Range of cumulative total LULUC emissions**				**(Max – Min) value**			
	RCP 2.6	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**	**RCP 2.6**	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**
Global	116–180	107–165	91–150	139–227	64	60	59	88
Tropics	33–72	21–50	32–59	48–96	39	29	27	48
Non-Tropics	67–119	74–126	51–99	77–146	52	52	48	69

Table 4.6 Comparison of biomass harvested from forests between Hurtt et al. (2011) and this study. The numbers provided within brackets are estimates of biomass harvested from non-forested tree types which we do not account for. The range of estimates provided for the historical period (corresponding to this study) is obtained using the three different LULUC reconstructions. The range of estimates provided for the RCPs, are based on estimates with and without the effects of nitrogen limitation. Lower end values are generally the estimates that include the effect of nitrogen limitation. Both the estimates with and without the effect of nitrogen limitation for the RCPs are mean estimates obtained by driving the Integrated Science Assessment Model (ISAM) using multiple climate model projections (Table S1). Units are in PgC/century.

	Hurtt et al. (2011)					**This study**				
Region	**20th century**	**21st century**				**20th century**	**21st century**			
	Historical	**RCP 2.6**	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**	**Historical**	**RCP 2.6**	**RCP 4.5**	**RCP 6.0**	**RCP 8.5**
Global	70 (3)	144 (22)	165 (17)	150 (35)	182 (68)	69–76	88–93	113–116	69–77	143–146
Tropics	18 (2)	38 (18)	54 (14)	53 (22)	61 (30)	11–13	28–33	33–37	37–41	57–62
Non-Tropics	52 (1)	106 (4)	111 (3)	97 (13)	121 (38)	56–65	56–65	77–83	29–41	82–90

Table 4.7 Global comparison of our model estimated cumulative (2001-2100) LULUC fluxes with previous studies (PgC/yr).

Reference	RCP2.6	RCP4.5	RCP6.0	RCP8.5	Notes	Nitrogen cycle	Wood harvest
This study							
Data from Table 4.4	14 9	-39 -64	-31 -25	31 18	Direct emissions Total emissions	x	✓
Data from Figure 4.4	25	-46	4	31	Total emissions	x	x
Data from Table 4.4	81 137	68 123	35 112	96 167	Direct emissions Total emissions	✓	✓
Other studies							
IAMs that produced the RCPs	68	30	6	60	Data from http://cmip-pcmdi.llnl.gov/cmip5/forcing.html	x	✓
Brovkin et al. (2013): LUCID-CMIP5	24 to 180 24 to 70	- -	- -	30 to 210 30 to 67	Range from five Earth System Models (one model included wood harvest) Range excluding MPI-ESM-LR	x x	See notes x
Boysen et al. (2014): LUCID-CMIP5 (extension to Brovkin et al.)	-	-	-	34 to 218 34 to 57	Range from four Earth System Models (one model included wood harvest) Range excluding MPI-ESM-LR	x x	See notes x
Kato et al. (2013)	118	-36	16	82	Data from their Figure 7	x	x
Lawrence et al. (2012)	185	158	191	266	Data from their Figure 8a	✓	✓
Stocker et al. (2014)	91 111	30 33	91 103	127 157	Direct emissions Total emissions	✓	✓
Wang et al. (2015)	-	-16	-	61	They provide estimates for 2006-2100 to which we added estimates for 2001-2005 based on the same model provided in Zhang et al. (2013).	✓	✓
Range	**24 to 185**	**-36 to 158**	**6 to 191**	**30 to 266**	Range across "Other studies"	-	-

Table 4.8 Summary of relative uncertainties in estimated 'total LULUC emissions' due to: (1) uncertainty in climate projections underlying future scenarios ('Climate'), (2) including nitrogen cycle ('Nitrogen cycle'), and (3) including wood harvest ('LULUC activities') under both with and without nitrogen limitation cases ('N lim' and 'No N lim' respectively). The ranges shown are minimum and maximum values of uncertainty estimated across the three historical reconstructions (for the 20th century), and across the four RCP scenarios (for the 21st century). We estimate the uncertainty for each LULUC history and RCP as follows. For 'Climate', the uncertainty values correspond to '(Max-Min) value' column in Table 4.5. For 'Nitrogen cycle', we calculated the difference in LULUC emission estimates between with and without nitrogen limitation case (from Table 4.4). For 'LULUC activities', we extracted the values corresponding to 'wood harvest' from figure 4.4 (brown bars). Units are in PgC/century.

Region	**Climate**	**Nitrogen cycle**	**LULUC activities (N lim)**	**LULUC activities (No N lim)**
20th century				
Global	-	34 to 52	55 to 64	-20 to 40
Tropics	-	0 to 16	8 to 11	8 to 10
Non-Tropics	-	34 to 53	47 to 56	-28 to 30
21st century				
Global	59 to 88	128 to 187	38 to 125	-29 to 11
Tropics	27 to 48	21 to 46	13 to 41	-8 to 11
Non-Tropics	48 to 69	100 to 141	17 to 84	-28 to 0

4.7 Figures

Figure 4.1 Annual rates of change in cropland and pastureland, annual wood harvest area from forests and annual net deforestation rates (net forest area loss including afforestation, and forest regrowth following cropland and pastureland abandonment; negative values indicate net forest loss) for 1900-2100. Figure legends are shown in panel (a). All units are in million ha/yr (1 ha = 0.01 km^2).

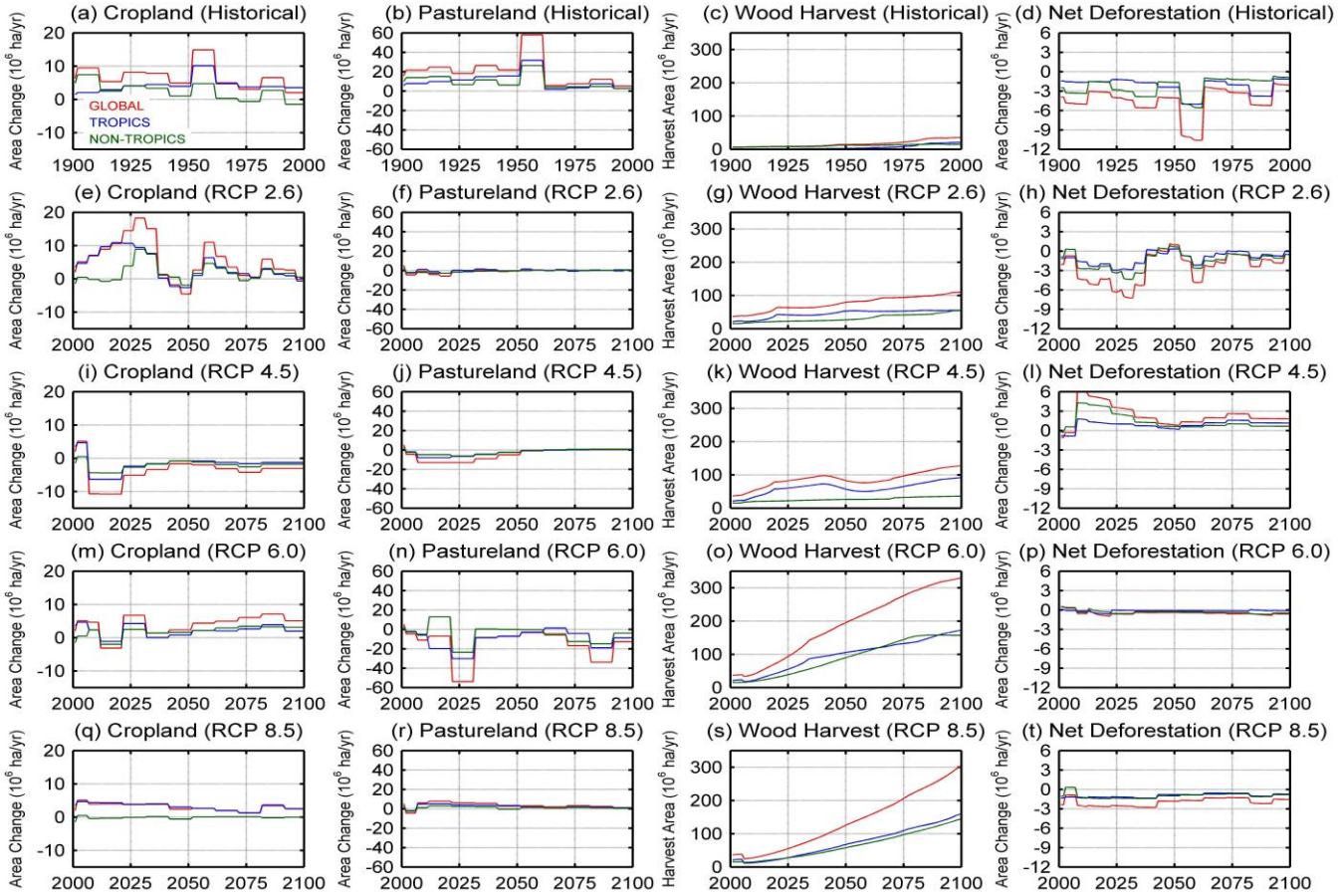


Figure 4.2 Zonal breakdown of above-ground carbon in forests (year 2000) based on: (1) our historical model simulations (averaged across estimates obtained using three LULUC datasets; including both primary and secondary forests), and (2) global gridded estimates based on FAO statistics (Kindermann et al., 2008). Darker grey shades in the background indicate larger forest area fraction along the latitude based on FAO statistics.

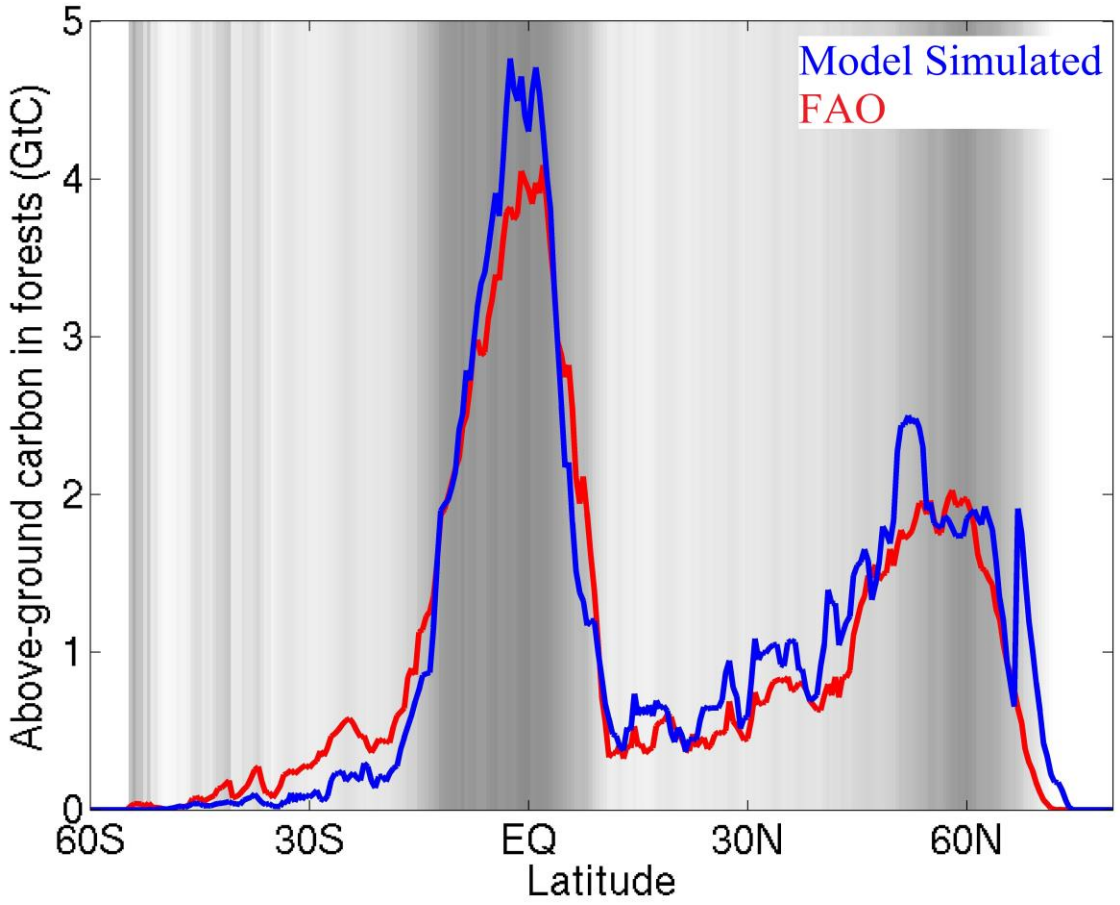


Figure 4.3 NPP estimated for different land cover types averaged globally over the period 2001-2005. The results are compared between (1) our historical model simulations, and (2) radiation-based modeled estimates of NPP derived from MODIS (Zhao and Running, 2010). The error bars indicate the standard deviation across the 5-year annual estimates. Our model-based error bars also encompass differences induced by LULUC datasets.

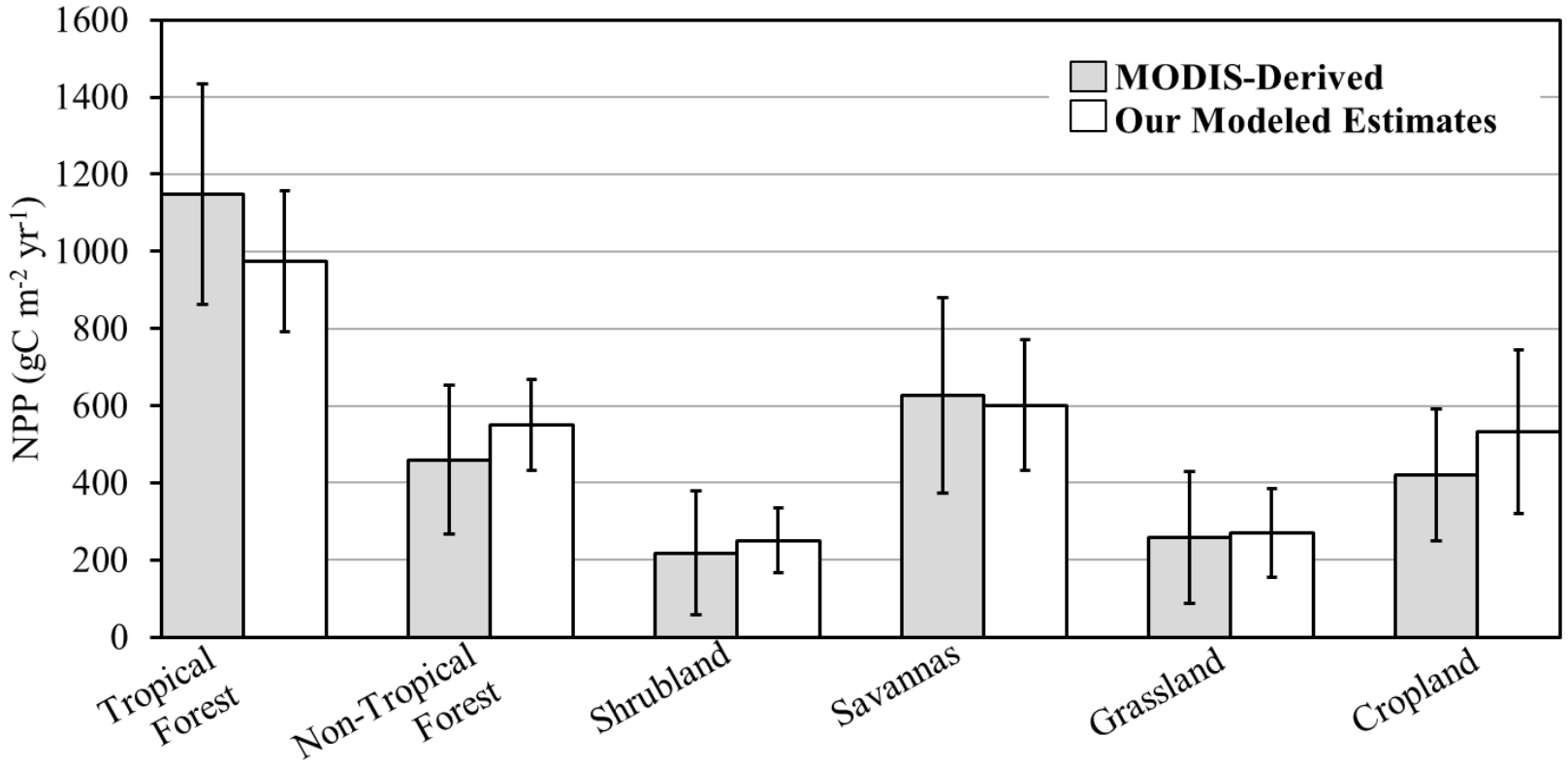


Figure 4.4 Break-down of LULUC emission fluxes attributable to land-use change (green bars), and forest wood harvest (brown bars). The fluxes shown are direct, indirect, and total (direct + indirect) emissions. Two sets of estimates are shown, one with and the other without the effect of nitrogen limitation. The first three bars in each panel are estimates that include the effect of nitrogen limitation ('With N lim' - see panel 'a'), and the other three bars are estimates without nitrogen limitation effect ('No N lim'). The historical estimates are averages of the three LULUC reconstructions. For each RCP, an array of LULUC fluxes were estimated, using outputs from a suite of climate model projections from the CMIP5 multi-model ensemble database (Table S1). The estimates shown for the RCPs are the mean across the array of estimates. Positive values indicate a land to atmosphere flux. Units are in PgC/century.

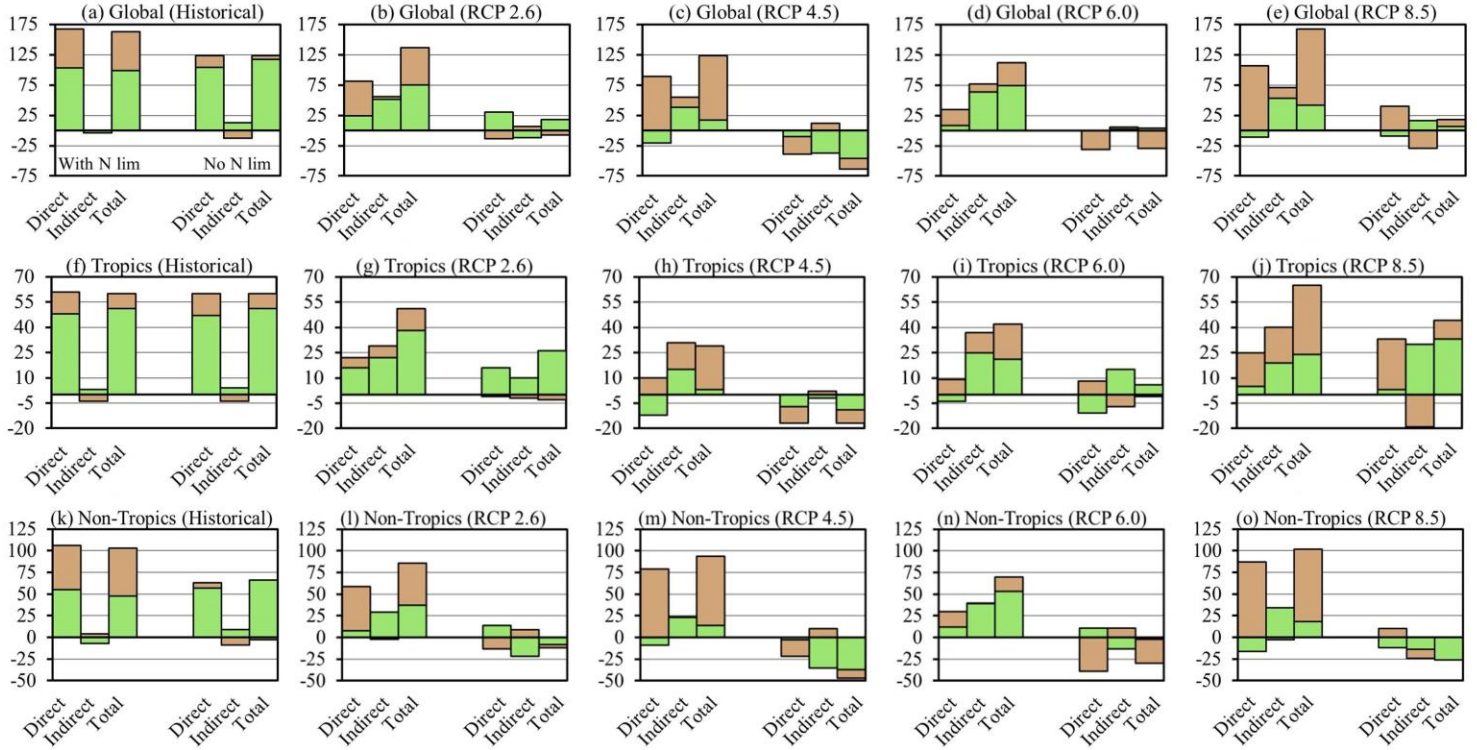


Figure 4.5 Estimates of direct, indirect and total (direct + indirect) emissions from LULUC averaged over each decade. The dark lines indicate the estimates that include the effect of nitrogen limitation. The grey lines are estimates obtained without nitrogen limitation effect. The historical estimates are averages based on the three LULUC reconstructions. For each RCP, an array of LULUC fluxes were estimated using outputs from a suite of climate model projections from the CMIP5 multi-model ensemble database (Table S1). The lines represent the mean across the array of estimates (for both with and without nitrogen limitation effect). The error bars indicate the uncertainty range in simulated indirect LULUC emissions (for with nitrogen limitation case) that results due to uncertainties in projecting future climate. Uncertainties in simulating indirect LULUC emissions will also introduce uncertainties in estimates of total LULUC emissions. For clarity, uncertainty estimates for total LULUC emissions are not shown, instead provided in Table 4.5. Units are in PgC/yr. The figure legends are shown in panel (d). See Figure S6 for figures corresponding to RCP2.6 and RCP6.0.

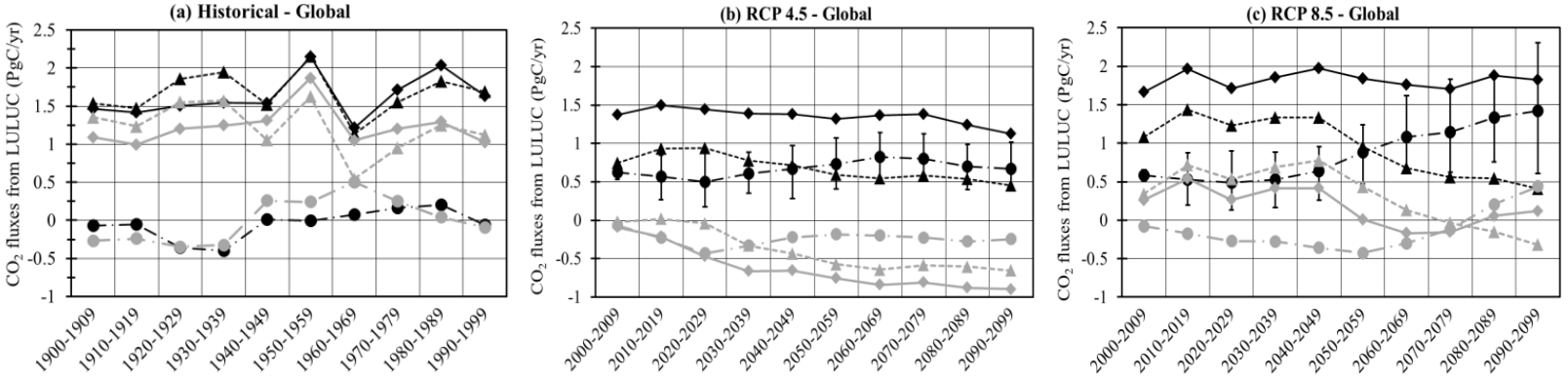


Figure 4.5 (Cont.)

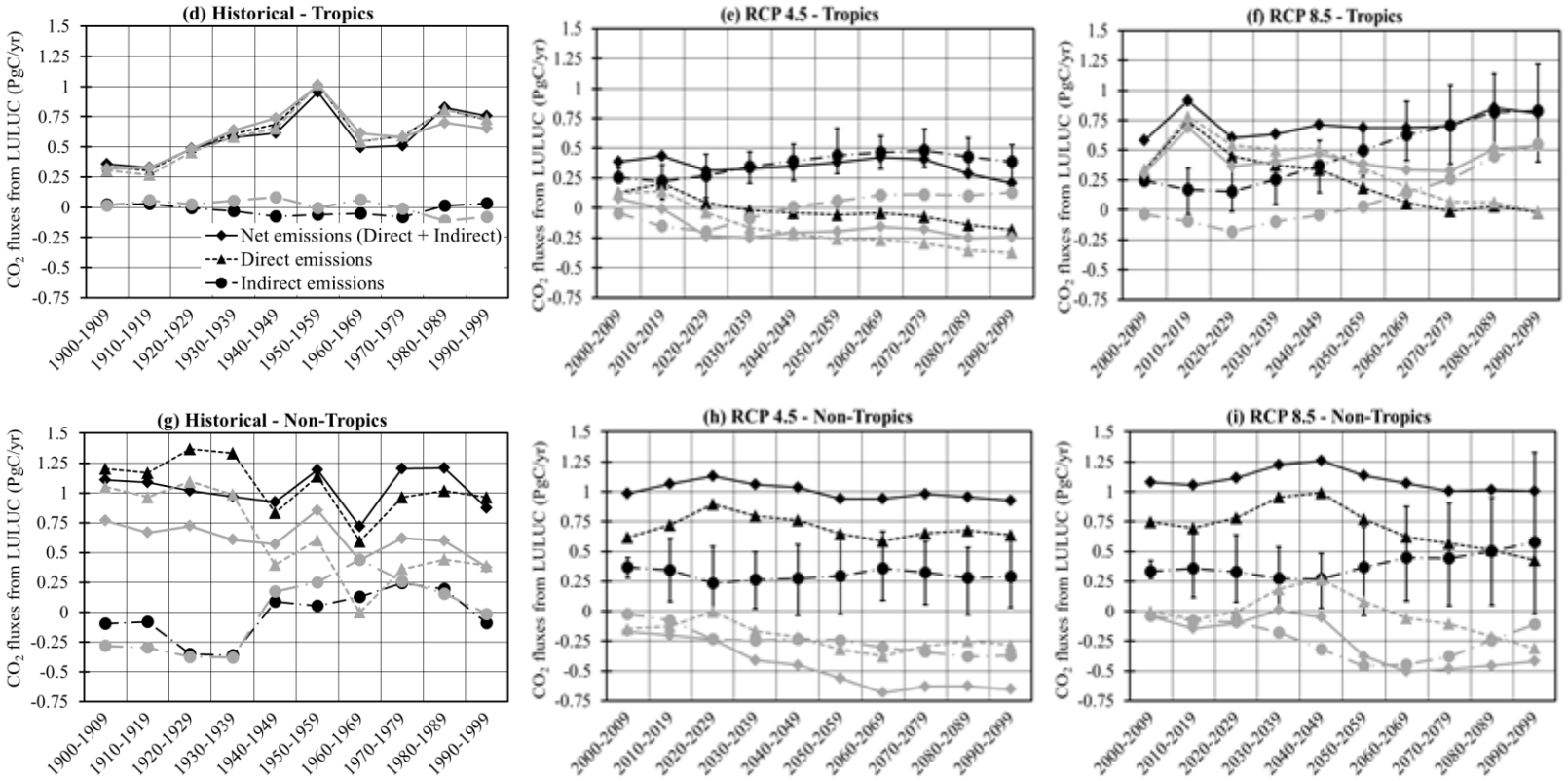


Figure 4.6 Net Primary Productivity (NPP) of secondary forests simulated by ISAM corresponding to RCP8.5. Results are compared between with and without nitrogen limitation case. The historical estimates are based on HYDE LULUC reconstruction. The future estimates shown are mean across the array of estimates obtained by driving ISAM with different climate model projections (for both with and without nitrogen limitation effect).

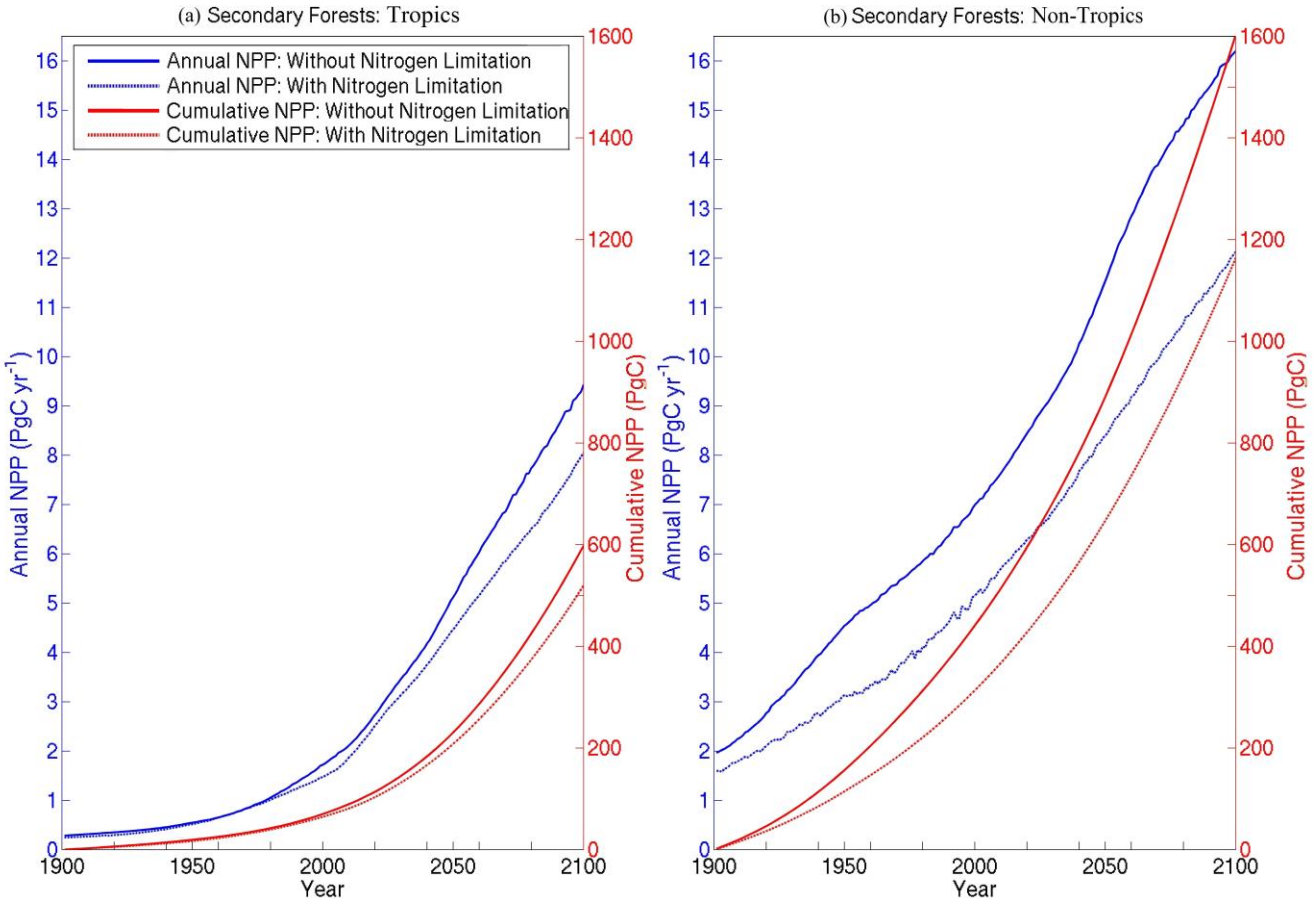


Figure 4.7 Model simulated response to key nitrogen variables illustrated using RCP8.5 simulations (with transient environmental factors) as example. First and second column figures correspond to tropics and the non-tropics respectively, and show the fluxes averaged over primary (unmanaged) and secondary forests (resulting from LULUC). Third column figures show fluxes by region, and includes all land cover types. Panel (a-i, l) is from simulations that include LULUC effect. Panel (j-k) is obtained by differencing fluxes obtained between "with" and "without" LULUC simulations. (a, b) Biological Nitrogen Fixation which a source of nitrogen to terrestrial ecosystems. (g, h) Nitrogen-Use Efficiency defined as the Net Primary Productivity (NPP; panels c-d) per unit uptake of nitrogen by plants (panels e-f). (i) Anthropogenic nitrogen deposition (NH_x +No_y) over land areas (source of nitrogen to both managed and unmanaged land). (j) Denitrification loss attributable to LULUC. (k) Nitrogen leaching loss attributable to LULUC. (k) Nitrogen loss from LULUC disturbance (product pool decays, slash burning and removals).

Figure 4.7 (Cont.)

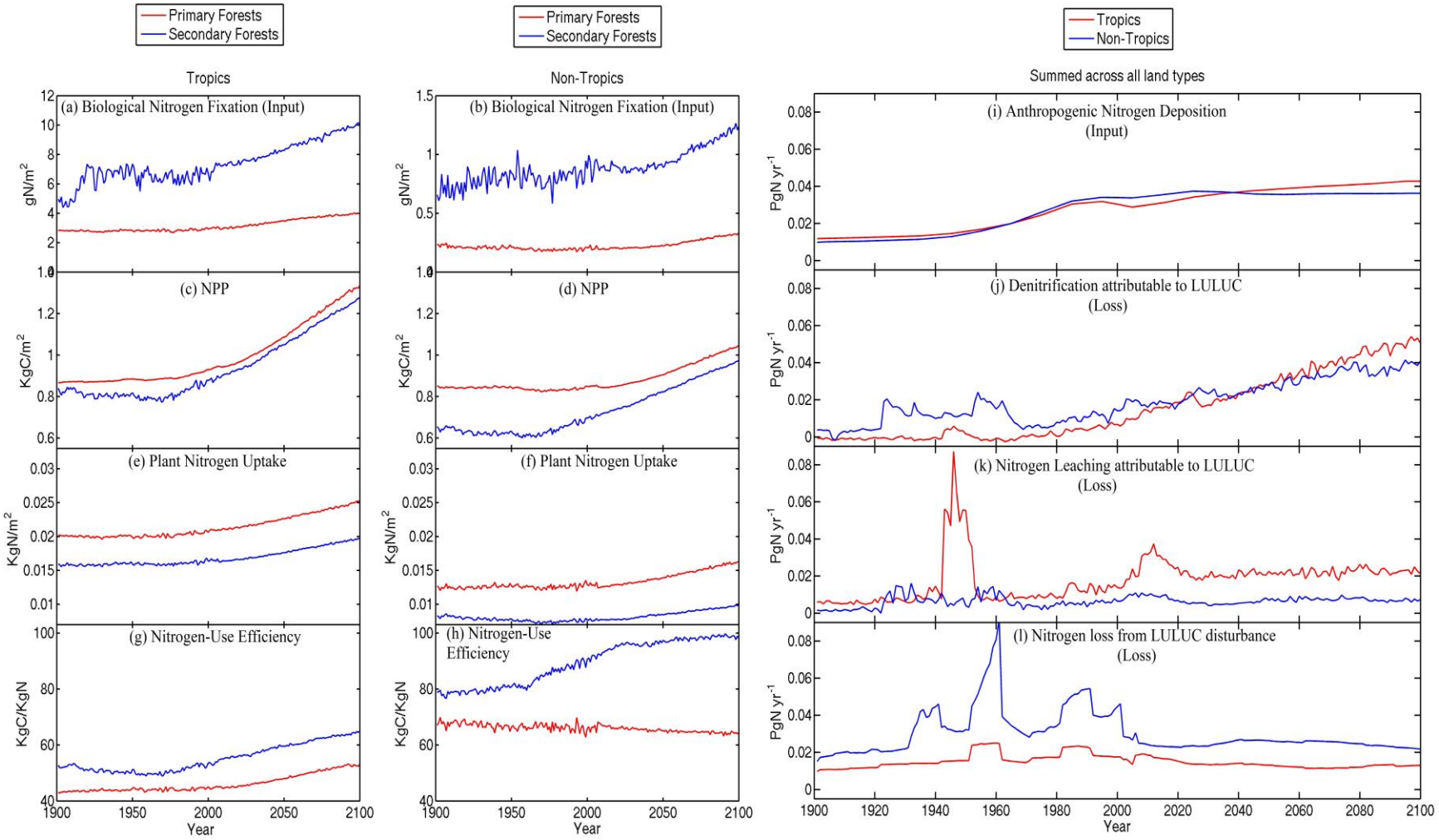


Figure 4.8 Two site-specific simulations (tropical and non-tropical forest site) showing our modeled response to the rate of vegetation carbon accumulation followed wood harvest. The "steady state" indicates the time taken to attain full maturity under ideal conditions (environmental factors unchanged from current site-specific conditions and no LULUC disturbance following wood harvest). An explanation of this figure is provided in text S7.

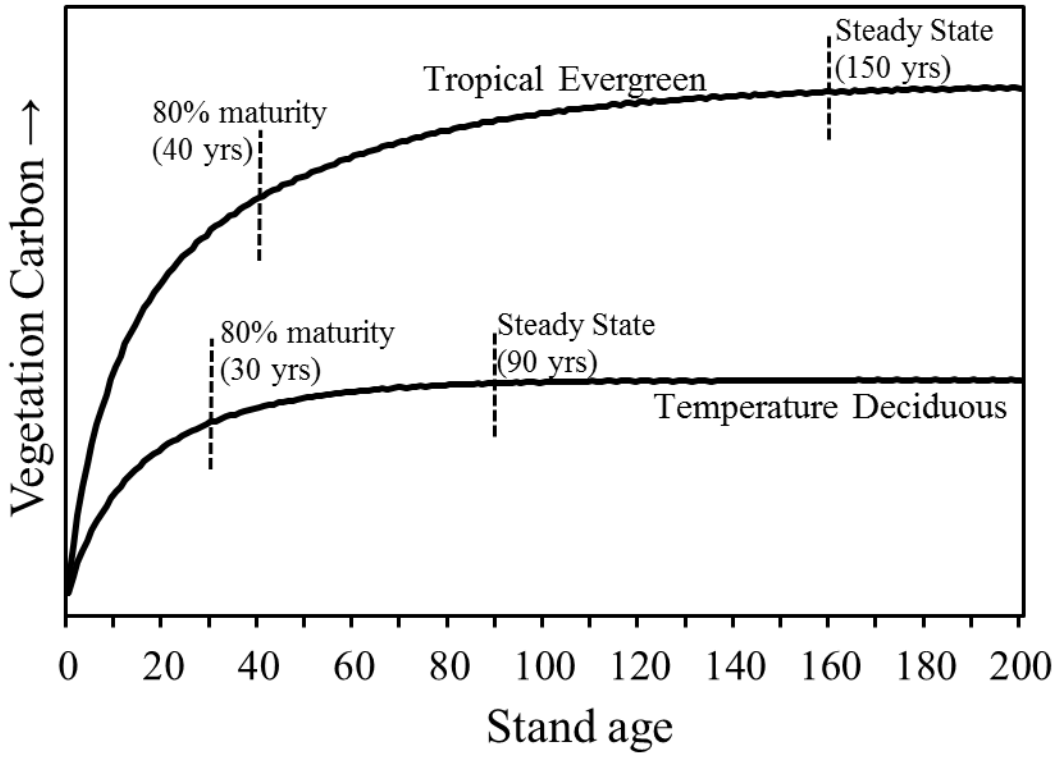
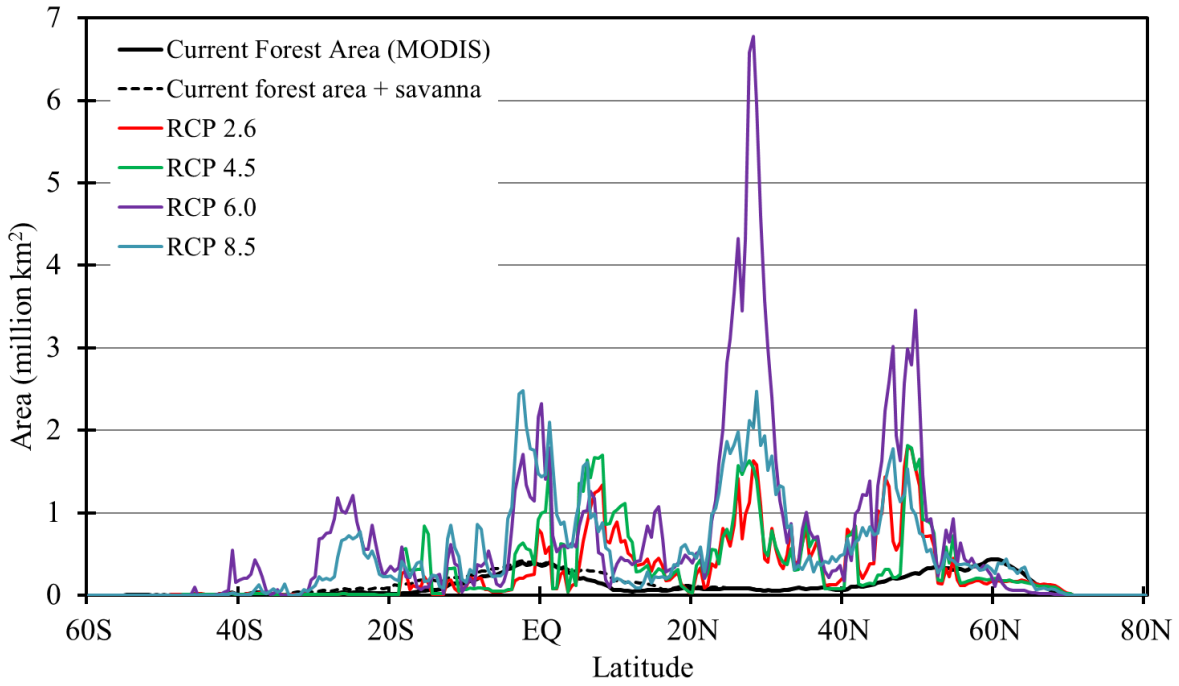


Figure 4.9 Zonal breakdown of prescribed forest harvest area for the four RCPs cumulated over the 21st century (data based on Hurtt et al. (2011)). For comparison, the contemporary (2005 AD) forest areas (and savannas) based on MODIS satellite data (Friedl et al., 2010) are shown. Units are in million km^2.



4.8 References

Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., DeGrandpre, M., Tans, P. P., Miller, J. B., Alden, C., and White, J. W. C.: Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty, Biogeosciences, 12, 2565-2584, 2015.Batterman, S. A., Hedin, L. O., Van Breugel, M., Ransijn, J., Craven, D. J., & Hall, J. S. (2013). Key role of symbiotic dinitrogen fixation in tropical forest secondary succession. Nature, 502(7470), 224-227.

Boden, T.A., G. Marland, and R.J. Andres. (2013). Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2013.

Boysen, L., Brovkin, V., Arora, V. K., Cadule, P., de Noblet-Ducoudré, N., Kato, E., ... & Gayler, V. (2014). Global and regional effects of land-use change on climate in 21st century simulations with interactive carbon cycle. *Earth System Dynamics, 5*, 309-319.

Brovkin, V., and Coauthors. (2013). Effect of Anthropogenic Land-Use and Land-Cover Changes on Climate and Land Carbon Storage in CMIP5 Projections for the Twenty-First Century. J. Climate, 26, 6859–6881.

Chapin III, F. S., Woodwell, G. M., Randerson, J. T., Rastetter, E. B., Lovett, G. M., Baldocchi, D. D., ... & Schulze, E. D. (2006). Reconciling carbon-cycle concepts, terminology, and methods. *Ecosystems*, *9*(7), 1041-1050.

Ciais P, Sabine C, Bala G, et al. (2013) Carbon and other biogeochemical cycles. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the International Panel on Climate Change (eds Stocker T, Qin D, Plattner G-K, et al.) Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Cleveland, C. C., Taylor, P., Chadwick, K. D., Dahlin, K., Doughty, C. E., Malhi, Y., ... & Townsend, A. R. (2015). A comparison of plot-based, satellite and Earth system model estimates of tropical forest net primary production. *Global Biogeochemical Cycles*. Doi: 10.1002/2014GB005022.

Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., ... & Wasson, M. F. (1999). Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. Global biogeochemical cycles, 13(2), 623-645.

Davidson, E. A., Reis de Carvalho, C. J., Vieira, I. C., Figueiredo, R. D. O., Moutinho, P., Yoko Ishida, F., ... & Tuma Sabá, R. (2004). Nitrogen and phosphorus limitation of biomass growth in a tropical secondary forest. *Ecological Applications*, *14*(sp4), 150-163.

Davidson, E. A., de Carvalho, C. J. R., Figueira, A. M., Ishida, F. Y., Ometto, J. P. H., Nardoto, G. B., ... & Martinelli, L. A. (2007). Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. Nature, 447(7147), 995-998.

Di Vittorio, A. V., Chini, L. P., Bond-Lamberty, B., Mao, J., Shi, X., Truesdale, J., ... & Thomson, A. (2014). From land use to land cover: restoring the afforestation signal in a coupled integrated assessment–earth system model and the implications for CMIP5 RCP simulations. *Biogeosciences*, *11*(22), 6435-6450.

Don, A., Schumacher, J., & Freibauer, A. (2011). Impact of tropical land-use change on soil organic carbon stocks–a meta-analysis. *Global Change Biology*, *17*(4), 1658-1670.

El-Masri, B., Barman, R., Meiyappan, P., Song, Y., Liang, M., & Jain, A. K. (2013). Carbon dynamics in the Amazonian Basin: Integration of eddy covariance and ecophysiological data with a land surface model. Agricultural and Forest Meteorology, 182, 156-167.

Exbrayat, J. F., Pitman, A. J., Zhang, Q., Abramowitz, G., & Wang, Y. P. (2013). Examining soil carbon uncertainty in a global model: response of microbial decomposition to temperature, moisture and nutrient limitation. Biogeosciences, 10(11), 7095-7108.

Fang, J., Guo, Z., Hu, H., Kato, T., Muraoka, H. and Son, Y. (2014), Forest biomass carbon sinks in East Asia, with special reference to the relative contributions of forest expansion and forest growth. Global Change Biology. doi: 10.1111/gcb.12512

FAO (2006) Global Forest Resources Assessment 2005 . FAO forestry Paper 147, Rome.

FAO (2010). Global Forest Resources Assessment 2010. FAO forestry Paper 13, Rome.

Finzi, A. C., Moore, D. J., DeLucia, E. H., Lichter, J., Hofmockel, K. S., Jackson, R. B., ... & Schlesinger, W. H. (2006). Progressive nitrogen limitation of ecosystem processes under elevated CO2 in a warm-temperate forest. *Ecology*, *87*(1), 15-25.

Finzi, A. C., Norby, R. J., Calfapietra, C., Gallet-Budynek, A., Gielen, B., Holmes, W. E., ... & Ceulemans, R. (2007). Increases in nitrogen uptake rather than nitrogen-use efficiency

support higher rates of temperate forest productivity under elevated CO2. *Proceedings of the National Academy of Sciences*, *104*(35), 14014-14019.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sensing of Environment, 114(1), 168-182.

Friedlingstein, P., & Prentice, I. C. (2010). Carbon–climate feedbacks: a review of model and observation based estimates. *Current Opinion in Environmental Sustainability*, *2*(4), 251-257.

Fuchs, R., Herold, M., Verburg, P., Clevers, J., & Eberle, J. 2014. Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. Global Change Biology. doi: 10.1111/gcb.12714.

Gerber, S., Hedin, L. O., Oppenheimer, M., Pacala, S. W., & Shevliakova, E. (2010). Nitrogen cycling and feedbacks in a global dynamic land model. Global Biogeochemical Cycles, 24(1).

Goll, D. S., Brovkin, V., Parida, B. R., Reick, C. H., Kattge, J., Reich, P. B., ... & Niinemets, Ü. (2012). Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. Biogeosciences, 9, 3547-3569.

Guo, L. B., & Gifford, R. M. (2002). Soil carbon stocks and land use change: a meta analysis. Global change biology, 8(4), 345-360.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. International Journal of Climatology, 34(3), 623-642.

Hanson, P. J., Amthor, J. S., Wullschleger, S. D., Wilson, K. B., Grant, R. F., Hartley, A., ... & Cushman, R. M. (2004). Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. *Ecological Monographs*, *74*(3), 443-489.

Hayes, D. J., McGuire, A. D., Kicklighter, D. W., Gurney, K. R., Burnside, T. J., & Melillo, J. M. (2011). Is the northern high-latitude land-based CO2 sink weakening?. Global Biogeochemical Cycles, 25(3).

Herbert, D. A., Williams, M., & Rastetter, E. B. (2003). A model analysis of N and P limitation on carbon accumulation in Amazonian secondary forest after alternate land-use abandonment. *Biogeochemistry*, *65*(1), 121-150.

Houghton, R. A., & Goodale, C. L. (2004). Effects of Land-Use Change on the Carbon Balance of Terrestrial Ecosystems. *Ecosystems and land use change*, 85-98.

Houghton, R.A., and J.L. Hackler. (2001). *Carbon Flux to the Atmosphere From Land-use Changes: 1850 to 1990*. NDP-050/R1, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

Houghton RA (2003) Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. Tellus B, 55, 378–390.

Houghton RA (2008) Carbon flux to the atmosphere from land-use changes: 1850-2005. In: TRENDS: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Houghton, R. A. (2013a). Keeping management effects separate from environmental effects in terrestrial carbon accounting. Global change biology, 19(9), 2609-2612.

Houghton, R. A. (2013b). The emissions of carbon from deforestation and degradation in the tropics: past trends and future potential. Carbon Management, 4(5), 539-546.

Houghton, R. A., House, J. I., Pongratz, J., Werf, G. R., DeFries, R. S., Hansen, M. C., ... & Ramankutty, nitrogen. (2012). Carbon emissions from land use and land-cover change. Biogeosciences, 9(12), 5125-5142.

Houlton, B. Z., Wang, Y. P., Vitousek, P. M., & Field, C. B. (2008). A unifying framework for dinitrogen fixation in the terrestrial biosphere. *Nature*, *454*(7202), 327-330.

Hungate, B. A., Dukes, J. S., Shaw, M. R., Luo, Y., & Field, C. B. (2003). Nitrogen and climate change. *Science*, *302*(5650), 1512-1513.

Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., ... & Cook, R. (2012). North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, *232*, 144-157.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., ... & Wang, Y. P. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change, 109(1-2), 117-161.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex

and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Ito, A., Penner, J. E., Prather, M. J., De Campos, C. P., Houghton, R. A., Kato, T., ... & Price, D. T. (2008). Can we reconcile differences in estimates of carbon fluxes from land-use change and forestry for the 1990s?. *Atmospheric Chemistry and Physics*, *8*(12), 3291-3310.

Jain AK, Yang X, Kheshgi H, McGuire AD, Post W, Kicklighter D (2009) Nitrogen attenuation of terrestrial carbon cycle response to global environmental factors. Global Biogeochem.Cycles, 23, GB4028, doi:10.1029/2009GB003519.

Jain, A. K., Meiyappan, P., Song, Y. and House, J. I. (2013). CO2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data. Global Change Biology, 19: 2893–2906.

Jain, AK, and Yang, X (2005) Modeling the effects of two different land-cover change data sets on the carbon stocks of plants and soils in concert with CO2 and climate change. Global Biogeochem.Cycles, 19(2), 1-20.

Jones, C., Robertson, E., Arora, V., Friedlingstein, P., Shevliakova, E., Bopp, L., ... & Tjiputra, J. (2013). Twenty-first-century compatible CO2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways. Journal of Climate, 26(13), 4398-4413.

Kato, E., Kinoshita, T., Ito, A., Kawamiya, M., & Yamagata, Y. (2013). Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model. Journal of Land Use Science, 8(1), 104-122.

Kindermann, G. E., McCallum, I., Fritz, S., & Obersteiner, M. (2008). A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*, *42*(3), 387.

Klein Goldewijk, K., & Verburg, P. H. (2013). Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set. Landscape ecology, 28(5), 861-877.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography, 20(1), 73-86.

Knutti, R., & Sedláček, J. (2012). Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change, 3, 369-373.

Kumar, S., Dirmeyer, P. A., Merwade, V., DelSole, T., Adams, J. M., & Niyogi, D. (2013). Land use/cover change impacts in CMIP5 climate simulations–a new methodology and 21st century challenges. Journal of Geophysical Research: Atmospheres, 118, 6337–6353.

Lamarque, J. F., Kyle, G. P., Meinshausen, M., Riahi, K., Smith, S. J., van Vuuren, D. P., ... & Vitt, F. (2011). Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways. Climatic change, 109(1-2), 191-212.

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., ... & Thornton, P. E. (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. Journal of Climate, 25(9), 3071-3095.

LeBauer, D. S., & Treseder, K. K. (2008). Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, *89*(2), 371-379.

Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Arneth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chini, L. P., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A. K., Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, T., Peng, S., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Séférian, R., Segschneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Werf, G. R., Viovy, N., Wang, Y.-P., Wanninkhof, R., Wiltshire, A., and Zeng, N.: Global carbon budget 2014, Earth Syst. Sci. Data, 7, 47-85, doi:10.5194/essd-7-47-2015, 2015.

Loveland, T. R. and Belward, A. S. (1997). The IGBP-DIS Global 1 km Land Cover Data Set, DISCover First Results. International Journal of Remote Sensing, v. 18, no. 5, p. 3,289-3,295.

Luo, Y., Hui, D., & Zhang, D. (2006). Elevated CO2 stimulates net accumulations of carbon and nitrogen in land ecosystems: a meta-analysis. Ecology, 87(1), 53-63.

Luo, Y., Su, B. O., Currie, W. S., Dukes, J. S., Finzi, A., Hartwig, U., ... & Field, C. B. (2004). Progressive nitrogen limitation of ecosystem responses to rising atmospheric carbon dioxide. Bioscience, 54(8), 731-739.

Mahmood, R., Pielke, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., ... & Fall, S. (2013). Land cover changes and their biogeophysical effects on climate. International Journal of Climatology. doi: 10.1002/joc.3736

Mathers, N. J., Harms, B., & Dalal, R. C. (2006). Impacts of land-use change on nitrogen status and mineralization in the Mulga Lands of Southern Queensland. *Austral ecology*, *31*(6), 708-718.

McGuire AD, Stitch, S, Clein, JS et al. (2001) Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO2. climate and land use effects with four process-based ecosystem models, Global Biogeochemical Cycles, 15, 183-206.

McGuire, A. D., Chapin Iii, F. S., Wirth, C., Apps, M., Bhatti, J., Callaghan, T., ... & Vaganov, E. (2007). Responses of high latitude ecosystems to global change: Potential consequences for the climate system. In *Terrestrial Ecosystems in a Changing World* (pp. 297-310). Springer Berlin Heidelberg.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., ... & van Vuuren, D. P. P. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Climatic Change, 109(1-2), 213-241.

Meiyappan P and Jain, AK (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years. Front. Earth Sci., 6(2): 122–139, DOI 10.1007/s11707-012-0314-2.

Meiyappan, P., Dalton, M., O'Neill, B. C., and Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. Ecological Modelling. 291: 152-174.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., ... & Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. Nature, 463(7282), 747-756.

Murty, D., Kirschbaum, M. U., Mcmurtrie, R. E., & Mcgilvray, H. (2002). Does conversion of forest to agricultural land change soil carbon and nitrogen? A review of the literature. Global Change Biology, 8(2), 105-123.

Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E., & McMurtrie, R. E. (2010). CO2 enhancement of forest productivity constrained by limited nitrogen availability. Proceedings of the National Academy of Sciences, 107(45), 19368-19373.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., ... & Yang, Z. L. (2013). Technical description of version 4.5 of the Community Land Model (CLM). Ncar Tech. Note NCAR/TN-503+ STR. National Center for Atmospheric Research, Boulder, CO, 422 pp. doi: 10.5065/D6RR1W7M.

Oren, R., Ellsworth, D. S., Johnsen, K. H., Phillips, N., Ewers, B. E., Maier, C., ... & Katul, G. G. (2001). Soil fertility limits carbon sequestration by forest ecosystems in a CO2-enriched atmosphere. *Nature*, *411*(6836), 469-472.

Pastor, J., & Post, W. M. (1985). *Development of a linked forest productivity-soil process model* (No. ORNL/TM-9519). Oak Ridge National Lab., TN (USA).

Pitman, A. J., de Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., ... & Voldoire, A. (2009). Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. Geophysical Research Letters, 36(14).

Pongratz, J., Reick, C. H., Raddatz, T., & Claussen, M. (2009). Effects of anthropogenic land cover change on the carbon cycle of the last millennium. *Global Biogeochemical Cycles*, *23*(4).

Pongratz, J. (2013). Climate science: Plant a tree, but tend it well. Nature, 498(7452), 47-48.

Pongratz, J., Reick, C. H., Houghton, R. A., & House, J. I. (2014). Terminology as a key uncertainty in net land use and land cover change carbon flux estimates. Earth Syst. Dynam., 5, 177-195, doi:10.5194/esd-5-177-2014.

Post, W. M., & Kwon, K. C. (2000). Soil carbon sequestration and land-use change: processes and potential. Global change biology, 6(3), 317-327.

Ramankutty, N. (2012). Global Cropland and Pasture Data from 1700 - 2007. Available online at (http://www.geog.mcgill.ca/~nramankutty/Datasets/Datasets.html) from the LUGE (Land Use and the Global Environment) Laboratory, Department of Geography, McGill University, Montreal, Quebec, Canada. [Accessed on 15 May 2013].

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), 997-1027.

Reich, P. B., Hungate, B. A., & Luo, Y. (2006). Carbon-nitrogen interactions in terrestrial ecosystems in response to rising atmospheric carbon dioxide. *Annual Review of Ecology, Evolution, and Systematics*, 611-636.

Richter, Daniel deB, Jr., & Houghton, R. A. (2011). Gross CO2 fluxes from land-use change: implications for reducing global emissions and increasing sinks. *Carbon Management*, *2*(1), 41-47.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, *54*(6), 547-560.

Schimel, D., Stephens, B. B., & Fisher, J. B. (2015). Effect of increasing CO2 on the terrestrial carbon cycle. *Proceedings of the National Academy of Sciences*, *112*(2), 436-441.

Schimel, D. S., Braswell, B. H., McKeown, R., Ojima, D. S., Parton, W. J., & Pulliam, W. (1996). Climate and nitrogen controls on the geography and timescales of terrestrial biogeochemical cycling. *Global Biogeochemical Cycles*, *10*(4), 677-692.

Schipper, L. A., Baisden, W. T., Parfitt, R. L., Ross, C., Claydon, J. J., & Arnold, G. (2007). Large losses of soil C and N from soil profiles under pasture in New Zealand during the past 20 years. *Global Change Biology*, *13*(6), 1138-1144.

Scholes, R. J., & Archer, S. R. (1997). Tree-grass interactions in savannas 1. Annual review of Ecology and Systematics, 28(1), 517-544.

Shevliakova, E., Stouffer, R. J., Malyshev, S., Krasting, J. P., Hurtt, G. C., & Pacala, S. W. (2013). Historical warming reduced due to enhanced land carbon uptake. Proceedings of the National Academy of Sciences, 201314047.

Stocker, B. D., Feissli, F., Strassmann, K. M., Spahni, R., & Joos, F. (2014). Past and future carbon fluxes from land use change, shifting cultivation and wood harvest. *Tellus B*, *66*, 23188.

Strassmann, K. M., Joos, F., & Fischer, G. (2008). Simulating effects of land use changes on carbon fluxes: past contributions to atmospheric CO2 increases and future commitments due to losses of terrestrial sink capacity. *Tellus B*, *60*(4), 583-603.

Sullivan, B. W., Smith, W. K., Townsend, A. R., Nasto, M. K., Reed, S. C., Chazdon, R. L., & Cleveland, C. C. (2014). Spatially robust estimates of biological nitrogen (N) fixation imply substantial human alteration of the tropical N cycle. Proceedings of the National Academy of Sciences, 111(22), 8101-8106.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485-498.

Tian, H., Lu, C., Yang, J., Banger, K., Huntinzger, D. N., Schwalm, C. R., ... & Zeng, N. (2015). Global Patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: current status and future directions. *Global Biogeochemical Cycles*. doi: 10.1002/2014GB005021

Thomas, R. Q., Brookshire, E. N., & Gerber, S. (2015). Nitrogen limitation on land: how can it occur in Earth system models?. Global change biology, 21(5), 1777-1793.

Todd-Brown, K. E. O., Randerson, J. T., Hopkins, F., Arora, V., Hajima, T., Jones, C., ... & Allison, S. D. (2014). Changes in soil organic carbon storage predicted by Earth system models during the 21st century. Biogeosciences, 11(8), 2341-2356.

UNEP (2013). The Emissions Gap Report 2013 - A UNEP Synthesis Report, Nairobi, Kenya: UNEP. Available at http://www.unep.org/emissionsgapreport2013/

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., ... & Rose, S. K. (2011). The representative concentration pathways: an overview. Climatic Change, 109(1-2), 5-31.

Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape ecology*, *24*(9), 1167-1181.

Vitousek, P. M., Porder, S., Houlton, B. Z., & Chadwick, O. A. (2010). Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions. Ecological Applications, 20(1), 5-15.

Vitousek, P. M., Menge, D. N., Reed, S. C., & Cleveland, C. C. (2013). Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), 20130119.

Walker, A. P., Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Asao, S., Hickler, T., ... & Norby, R. J. (2015). Predicting long-term carbon sequestration in response to CO2 enrichment: How and why do current ecosystem models differ?. *Global Biogeochemical Cycles*. DOI: 10.1002/2014GB004995.

Walker, A. P., Hanson, P. J., Jain, A., Lomas, M., Luo, Y., McCarthy, H., ... & Iversen, C. M. (2014). Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO_2 enrichment experiments: model performance at ambient CO_2 concentration.

Wang, Y. P., Zhang, Q., Pitman, A. J., & Dai, Y. (2015). Nitrogen and phosphorous limitation reduces the effects of land use change on land carbon uptake or emission. *Environmental Research Letters*, *10*(1), 014001.

Wang, Y. P., Law, R. M., & Pak, B. (2010). A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. Biogeosciences, 7(7), 2261-2282.

Wang, Y. P., & Houlton, B. Z. (2009). Nitrogen constraints on terrestrial carbon uptake: Implications for the global carbon-climate feedback. Geophysical Research Letters, 36(24).

Wania, R., Meissner, K. J., Eby, M., Arora, V. K., Ross, I., & Weaver, A. J. (2012). Carbon-nitrogen feedbacks in the UVic ESCM. Geoscientific Model Development, 5(5), 1137-1160.

Wårlind, D., Smith, B., Hickler, T., & Arneth, A. (2014). Nitrogen feedbacks increase future terrestrial ecosystem carbon uptake in an individual-based dynamic vegetation model. Biogeosciences, 11, 6131-6146.

Wieder, W. R., Cleveland, C. C., Smith, W. K., & Todd-Brown, K. (2015a). Future productivity and carbon storage limited by terrestrial nutrient availability. *Nature Geoscience*. doi:10.1038/ngeo2413

Wieder, W. R., Cleveland, C. C., Lawrence, D. M., & Bonan, G. B. (2015b). Effects of model structural uncertainty on carbon cycle projections: biological nitrogen fixation as a case study. Environmental Research Letters, 10(4), 044016.

Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T., & Reick, C. (2014). Comparing the influence of net and gross anthropogenic land use and land cover changes on the carbon cycle in the MPI-ESM. Biogeosciences, 11, 4817-4828.

Yang X, Richardson TK, and Jain AK (2010) Contributions of secondary forest and nitrogen dynamics to terrestrial carbon uptake Biogeosciences, 7, 3041–3050.

Yang X, Wittig V, Jain A, and Post W (2009) Integration of Nitrogen Dynamics into a Global Terrestrial Ecosystem Model Global Biogeochem. Cycles, 23, GB4028, doi:10.1029/2009GB003519.

Yang, X., Thornton, P. E., Ricciuto, D. M., & Post, W. M. (2013). The role of phosphorus dynamics in tropical forests–a modeling study using CLM-CNP. Biogeosciences, 11, 1667-1681.

Zaehle, S., & Friend, A. D. (2010). Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. Global Biogeochemical Cycles, 24(1).

Zhang, Q., Pitman, A. J., Wang, Y. P., Dai, Y., & Lawrence, P. J. (2013). The impact of nitrogen and phosphorous limitation on the estimated terrestrial carbon balance and warming of land use change over the last 156 yr. Earth System Dynamics, 4, 333-345.

Zhao, M., & Running, S. W. (2010). Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. *science*, *329*(5994), 940-943.

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote sensing of Environment*, *95*(2), 164-176.

CHAPTER 5

Spatial modeling of agricultural land use change at global scale

5.1 Abstract

Long-term modeling of agricultural land use is central in global scale assessments of climate change, food security, biodiversity, and climate adaptation and mitigation policies. We present a global-scale dynamic land use allocation model and show that it can reproduce the broad spatial features of the past 100 years of evolution of cropland and pastureland patterns. The modeling approach integrates economic theory, observed land use history, and data on both socioeconomic and biophysical determinants of land use change, and estimates relationships using long-term historical data, thereby making it suitable for long-term projections. The underlying economic motivation is maximization of expected profits by landowners within each grid cell. The model predicts fractional land use for cropland and pastureland within each grid cell based on socioeconomic and biophysical driving factors that change with time. The model explicitly incorporates the following key features: (1) land use competition, (2) spatial heterogeneity in the nature of driving factors across geographic regions, (3) spatial heterogeneity in the relative importance of driving factors and previous land use patterns in determining land use allocation, and (4) spatial and temporal autocorrelation in land use patterns.

We show that land use allocation approaches based solely on previous land use history (but disregarding the impact of driving factors), or those based on mechanistically fitting models for the spatial processes of land use change do not reproduce well long-term historical land use patterns. With an example application to the terrestrial carbon cycle, we show that such inaccuracies in land use allocation can translate into significant implications for global environmental assessments. The modeling approach and its evaluation provide an example that can be useful to the land use, Integrated Assessment, and the Earth system modeling communities.

5.2 Introduction

Changes in land use are driven by non-linear interactions between socioeconomic conditions (e.g. population, technology, and economy), biophysical characteristics of the land

(e.g. soil, topography, and climate), and land use history (Lambin et al., 2001, 2003). The spatial heterogeneity in driving factors has led to spatially distinct land use patterns. Land use change models exploit techniques to understand the spatial relationship between historical changes in land use and its driving factors (or proxies for them). Such models are also used to project spatial changes in land use based on scenarios of changes in its drivers. The importance of land use change models is evident from the wide range of existing modeling approaches and applications (see reviews by NRC, 2014; Heistermann et al., 2006; Verburg et al., 2004; Parker et al., 2003; Agarwal et al., 2002; Irwin and Geoghegan, 2001; Briassoulis, 2000; U.S. EPA, 2000). However, most land use change models are designed for local to regional scale studies (typically sub-national to national level); global-scale modeling approaches are scarce (Rounsevell and Arneth, 2011; Heistermann et al., 2006).

Global-scale land use modeling is challenging compared to smaller-scale approaches for three main reasons. First, the set of driving factors and their spatial characteristics of change are diverse across the globe, and models need to represent this variability (van Asselen and Verburg, 2012). Second, the various factors that affect land use decisions operate at different spatial scales. For example, landowners make decisions at local scale, whereas factors like governance, institutions, and enforcement of property rights operate at much larger scales. Ideally, global-scale models should incorporate the effects of driving factors at multiple scales (Rounsevell et al., 2014; Heistermann et al., 2006). However, an integrated understanding of how the multi-scale drivers combine to cause land use change is far from complete (Lambin et al., 2001; Meyfroidt, 2012). Third, spatially and temporally consistent data for many important driving factors (e.g. market influence) are not readily available at a global scale and at the required spatial resolution (Verburg et al., 2011, 2013).

Despite these challenges, there are three reasons for modeling land use at a global scale. First, several key drivers of land use (e.g. climate) and their impacts on land use have no regional demarcations and substantial feedback exists between them (Rounsevell et al., 2014). Addressing the feedback between land use and socioecological systems requires a globally consistent framework. Second, regions across the world are interconnected through global markets and trade that can shift supply responses to demands for land across geopolitical regions (Meyfroidt et al., 2013). Modeling such complex interactions among economies demands a global scale approach. Third, the aggregate consequences of land use at the global scale have significant

consequences for climate change (Pielke et al., 2011), global biogeochemical cycles (Jain et al., 2013), water resources (Bennett et al., 2001) and biodiversity (Phalan et al., 2011), making global land use modeling a useful component of analyses of these issues.

These reasons have motivated global scale assessments using Integrated Assessment Models (IAMs) that seek to treat the interactions between land and other socioecological systems in a fully coupled manner (Sarofim and Rielly, 2011). In IAMs, socioeconomic models are coupled with biophysical models (process-based vegetation models and/or climate models) to translate socioeconomic scenarios into changes in land cover and its impacts on environmental variables of interest (van Vuuren et al., 2012). IAMs typically disaggregate the world into 14-24 regions (van Vuuren et al., 2011), and land use decisions are made at this regional scale. Some IAMs have spatially explicit biophysical components, and in these cases land use information on geographic grids at a much higher spatial resolution is required (typically 0.5° x 0.5° lat/lon). To provide this information, spatial land use allocation approaches are employed to downscale aggregate land demands for large world regions to individual grid cells. Examples of such global scale land use allocation approaches can be found in the Global Forest Model (Rokityanskiy et al., 2007), IMAGE (Bouwman et al., 2006), MagPie (Lotze-Campen et al., 2010), KLUM (Ronneberger et al., 2005, 2009), MIT-IGSM (Reilly et al., 2012; Wang, 2008), GLOBIO3 (Alkemade et al., 2009), GLOBIOM (Havlik et al., 2011), Nexus land use model (Souty et al., 2012, 2013), and the Global Land use Model (GLM) (Hurtt et al., 2011).

In this article, we develop a new global land use allocation model specifically to downscale agricultural (cropland and pastureland) land use from large world regions to the grid cell level. Agricultural land use merits special attention because it is associated with the majority of land use-related environmental consequences (Green et al., 2005), currently occupying ~40% of Earth's land area (Foley et al., 2005). There are two novel features of our approach that distinguish it from previous approaches.

First, our model predicts fractional land use within each grid cell (continuous field approach) driven by time-varying socioeconomic and biophysical factors. In contrast, most existing models do one or the other but not both. For example, many downscaling methods represent land use in each grid cell (0.5° x 0.5° lat/lon or coarser) by the dominant land cover category (e.g. MagPie, IMAGE, GLOBIOM, and the Nexus land use model). This simplified representation in land cover underestimates land cover heterogeneity and is a major source of

uncertainty in impact assessments (Verburg et al., 2013). Some recent efforts (e.g. Letourneau et al., 2012; Schaldach et al., 2011) have addressed this problem by increasing spatial resolution, for example using 5-minute grid cells that represent dominant land cover types. While such approaches are an improvement, they are also much more computationally intensive and do not escape the problem that for many variables representing land use drivers, high resolution data at the global scale are unavailable (Verburg et al., 2013). In other approaches (e.g., GLOBIO3 and GLM) land cover is represented as fractional units within each grid cell (again 0.5° x 0.5° lat/lon), but the approach to allocation is overly simplified, proportionally allocating land use projections for aggregate regions to grid cells as closely as possible to existing land use patterns. Such an approach does not account for the effect of changes over time in land use drivers, which can lead to land use projections that are inconsistent with those drivers (as will be shown later).

Second, we carry out the first global scale evaluation of a spatial land use allocation model over a long historical period (>100 years), reproducing the broad spatial features of the long-term evolution of agricultural land use patterns. Evaluation of global-scale spatial land use models is important because they are used to generate scenarios for 50-100 years into the future, for example to explore issues related to greenhouse gas emissions and mitigation possibilities (Moss et al., 2010; Kindermann et al., 2008), climate change impacts on ecosystems (MEA, 2005; UNEP, 2012), biodiversity (TEEB, 2010; Pereira et al., 2010), or adaptation options involving land use (OECD, 2012; Phalan et al., 2011). While evaluation of model performance over the past 100 years is no guarantee of good performance over the next 100 years, demonstrating the ability of a model to reproduce long-term historical patterns increases confidence in its suitability for application to long-term scenarios of future change. The model evaluation presented here could serve as an example for how evaluation of other downscaling methodologies could be carried out (O'Neill and Verburg, 2012; Hibbard et al., 2010).

5.3 Methods and data

Overview of the approach

Our land use allocation model simulates the spatial and temporal development of cropland and pastureland at a spatial resolution of 0.5° x 0.5° lat/lon and at an annual time-step.

The model operates at two different spatial levels. On the regional level, the aggregate regional demand for cropland and pastureland is provided as input to the model. The model then allocates this demand to individual grid cells within that region. We use a constrained optimization technique to allocate a fraction of each grid cell to cropland and pastureland while meeting the aggregate regional demand for each type of land. The optimization technique selects the most profitable land to grow crops and pasture based on (1) the suitability of each grid cell for crop or pasture production, determined by a set of 46 biophysical and socioeconomic factors (Table 5.1), (2) historical land use patterns (temporal autocorrelation) and (3) the land use predicted for neighboring grid cells (spatial autocorrelation).

A primary intended application of this model is as one component of a larger modeling framework that includes a global, regionally resolved economic model that generates scenarios of future demand for land at the regional level, similar to the approach taken in other IAMs or land use models as discussed above. However, the main aim of this paper is to present and evaluate our model in a historical simulation against 20th century gridded data of cropland and pastureland. Ideally, the model should be evaluated against observational data. However, purely observational data for global, spatially resolved land use data do not exist. Rather, existing gridded land use reconstructions are modeled estimates that draw on national and sub-national data to the extent possible (see Appendix B). For practical purposes, we assume existing land use reconstructions represent the "truth" for the purpose of our model calibration and evaluation. In this respect, we face the same limitations as all global, spatially resolved models of land use, which typically use such reconstructions as maps representing current land use as a basis for future projections (e.g., Hurtt et al., 2011). However, it must be kept in mind that such reconstructions are estimates, and that estimates can and do differ from one another based on different uses of the underlying data and methods for estimation of gridded outcomes.

We break down the overall approach discussed above into three components for explanatory purpose. First, we formulate a land use allocation model based on profit maximization using mathematical programming methods for constrained optimization with respect to spatial and temporal distribution of land use types. Second, we derive an estimation procedure for the unknown parameters in the land use allocation model. The estimation procedure accounts for the heterogeneous nature and importance of driving factors across

geographic regions. Finally, we evaluate the land use allocation model and estimation procedure using a historical global cropland and pastureland dataset.

The land use allocation model

Theoretical framework

The economic motivation for the land use allocation model is maximization of expected profit by hypothesized landowners within each grid cell (Lubowski et al., 2008). This motivation is consistent with the structure of most IAMs at the regional scale, which generally assume some form of optimization for economic sectors (e.g. profit maximization, cost minimization) that generate aggregate demand for land. We formulate the land use allocation model as a dynamic profit maximization function that consists of two components: a static profit maximization function and a dynamic adjustment cost model. The static profit maximization function maximizes the achievable profit within each grid cell by selecting the most productive land for growing crops and pastures. The dynamic adjustment cost model accounts for the adjustment cost associated with changes in land use patterns over time (Golub et al., 2008). For example, expanding cropland into unmanaged ecosystems would entail some cost to clear unmanaged land and build roads and other infrastructure. This adjustment cost tends to create inertia in land use patterns over time.

For ease of understanding, we introduce the model component-wise. The components are then eventually combined to form the final land use allocation model. In this description of the theoretical basis of the model, we differentiate two broad categories of land: managed and unmanaged. Managed land is the sum of cropland and pastureland, and we define all other land types as unmanaged. Because our focus is on cropland and pastureland, all equations we describe in this (and the next) section refer to managed land area. We account for unmanaged land later in the estimation procedure.

The static profit function for each grid cell 'g' is expressed as:

$$\underset{\{Y_{\mathrm{lg}}^{t}\}}{Maximize} \sum_{l=1}^{2}\left(P_{\mathrm{lg}}^{t}-W_{\mathrm{lg}}^{t}\right)Y_{\mathrm{lg}}^{t}-R_{\mathrm{lg}}\left(Y_{\mathrm{lg}}^{t}\right)^{2} \qquad (1)$$

In Eq. (1), Y_{lg}^{t} represents the area to be estimated of land use type 'l' (=1 for cropland, and 2 for pastureland) in grid cell 'g' at time step 't'. P_{lg}^{t} denotes the price per unit area for commodities produced by land use activity 'l' and W_{lg}^{t} represents the cost per unit area for producing those commodities. The linear term $\left(P_{lg}^{t} - W_{lg}^{t}\right)Y_{lg}^{t}$ represents the 'net profit' for each grid cell, which can be thought of as a measure of land suitability for land use activity 'l'. The second term $-R_{lg}\left(Y_{lg}^{t}\right)^{2}$ represents the non-linear cost associated with decreasing returns to scale; i.e. output increases less than proportionately to an increase in inputs (land use area) and the rate of increase in output decreases progressively with additional inputs. The non-linear cost term is included because land profitability is assumed to vary within each grid cell, and the most profitable land is used first. Therefore, in the long run, the profitability of each additional hectare of land brought under production within a grid cell declines (Gouel and Hertel, 2006). $R_{lg}\left(>0\right)$ is a productivity/returns constant and is a function of land use type 'l' and location 'g'.

Eq. (1) is considered a 'static' profit function because all variables are based on the current time step 't' with no reference to history. Eq. (1) can be simplified as a quadratic function (see Appendix A.1 for detailed steps):

$$\underset{\{Y_{lg}^{t}\}}{Maximize}\sum_{l=1}^{2}\left(-R_{lg}\right)\left(Y_{lg}^{t} - d_{lg}S_{lg}^{t}\right)^{2} \tag{2}$$

In Eq. (2), $d_{lg} = \frac{1}{2R_{lg}}$ is a constant for decreasing returns to scale and S_{lg}^{t} (suitability) equals the net price term $P_{lg}^{t} - W_{lg}^{t}$. Eq. (2) is subject to two constraints:

$$Y_{lg}^{t} \geq 0 \tag{3}$$

$$\sum_{l=1}^{2} Y_{lg}^{t} \leq GA_{g} \tag{4}$$

Eq. (3) avoids negative allocations. Eq. (4) implies the total of cropland and pastureland area allocated within each grid cell 'g' should not exceed the grid cell area GA_{g}.

We formulate a dynamic adjustment cost model for each grid cell 'g' as follows:

$$\underset{\{Y_{\lg}^{t}\}}{Minimize} \sum_{l=1}^{2} Q_{\lg}\left(Y_{\lg}^{t} - Y_{\lg}^{\bar{t}}\right)^{2} \tag{5}$$

Eq. (5) is also constrained by Eqs. (3)-(4). Eq. (5) represents a constrained least-squares optimization that tends to minimize the adjustment cost by minimizing the changes in land use allocation between the current and a previous time step $\bar{t}$ $(\bar{t} < t)$. Criteria for selecting the value of $\bar{t}$ are explained in a subsequent sub-section. $Q_{\lg}(>0)$is a constant that indicates the adjustment cost per unit area and is a function of land use type 'l' and location 'g'. In Eq. (5) we assume an exponent of 2 because: (1) our land use allocation method is based on quadratic programming, and (2) our model parameter estimation involves differentiating the quadratic program, which results in linear equations that are convenient to solve.

We combine Eq. (2) and Eq. (5), to write the overall objective function as a minimization problem for each grid cell 'g':

$$\underset{\{Y_{\lg}^{t}\}}{Minimize} \sum_{l=1}^{2} \left[Q_{\lg}\left(\frac{Y_{\lg}^{t}}{GA_{g}} - \frac{Y_{\lg}^{\bar{t}}}{GA_{g}} \right)^{2} + R_{\lg}\left(\frac{Y_{\lg}^{t}}{GA_{g}} - d_{\lg} \frac{S_{\lg}^{t}}{GA_{g}} \right)^{2} \right] \tag{6}$$

For convenience we have divided Eq. (6) by a constant that equals the square of grid cell area. The minimization problem is unaffected by this modification. Later, it will become evident that treating variables as fractions instead of areas is convenient in the parameter estimation procedure.

Our aim is to allocate the aggregate land demand among the grid cells within a given region such that the total profits are maximized. Therefore, we stack the individual grid cell level optimizations (Eq. (6)) over the aggregate region and write in matrix notation:

$$\underset{\{Y^{t}\}}{Minimize} \left(Y^{t} - Y^{\bar{t}}\right)' \mathbf{A} \left(Y^{t} - Y^{\bar{t}}\right) + \left(Y^{t} - \mathbf{D}S^{t}\right)' \left(Y^{t} - \mathbf{D}S^{t}\right) \tag{7}$$

In Eq. (7), primes denote the matrix transpose operator. Y^{t} is a column vector of size '$2N \times 1$' with elements $\frac{Y_{\lg}^{t}}{GA_{g}}$, i.e. $Y^{t} = \left[\begin{array}{cccc} \frac{Y_{l1}^{t}}{GA_{1}} & \frac{Y_{21}^{t}}{GA_{1}} & \ldots & \frac{Y_{lN}^{t}}{GA_{N}} \quad \frac{Y_{21}^{t}}{GA_{N}} \end{array} \right]'$, where '$N$' represents the total number of grid cells within the aggregate region. Therefore, elements in vector Y^{t} are

normalized by the grid cell area and will therefore range from zero to one. Similarly $Y^{\bar{t}}$, d, and S^t are vectors of $\frac{Y_{\mathrm{lg}}^{\bar{t}}}{GA_g}$, d_{lg}, and $\frac{S_{\mathrm{lg}}^{t}}{GA_g}$ respectively. The term $\mathbf{D}$ is a diagonal matrix of size $'2N \times 2N'$ given by

$$\mathbf{D} = d * \mathbf{I}_{2N \times 2N}$$

where $\mathbf{I}$ is an identity matrix. The term $\mathbf{A}$ represents a constant diagonal matrix.

$$\mathbf{A} = \begin{bmatrix} \frac{Q_{11}}{R_{11}} & 0 & \dots & \dots & 0 \\ 0 & \frac{Q_{21}}{R_{21}} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{Q_{1N}}{R_{1N}} & 0 \\ 0 & \dots & \dots & 0 & \frac{Q_{2N}}{R_{2N}} \end{bmatrix}_{2N \times 2N} .$$

The matrix $\mathbf{A}$ can be usefully interpreted as representing the balance between the importance of adjustment costs and of land suitability in determining land allocation across grid cells. Eq. (7) can be regarded as a balance between a dynamic (time-series aspect) and a static (cross-sectional aspect) term. The dynamic term is implicitly minimizing adjustment costs by trying to keep land use similar to the historical (already existing) land use patterns, whereas the static term selects the most suitable land to maximize the net profit regardless of history. The balance between the static and dynamic term is determined by the values of the matrix $\mathbf{A}$. In extreme cases, when R_{lg} is zero, no explicit account is taken of the relative suitability of land across grid cells and outcomes are determined entirely by land use history; when Q_{lg} is zero, no account is taken of past land use patterns and outcomes are determined entirely by suitability across grid cells.

We simplify the matrix $\mathbf{A}$ by assuming the diagonal elements are equal (i.e. $\frac{Q_{\mathrm{lg}}}{R_{\mathrm{lg}}}$ is same for all $'l'$ and $'g'$). Hence $\mathbf{A} = a\mathbf{I}$, where $'a'$ is a positive scalar $\left(= \frac{Q_{\mathrm{lg}}}{R_{\mathrm{lg}}} \right)$ and $\mathbf{I}$ is an identity matrix of size $'2N \times 2N'$. Substantively, this implies that while adjustment costs and land

suitability can vary across grid cells, their relative importance to land allocation decisions is held fixed across grid cells within a given region. This is not an unreasonable assumption and a minor concession given its practical benefits: it is both unrealistic and undesirable to estimate $\frac{Q_{lg}}{R_{lg}}$ for each 'l' and 'g'. It is unrealistic because the number of unknown parameters will increase with the number of grid cells resulting in the incidental parameters problem (see Lancaster, 2000). It is undesirable because the historical data for land use and its driving factors available to constrain the model is limited by both availability and grid level accuracy (see section 2.6 and the appendix information cited therein).

Eq. (7) is quadratic in the Y^t vector and is subject to two grid cell area constraints (Eqs. (8)-(9)) that are the vector forms of Eqs. (3)-(4) respectively.

$$Y^t \geq \mathbf{0} \tag{8}$$

$$\left(Y^t_{2g-1} + Y^t_{2g}\right) \leq 1 \quad \forall g \in [1, N] \tag{9}$$

$$\sum_{g=1}^{N} GA_g Y^t_{2g-1} = \text{Regional area demand for cropland} \tag{10}$$

$$\sum_{g=1}^{N} GA_g Y^t_{2g} = \text{Regional area demand for pastureland} \tag{11}$$

Eq. (7) is also subject to regional scale constraints (Eqs. (10)-(11)) that ensure that aggregate regional demand for each land use activity is equal to the total grid cell allocations of that land use activity within that region. Therefore, Eq. (7) is a quadratic program. Competition between land use types is accounted for in Eq. (7) because the profits within each grid cell are maximized by simultaneously weighing both the cropland and pastureland benefits. The first term (adjustment cost) in Eq. (7) accounts for temporal autocorrelation in land use datasets. In the following section, we update Eq. (7) to account for spatial autocorrelation.

Accounting for spatial autocorrelation

Land use area in a grid cell tends to be more similar to the values at surrounding grid cells than to those farther away, a feature known as spatial autocorrelation (Overmars et al., 2003). When spatial autocorrelation is not accounted for, we violate a key assumption in statistical analysis that the residuals are independent and identically distributed (Dormann et al.,

2007). We account for spatial autocorrelation by introducing a spatial weight matrix with the neighborhood size and weighting scheme selected based on trial and error (Augustin et al., 1996). We chose the neighborhood region to be the surrounding eight grid cells (first-order Moore's neighborhood) all with equal weight.

The land use allocation model (Eq. (7)) with a spatial weights matrix **B** for spatial autocorrelation is:

$$\underset{\{Y^t\}}{Minimize}\ \left(Y^t - \bar{Y^t}\right)^{¢}\left(\mathbf{A}+\mathbf{B}\right)\left(Y^t - \bar{Y^t}\right) + \left(Y^t - \mathbf{D}S^t\right)^{¢}\left(Y^t - \mathbf{D}S^t\right) \quad (12)$$

The **B** matrix is proportional to a **W** matrix of spatial weights that is assumed to be symmetric and have zeros along its main diagonal. Note the **A** matrix is diagonal. We represent the spatial weights in **W** matrix by a constant scalar '*b*' that can be positive (if positively correlated), negative (negatively correlated), or zero (uncorrelated). We assume zero spatial autocorrelation between the two land use activities for a practical benefit: the resulting matrix structure allows us to use specialized techniques to perform matrix inversions quickly that are required for estimating model parameters (described in a subsequent sub-section), which otherwise is computationally expensive. We provide some examples in Appendix A.2 to help illustrate the structure of matrices **A** and **B**. For grid cells lying along political boundaries, slight deviations in averaging could arise due to edge effects.

Eq. (12) is our final land use allocation model and is subject to two grid level constraints (Eqs. (8)-(9)) and two regional constrains (Eqs. (10)-(11)). There are four unknown components in Eq. (12) that need to be estimated from historical data: the potential land suitability vector S^t, the scalar constants '*a*' and '*b*', and the constant vector for decreasing returns '*d*'. For consistency, we estimate all the unknown parameters simultaneously using the following procedure.

Estimation method for unknown parameters

Consistent estimates for the parameters in Eq. (12) can be obtained with historical data for land use and its driving factors by treating Eq. (12) as a least-squares problem that combines first-order autoregressive stochastic processes (for first-order spatial autocorrelation and dynamic adjustment costs) and a logit function (with explanatory factors) for the term S^t. A restriction on

the error process for each grid cell insures that the sum of Y^t elements for each grid cell $\left(i.e. \frac{Y_{1g}^t}{GA_g} + \frac{Y_{2g}^t}{GA_g} \right)$ is bounded between zero and one, or they may take a value of zero or one.

Logit function

We assume S^t (dependent variable) to be a function of a matrix $\mathbf{X}_{gt}$ of potential driving factors (exogenous explanatory variables; see table 5.1) that is specific to grid cell 'g' and time 't'. The matrix $\mathbf{X}_{g0}$ (i.e. for t=0) refers to potential driving factors that are time-stationary (e.g. soil and terrain conditions). We model the relationship between the dependent and explanatory variables as a binomial logistic regression (see Lesschen et al. (2005) for regression approaches used in spatial land use models). For each grid cell 'g' and time 't', the logit functions for cropland $(l=1)$ and pastureland $(l=2)$ are given by Eqs. (13)-(14).

$$S_{1g}^t = \frac{1}{1+e^{\beta_0 + \mathbf{X}'_{gt}\beta}} \quad (13)$$

$$S_{2g}^t = \frac{e^{\beta_0 + \mathbf{X}'_{gt}\beta}}{1+e^{\beta_0 + \mathbf{X}'_{gt}\beta}} \quad (14)$$

In Eqs. (13)-(14), β_0 is a constant coefficient and β is a vector of coefficients with a component for each explanatory variable. β_0 and β need to be estimated. The sum of Eqs. (13)-(14) implies the index of land suitability summed for cropland and pastureland for each grid cell equals one (recall that these equations apply only to managed land). Therefore, Eqs. (13)-(14) can be interpreted to partition the total land use area in grid cell 'g' as proportions of cropland and pastureland.

Error process

Formally, the unconstrained version of the minimization problem in Eq. (12) implies a set of first-order necessary conditions. Therefore, we differentiate Eq. (12) and equate to zero:

$$\left(\mathbf{A}+\mathbf{B}\right)\left(Y^t - Y^{\bar{t}}\right) + \left(Y^t - \mathbf{D}S^t\right)' = 0 \Leftrightarrow Y^t = \left(\mathbf{I}+\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{D}S^t + \left(\mathbf{I}+\mathbf{A}+\mathbf{B}\right)^{-1}\left(\mathbf{A}+\mathbf{B}\right)Y^{\bar{t}} \quad (15)$$

In Eq. (15), $\mathbf{I}$ is an identity matrix. Let $\boldsymbol{\Psi}=(\mathbf{I}+\mathbf{A}+\mathbf{B})^{-1}$, $\boldsymbol{\Omega}=\boldsymbol{\Psi}(\mathbf{A}+\mathbf{B})$, and ε_{1g}^{t} denote random variables, each with a mean zero that satisfy $\varepsilon_{1g}^{t}+\varepsilon_{2g}^{t}=0$. Non-linear regression equations associated with Eq. (15) are

$$Y_{1g}^{t}=\sum_{k=1}^{N}\boldsymbol{\Psi}_{2g-1,2k-1}d_{2k-1}\frac{1}{1+e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}+\sum_{k=1}^{N}\boldsymbol{\Omega}_{2g-1,2k-1}Y_{1k}^{\bar{t}}+\varepsilon_{1g}^{t} \tag{16}$$

$$Y_{2g}^{t}=\sum_{k=1}^{N}\boldsymbol{\Psi}_{2g,2k}d_{2k-1}\frac{e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}{1+e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}+\sum_{k=1}^{N}\boldsymbol{\Omega}_{2g,2k}Y_{2k}^{\bar{t}}+\varepsilon_{2g}^{t} \tag{17}$$

However in the above formulation, estimation of the 'd' parameters (which, as noted above, reflect returns to scale) for each grid cell is inconsistent because of incidental parameter bias (Lancaster, 2000). For consistency, we treat the 'd' parameters as random effects. For random effects, a logit function that differentiates managed (crop + pasture) from unmanaged land (e.g. forests, grasslands, and bare land that occupy rest of the grid cell area) is a natural specification that builds on a nested logit structure.

The 'd' parameter corresponding to managed land fraction is:

$$d_g=\frac{1}{1+e^{\gamma_0+\mathbf{X}'_{g0}\gamma}} \tag{18}$$

In Eq. (18), γ_0 is a constant coefficient and γ is a vector of coefficients to be estimated for the set of explanatory variables specified by $\mathbf{X}_{g0}$.

Substituting Eq. (18) into Eqs. (16)-(17) gives Eqs. (19)-(20) respectively.

$$Y_{1g}^{t}=\sum_{k=1}^{N}\boldsymbol{\Psi}_{2g-1,2k-1}\frac{1}{1+e^{\gamma_0+\mathbf{X}'_{2k-1,0}\gamma}}\frac{1}{1+e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}+\sum_{k=1}^{N}\boldsymbol{\Omega}_{2g-1,2k-1}Y_{1k}^{\bar{t}}+\varepsilon_{1g}^{t} \tag{19}$$

$$Y_{2g}^{t}=\sum_{k=1}^{N}\boldsymbol{\Psi}_{2g,2k}\frac{1}{1+e^{\gamma_0+\mathbf{X}'_{2k-1,0}\gamma}}\frac{e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}{1+e^{\beta_0+\mathbf{X}'_{2k-1,t}\beta}}+\sum_{k=1}^{N}\boldsymbol{\Omega}_{2g,2k}Y_{2k}^{\bar{t}}+\varepsilon_{2g}^{t} \tag{20}$$

An important remark is that β and γ are not identified in Eqs. (19)-(20) unless $\mathbf{X}_{g0}=\mathbf{X}_{gt}$ for some 't'. We therefore specify the explanatory variables that are time-stationary (factors such as soil and terrain conditions) within $\mathbf{X}_{g0}$ and the transient explanatory variables (e.g. climate and socioeconomics) within $\mathbf{X}_{gt}$.

In Eqs. (19)-(20), for each grid cell 'g' the logit function for the 'd' parameter accounts for the fraction of managed land area, whereas the logits for 'S_{1g}^{t}' and 'S_{2g}^{t}' further splits the managed land fraction into proportions of cropland and pastureland respectively.

A third equation in this system applies to unmanaged land fractions, which are the random effects:

$$1- \left(Y_{1g}^{t} + Y_{2g}^{t}\right) = \frac{e^{g_0 + \mathbf{X}_{g,0}' g}}{1+e^{g_0 + \mathbf{X}_{g,0}' g}} + h_g \tag{21}$$

Eq. (21) implies that we can deduce the unmanaged land fraction using Eqs. (19)-(20). This is implied from the assumption built into the model's error process such that the sum of managed and unmanaged land fractions adds up to one for each grid cell. Therefore, Eq. (21) is redundant and dropped from the estimation.

In general for a spatial-weights matrix $\mathbf{W}$, a large number of grid cells implies the components of both $\mathbf{\Psi}(a,b)$ and $\mathbf{\Omega}(a,b)$ are high-order rational polynomials in powers of parameters 'a' and 'b'. These are derived from the functions $\mathbf{\Psi}(a,b) = ((a+1)\mathbf{I} + b\mathbf{W})^{-1}$, and $\mathbf{\Omega}(a,b) = \mathbf{\Psi}(a,b)(a\mathbf{I} + b\mathbf{W})$, which are applied to form a stacked system of regression equations where 'S^t' and 'd' parameters are logit functions of a vector of explanatory variables as discussed above. Eq. (19)-(20) can be combined as:

$$Y^t = \mathbf{\Psi}(a,b)\, \mathbf{D}(\mathbf{X}_0'\gamma)\, S^t(\mathbf{X}_t'\beta) + \mathbf{\Omega}(a,b)\, Y^{\bar{t}} + \varepsilon^t \tag{22}$$

In Eq. (22), $\mathbf{D} = d(\mathbf{X}_0'\gamma) * \mathbf{I}_{2N\times 2N}$ where $\mathbf{I}$ is an identity matrix.

Least-square estimation

The least-squares problem for the nonlinear regression to estimate the parameters 'β_0', 'β', 'γ_0','γ', 'a', and 'b' is obtained by minimizing ε^t in Eq. (22).

$$\underset{\{\beta,\gamma,a,b\}}{Minimize} \sum_t \left(Y^t - \left(\mathbf{\Psi}(a,b)\mathbf{D}(\mathbf{X}_0'\gamma)S^t(\mathbf{X}_t'\beta) + \mathbf{\Omega}(a,b)Y^{\bar{t}}\right)\right)' \left(Y^t - \left(\mathbf{\Psi}(a,b)\mathbf{D}(\mathbf{X}_0'\gamma)S^t(\mathbf{X}_t'\beta) + \mathbf{\Omega}(a,b)Y^{\bar{t}}\right)\right) \tag{23}$$

The summation over $'t'$ in Eq. (23) implies multiple years of data can be used to estimate parameters. There are no grid level or regional demand constraints imposed on Eq. (23). The only constraint imposed on Eq. (23) is that $a>0$.

Solving for estimates of $'\beta_0'$, $'\beta'$, $'\gamma_0'$,$'\gamma'$, $'a'$, and $'b'$ using Eq. (23) can be difficult when there are many grid cells, due to the numerical costs of inverting a large matrix to compute $\boldsymbol{\Psi}(a,b)$ for each iteration in the estimation procedure. Therefore, we use specialized techniques to invert the matrix efficiently, the details of which are spelled out in Appendix A.3 with examples.

Accounting for spatial heterogeneity in driving factors

The set of driving factors and their relative importance (i.e. values of $'\beta'$ and $'\gamma'$ for a given explanatory variable) often differ between geographic regions. Further, the strength of temporal and spatial autocorrelation (i.e. $'a'$ and $'b'$ parameters) may vary between geographic regions. To account for this spatial heterogeneity, we disaggregate the world into into 127 distinct sub-regions (Figure 5.1) based on administrative boundaries (see Appendix C for methods and rationale) and solve Eq. (23) separately for each sub-region. Though we had earlier assumed that $'a'$ and $'b'$ parameters do not vary across grid cells within a region, estimating Eq. (23) for the 127 sub-regions imply $'a'$ and $'b'$ parameters can vary across sub-regions.

Selecting lag-year $(\bar{t})$ *associated with the dynamic adjustment cost term*

There are two main considerations for selecting a value for $\bar{t}$:

1. In general, the value of $Y^t - Y^{\bar{t}}$ should not be negligible relative to the value $Y^{\bar{t}}$ for most grid cells. If such were the case, the least-square optimization (Eq. (23)) would tend to be biased towards the dynamic adjustment cost term $\left(R_{\text{lg}} \to 0\right)$. An exception applies to grid cells that are completely unsuitable for both cropland and pastureland where $Y_{\text{lg}}^t = Y_{\text{lg}}^{\bar{t}} = 0$ for any $'t'$. Typically, for global land use change datasets at 0.5° x 0.5° lat/lon, the grid cell level net changes in land use fractions between consecutive years is less than 10^{-3} for both cropland and pastureland when averaged globally over the 20$^{\text{th}}$ century (excluding grid cells unsuitable for agriculture and

computed based on the land use change data). Therefore, a lag on the order of one year ($\bar{t} = t - 1$) is not an appropriate choice.

2. Our specification of random errors in Eqs. (19)-(20) imply that they are uncorrelated with the explanatory variables. For this assumption to be valid, the value of the lag should be sufficiently large.

Based on experimentation, we found a lag of 10 years ($\bar{t} = t - 10$) satisfies the above requirements and also roughly matches the temporal autocorrelation present in the historical land use change dataset.

Explanatory variables and land use change data

We include a total of 46 variables as potential explanatory variables (or proxies for them) in our regression analysis (table 5.1). These variables are restricted to those expected to determine the spatial (as opposed to aggregate) determinants of land use patterns and they broadly align with our existing knowledge of land use dynamics (Lambin et al., 2001, 2003). At the global scale, the factors listed in table 5.1 are adequate to describe the major spatial patterns of agricultural land use (Ramankutty et al., 2002). However, this list is not exhaustive. For example, policies that would influence spatial land use patterns within a region are likely relevant but are not explicitly included here. Rather, their effect (present in historical land use patterns) would be captured only implicitly through proxy variables. In cases like this we also rely on the fact that such factors are incorporated at least at the level of aggregate regions in scenarios generated by IAMs.

We synthesize the information for the 46 explanatory variables from a wide range of sources (table 5.2). The data for each of the explanatory variables was either available for the time period 1900-2005 (annually) at a spatial resolution of 0.5° x 0.5° lat/lon, or was available for shorter time periods and/or coarser resolutions and we extended/refined it to a common time period and resolution. The rationale for selecting these variables and methodologies applied to extend/refine the raw data are detailed in Appendix D.

Historical reconstructions of cropland and pastureland were obtained from Ramankutty (2012) (hereafter referred as RF because it is an updated version of Ramankutty and Foley (1999) data). The reconstruction is available yearly (1700-2007) at 0.5° x 0.5° spatial resolution; we utilize data for the period 1891-2005 for historical model simulation, of which ~20% subset

is used for model estimation as explained below. In principle, the model can be estimated and evaluated with any historical reconstruction available in literature. The rationale for selecting RF data for our study and further background information on the land use change dataset is provided in Appendix B.

To estimate the parameters $'\beta_0'$, $'\beta'$, $'\gamma_0'$,$'\gamma'$, $'a'$, and $'b'$ we select 20 years of RF data for land use $\left(Y^t\right)$ and its explanatory variables $\left(\mathbf{X}_{gt}\right)$ over the 1895-2005 period; i.e. about 20% of the available (annual) data. In selecting particular years to use in estimating parameters, we balance two goals: capturing recent patterns of land use from which future projections will begin, and capturing larger, longer-term changes in land use and explanatory variables to better support use of the model in long-term future projections. We therefore choose two 10-year sets of data. The first set, to capture longer-term changes, consists of 10 years drawn between 1905 and 1995 at 10-year time steps (i.e. 1905, 1915 …1995). For each year, a corresponding 10-year lag data point $\left(Y^{t-10}\right)$is used in the estimation procedure. For example, for year 1905, the lag year data corresponds to 1895, and for 1995 the lag year data corresponds to 1985. The second set, to capture contemporary relationships, includes 10 years of data covering the period 1996-2005 at 1-year time steps. For 1996, the lag year data corresponds to 1986, and for 2005 the lag year data corresponds to 1995.

The explanatory variables (table 5.1) used in the analysis have different units and scales. Hence, the estimated regression coefficients (β and γ vectors) are of different scale and cannot be directly interpreted to infer the relative importance of explanatory variables on the dependent variable. To address this problem, we standardize all explanatory variables covering the period 1901-2005 before the parameter estimation and model simulation procedure. The standardization also prevents numerical difficulties that could arise due to scaling problems in the least-squares estimation (Eq. (23)). A standardized coefficient indicates how many standard deviations a dependent variable will change, per standard deviation increase in the explanatory variable (Hunter and Hamilton, 2002). For each explanatory variable associated with the vectors β and γ, we calculate its mean and standard deviation using five years of data (2001-2005) separately for each of the 127 sub-region. For a given explanatory variable and grid cell, we standardize the variable using the z-score which is computed as the difference between the value of the variable

at that grid cell and its mean value for the corresponding sub-region, divided by the standard deviation corresponding to that sub-region.

See Appendix E for a discussion on how we handle multicollinearity among explanatory variables and excess-zeros problem. Appendix A.4 provides details on the solvers used to implement the land use allocation model (Eq. (12)) and the least-squares optimization (Eq. (23)).

Simulation procedure to evaluate the land use allocation model

To test the land use allocation model, we compared results from model simulations for the historical period (1901-2005) to the historical data reconstruction (RF data) over that period. For this test, first we divided the world into nine regions (Figure 5.1), consistent with the regions used in a general equilibrium model of the global economy, the PET (Population-Environment-Technology) model (O'Neill et al., 2010). This regional mask will allow us to subsequently link the land use allocation model with the PET model for exploring future scenarios. For each year over the period 1901-2005, we aggregated the 0.5°x0.5° lat/lon reconstruction data for cropland and pastureland to these nine regions. This regionally aggregated land use information was then used as input to the land use allocation model (Eq. (12)) to form the annual regional-scale constraint on the total area demand for each land use type (through Eqs. (10)-(11)). Next, the land use allocation model (Eq. (12)) allocated the regionally aggregated land use information back to 0.5°x0.5° spatial resolution by applying time-dependent regional demand constraints and two local constraints (Eqs. (8)-(9)). The model-downscaled land use maps were finally compared to the original 0.5°x0.5° lat/lon RF data to evaluate model performance.

The evaluation test is rigorous given that most IAMs disaggregate the world into a larger number of smaller regions (14-24 regions; van Vuuren et al., 2011). Figure 5.2 depicts our evaluation strategy, which we implement with the algorithm discussed next. This algorithm is repeated separately for each of the nine aggregate world regions.

1. Form the matrices $\mathbf{A}$ and $\mathbf{B}$ to use in Eq. (12). Each of the nine aggregate regions has several sub-regions (Figure 5.1) with corresponding scalar constants $'a'$ and $'b'$. Therefore, each grid cell in matrices $\mathbf{A}$ and $\mathbf{B}$ is weighted based on the parameters $'a'$ and $'b'$ within the sub-region the grid cell belongs to. Matrices $\mathbf{A}$ and $\mathbf{B}$ are static and need to be computed just once at the start of the model simulation.

2. For each grid cell 'g' within the aggregate region, calculate decreasing returns constant d_g from Eq. (18). For this calculation, spatial data on the standardized explanatory variables X_{lg}^{0} are used. Note that for each sub-region within the aggregate region (Figure 5.1), values for 'γ_0' and 'γ' corresponding to that sub-region are used. This information is used to form the $\mathbf{D}$ matrix in Eq. (12) for the aggregate region. The term d_g in Eq. (18) is independent of time-step 't', and needs to be calculated only the first time.

3. Set model reference year '$t = 1901$' .

4. Use the RF spatial data for cropland and pastureland for year 1891 to form the $Y^{\bar{t}}$ terms in Eq. (12). The size of vectors Y^{t} and $Y^{\bar{t}}$ in Eq. (12) is two times the total number of grid cells within an aggregate region.

5. For time-step 't', and for each grid cell 'g' within the aggregate region, calculate land suitability S_{lg}^{t} from Eqs. (13)-(14). For this calculation, spatial data on the standardized explanatory variables X_{lg}^{t} are used. For each sub-region within the aggregate region (Figure 5.1), values for 'β_0' and 'β' corresponding to that sub-region are used. This information is used to form the vector S^{t} in Eq. (12) for the aggregate region.

6. For time-step 't', the land use allocation model computes Y^{t} (from Eq. (12)) using five variables: (1) matrices $\mathbf{A}$ and $\mathbf{B}$ from step 1, (2) matrix $\mathbf{D}$ from step 2, (3) S^{t} from step 5, (4) $Y^{\bar{t}}$ where $\bar{t} = t - 10$, and (5) the regional total area demand for each land use type in time-step 't' which is used as input for the regional constraints through Eqs. (10)-(11). This step, when carried out separately for each of the nine world regions, results in a global map of cropland and pastureland for 't'.

7. Increment model time-step by one year ('$t = t + 1$').

8. Repeat steps 5 to 7 until '$t = 2005$'. For the first ten years of model simulation $(1901 \le t \le 1910)$, RF data for the period 1891-1900 are used to form the $Y^{\bar{t}}$ term. For $t \ge 1911$, the model predicted maps are utilized to form the $Y^{\bar{t}}$ term.

In summary, the land use allocation model requires three inputs: (1) maps of cropland and pastureland from 1891-1900 to form the lagged $Y^{\bar{t}}$ term for the first 10 years of model simulation (cropland and pastureland maps for the 1900 RF data are presented in Figure 5.3), (2) annual maps (1901-2005) of explanatory variables, and (3) the annual (1901-2005) aggregate demands for cropland and pastureland for each of the nine world regions. With these inputs the model dynamically allocates aggregate land use information at 0.5° x 0.5° spatial resolution for each year starting from 1901 until 2005. The dependence of our allocation model on previous and neighboring land use through $Y^{\bar{t}}$ and matrix **B** respectively result in high path dependence of the simulated land use patterns.

5.4 Results

Land use allocation model simulation and historical land use patterns

Figures 5.4-8 show the model predicted maps for cropland and pastureland from the model simulation at 20-year time intervals. Table 5.3 summarizes the comparison in terms of adjusted Kappa coefficients. Kappa coefficient is a statistical measure of inter-rater agreement, and range from zero to one (unit less quantities). Greater magnitude of kappa indicates better agreement between the simulated and the actual values (RF data). Adjusted kappa coefficient (Mertens et al., 2003) is same as kappa coefficient, but is intended for sub-grid mapping and ignores grid cells where both predicted and actual values are zero (including zero grid cells would inflate the kappa values without adding much information about models prediction abilities).

Overall, the model predicted fractional areas for cropland and pastureland for the entire 20th century are broadly consistent with RF data (Figures 5.4-8; table 5.3). There are two points that stand out from the temporal trend in kappa coefficients (table 5.3). First, the land use patterns predicted by the model better match the RF data toward the start and end of the simulation period. Second, the prediction of historical pastureland patterns is consistently worse than that for cropland.

The reason for the better performance at the start and end of the simulation is that the model begins in 1901 with the observed land use pattern and projected values deviate over time, so that outcomes early in the century, all else equal, are likely to be more accurate than those later in the century. Towards the end of model simulation, the predicted maps converge towards

the actual RF data because our estimated parameters are weighted more towards the contemporary relationships (section 2.6).

The reason for the limited accuracy in predicting pastureland is likely because the spatial reconstructions of pastureland are highly uncertain, so that the explanatory factors used in our study have limited capacity in explaining the historical pastureland patterns. For example, the RF data used here estimates global pastureland area at 26.3 million km^2 during 2005. In comparison, the other well-known HYDE 3.1 reconstruction (see Appendix B) estimates pastureland area at 33.0 million km^2 during 2005, 26% higher than RF data. Therefore, at grid-level, the relative uncertainties in pastureland estimates are even higher (Figure 5.9), and increase as we go further back in time from 2005 (Meiyappan and Jain, 2012).

A key feature of the model is its ability to replicate the timing and magnitude of spatial shifts in land use patterns that occur in the RF data, even within an aggregate region. For example, Figure 5.10 shows the model predicted net transitions in cropland over the US for the period 1900-1960 (calculated as the difference between 1960 model predictions and the 1900 reference map divided by the number of years). The model is able to reproduce the decline in cropland in the eastern US and subsequent expansion to the mid-western US that occurred over this period (Fig 5.10, compare top left and top right panels). The model is able to reproduce the shift in these patterns mainly because we account for the heterogeneous nature of the driving factors within each aggregate region and their changes over time. Similarly, we show the model is able to replicate the key spatial patterns of land-use change (as indicated by RF data) for other world regions: (1) Europe and the western portion of the Former Soviet Union (FSU) between 1935-1960, during which period Europe experienced a gradual decline in cropland, and FSU experienced sharp cropland expansion associated with the opening up of "New Lands" (compare top two left panels in Figure 5.11), (2) in the same region, but for the period 1960-2005, when cropland abandonment was common to both Europe and FSU (top two right panels in Figure 5.11), and (3) widespread net cropland expansion in the tropics between 1920 and 1980 that resulted in significant deforestation (top two panels in Figure 5.12). Overall, results indicate that the general patterns of cropland expansion and abandonment are replicated well compared to RF data, with some exceptions (e.g. in Figure 5.12, we simulate cropland expansion in the Caribbean and parts of India where RF shows abandonment).

Estimated parameters

The 'a' parameters indicate the relative importance of the dynamic adjustment cost model (dependence of land use allocation on the previous land use patterns) compared to the static profit maximization function (dependence of land use allocation on potential land suitability). The 'b'parameters indicate the nature and magnitude of spatial autocorrelation in land use patterns.

Three key results stand out from the estimated 'a' and 'b'parameters (Figure 5.13). First, the values for both the parameters are spatially heterogeneous across the globe. This heterogeneity would be left unaccounted for if models were not parameterized at sub-global scales.

Second, the 'b'parameters are non-zero and significant for most regions across the globe indicating that global land use change datasets have significant spatial autocorrelation (note that $b=0$ indicates no autocorrelation). Therefore, disregarding the presence of spatial autocorrelation from estimation procedure will result in biased parameter estimates. The negative values for 'b' across most sub-regions indicate the bias would tend to inflate the importance of driving factors in these regions because the estimated 'a'parameters would be smaller compared to that in Figure 5.13a (smaller because when 'b' is disregarded, the 'a' parameter would reflect the net effect of the 'a' and 'b' parameters).

Third, the 'a'parameters indicate that temporal autocorrelation is strong for most regions; i.e., the dynamic adjustment cost term dominates land suitability in determining land allocation, leading to a highly path dependent process. Higher 'a' values generally coincide with regions where extensive agriculture is found as early as 1900 (e.g. cropland in Europe, India and China in Figure 5.3a, and pastureland in USA, west of the Mississippi river in Figure 5.3b). The spatial patterns of land use in these regions reflect a long history of changes in land use in response to socioeconomic and biophysical factors. The model therefore tends to rely more on previous land use patterns to explain subsequent changes in land use patterns in these regions.

The importance of the adjustment cost term does not imply that driving factors that determine land use suitability are insignificant in a dynamic allocation procedure. High path dependency implies inaccuracies in predicting land use allocations in one year will reduce the accuracy of predictions for subsequent years. As will be shown in the next section, it is this path dependency behavior that makes inclusion of driving factors important. If we exclude driving

factors from the land use allocation procedure, the inaccuracy in land use allocations for initial years of simulation would be negligible, but over time they would accumulate to produce land use maps that are substantially different from the historical reconstruction.

Comparison to other models

To evaluate the land use allocation model, we repeat the historical simulation (last sub-section in 5.3) with two other common land use allocation approaches and compare them with our results. We designed both these approaches as representative of general allocation procedures; they do not replicate specific existing models. Full methodologies are provided in Appendix A.5. Here, we highlight the key features of these approaches.

(1) Proportional downscaling approach – The aggregate land use projections are allocated to grid cells as closely as possible to previous year land use patterns. No account is taken of the impact of driving factors. Models following this general approach include GLOBIO3, and GLM used to downscale land use projections from the Global Change Assessment Model (GCAM) model corresponding to the RCP4.5 scenario of the IPCC (van Vuuren et al., 2011).

(2) Constrained proportional downscaling approach – The aggregate land use projections are allocated to grid cells as closely as possible to previous year land use patterns, but the direction of change is constrained by the direction of change in a measure of land suitability. Land suitability is determined with regression relationships driven by explanatory variables, and the constraint implies that grid cells in which suitability decreases (increases) must have a decrease (increase) in land use (or no change). The magnitude of decrease (increase) is restricted to vary between the estimated land suitability and previous land use area. This approach mechanistically fits models to explain spatial processes in land use change and update the allocation maps accordingly. Models following this general approach include MIT-IGSM (Wang, 2008).

Performance of previously published land use allocation approaches

Results show that the final predicted land use map (2005) using both allocation approaches to downscale the aggregate land demands derived from the RF data are less accurate than our model (compare Figure 5.14 and 5.15 with bottom panels in Figure 5.8). The adjusted kappa coefficients indicate that at the end of model simulation (2005 A.D.), both the allocation

approaches have a 59-66% accuracy in simulating cropland patterns, compared to 87% accuracy by our model (table 5.3). The accuracy in simulating pastureland patterns also differ by similar magnitudes between our model and the other two approaches. Inaccuracies using the proportional downscaling approach are driven by the fact that the allocation across grid cells within an aggregate region is homogeneous (Figure 5.14), and does not capture major shifts in agricultural patterns caused by changes in the spatial patterns of driving forces (Figure 5.10 compare mid-left panel with top-right panel; Figures 5.11 and 5.12 compare middle panels with top panels). This leads to severe overestimation of land use within some regions (e.g. ~40% in eastern US for cropland) and a corresponding underestimation in other parts of the same region (e.g. >50% in Great Plains) (Figure 5.10). In contrast, the constrained proportional downscaling approach tends to reproduce shifts in agricultural patterns in some regions, but not accurately. For example, the abandonment of cropland in the eastern US is reproduced to some extent (Figure 5.10; mid-right panel), but cropland expansion occurs not only in the Great Plains but also in the western US (Figure 5.10; mid-right panel) where pastureland hotspots are located (Figure 8d). In the case of Europe (Figure 5.11) and the tropics (Figure 5.12), we find both the proportional and constrained proportional downscaling approaches capture the general regions of cropland abandonment and expansion (in reference to RF); however, the hotspots are severely underestimated because the allocation across grid cells within an aggregate region is homogeneous. As evident from our analysis this pattern of allocation is explained by two reasons. First, the importance of driving factors is underrepresented in a constrained proportional downscaling approach for most regions, because driving factors affect the direction but not the magnitude of land use change. Second, the importance of driving factors in determining land use allocation is assumed to be homogeneous across the globe. Therefore, how we represent the role of driving factors within a land use allocation procedure is as important as including the driving factors itself.

We note that the cropland transitions seen in the alpine tundra of the Himalayas (in both proportional and constrained proportional allocation approach) is an artifact of our model reproduction methodology (Appendix A.5). In principle, we can force the model not to allocate croplands in such biophysically unfavorable regions through grid cell constraints. However, we have not imposed such restrictions as our aim was only to elucidate the general allocation behavior of the both the approaches.

Coupled land use allocation model and historical carbon emissions

To investigate the sensitivity of environmental impacts to alternative land use allocation models, we apply the historical downscaled land use data from our land use allocation model as input to an important type of study land use change models are used for: projecting CO_2 emissions from land use change. The accuracy of the emissions results depends on getting the spatial patterns of land use correct where it most matters (i.e. where the carbon consequences are highest). We compare the simulated emissions with those obtained using the two other land use allocation approaches, and with other existing land use reconstructions available in literature.

We use a land-surface model, the Integrated Science Assessment Model (ISAM) to estimate net CO_2 emissions from land use change at 0.5° x 0.5° lat/lon resolution annually for the period 1900-2005. Further background information on ISAM and the simulation protocol is detailed in Appendix A.6. As six separate experiments, we calculate the CO_2 emissions due to changes in the areas of cropland and pastureland from six different land use change datasets. Three of the six land use change datasets are downscaled land use information: (1) from our land use allocation model, (2) the proportional downscaling approach, and (3) the constrained proportional downscaling approach. The fourth is the RF data for cropland and pastureland (the reference case because we prescribed the aggregate land demands in datasets (1)-(3) from RF). The other two datasets are independent reconstructions of historical land use change summarized in Meiyappan and Jain (2012): HYDE 3.1 (Klein Goldewijk et al., 2011), and Houghton (Houghton, 2008). The RF data used in our study is just one realization of what could have happened in the past (Appendix B). Therefore, estimates based on multiple reconstructions are particularly helpful to understand the range of uncertainties among available historical reconstructions of agricultural land use. Our reported net emission excludes emissions from indirect environmental effects (e.g. changes in climate, CO_2 fertilization, and nitrogen deposition).

Figure 5.16 provides a comparison of the estimated carbon emissions across the six land use change datasets at aggregate regional scale and cumulated over the period 1900-2005. Four key points are evident. First, estimates based on our land use allocation model compare well with that from RF data, as expected since our modeled spatial land use history also compares well with the RF data, and are within the uncertainty range of three reconstructions. Second, at an aggregate global scale, the proportional and constrained proportional downscaling approach

overestimate carbon emissions on average by ~0.17 PgC/yr (26%) and ~0.14 PgC/yr (23%) respectively, compared to RF data, even though these two approaches used the same aggregate regional land use change as the RF data. This overestimate is significant given that the total uncertainty (from agricultural land use change, other land disturbance activities, and knowledge gaps in process understanding and modeling) in estimating historical carbon emissions from land-use and land-use change is ~0.5 PgC/yr (Le Quéré et al., 2014). Third, both proportional and constrained proportional allocation approaches result in much higher disagreement at a regional scale, compared to RF data. A striking example is North America, where estimates based on the proportional and constrained proportional downscaling approaches are ~4.4 and ~2.5 times higher than RF data respectively. This is a consequence of the higher inaccuracy in reproducing the changes in agricultural hotspots by both downscaling approaches. In the case of RF data and our land use allocation model, abandonment of cropland over the eastern US (Figure 5.10; top panels) causes a larger carbon sink (hence, smaller net emissions) due to subsequent forest regrowth (see Figure A.1 in online supplementary). This important feature is reproduced only to some limited extent in the constrained proportional downscaling approach, and is nonexistent for the proportional downscaling approach (Figure 5.10; mid panels). A consequence of overestimating cropland in the eastern US is a corresponding underestimation of cropland expansion in other parts of the US (consequently lower carbon emissions in these regions compared to RF data). Therefore, at grid level the disagreement in estimated net emissions is much higher for proportional and constrained proportional allocation approaches compared to RF data (not shown). Fourth, net emission estimates based on the three existing historical reconstructions of agricultural land use show significant disagreement, especially at the regional scale. This disagreement is largely explained by the difference in agricultural inventory data used by these reconstructions (Jain et al., 2013; Meiyappan and Jain, 2012). No reconstruction is clearly better than another, as is evident from the uncertainties in net transitions estimated between the two reconstructions (see Figure 5.9 for pastureland, and compare the top and bottom panels in Figures 5.10-5.13 for cropland). Therefore, land use allocation approaches need not closely emulate any one reconstruction. However, despite the large uncertainty range among historical reconstructions, the emission estimates based on both the proportional and constrained proportional downscaling approaches fall outside this range for many regions. This underscores the importance of a reliable approach to modeling land use allocation. It is also important to

continuously improve the quality of historical land use data to improve models and to more accurately predict future land use change.

5.5 Discussion

The land use allocation model

We present a statistical model for land use allocation with an econometric interpretation of land suitability that is based on profit maximization (or cost minimization). The approach integrates economic theory, observed land use, and data on both socioeconomic and biophysical determinants of land use change. It is global in scope and is estimated using long-term historical data, thereby making it suitable for long-term projections, such as in IAMs. The method accounts for spatial heterogeneity in the nature of driving factors across geographic regions. The allocation is modified by autonomous development (previous and neighboring land use patterns, thereby accounting for temporal and spatial autocorrelation), competition between land use types, and exogenous drivers that are treated as explanatory variables. The spatial and temporal resolution at which the model operates is flexible.

Given that the biophysical components in most global-scale land use models operate at a spatial resolution of 0.5°x0.5° lat/lon (or coarser), representing landscape heterogeneity at the sub-grid scale level, at least to some extent, is important. The method of fractional land use prediction developed here is a first step towards representing landscape heterogeneity in global scale land use models. Ultimately, a highly detailed representation of the composition of landscapes is necessary for certain environmental assessments, such as biodiversity (van Asselen and Verburg, 2012; 2013). Two other equally important aspects in land use allocation procedure are to account for the: (1) transient impacts of socioeconomic and biophysical driving factors, and (2) spatial heterogeneity in the relative importance between driving factors and past land use patterns in determining land use allocation. As demonstrated in this study, other downscaling approaches that disregard the first aspect, but predict fractional land use fail to reproduce the hotspots of historical agricultural patterns. A constrained proportional land use allocation approach that accounts for the first aspect, but disregards the second, also fails to reproduce the historical land use patterns. We also show that such downscaling approaches could have significant implications for global environmental assessments. Our approach is novel because we predict fractional land use within each grid cell and simultaneously account for the spatial

heterogeneity in the relative importance between past land use patterns, and driving factors that change with time.

As a first step, we apply the model framework to evaluate the land use patterns for two generic land use types: cropland and pastureland. However, the framework is extendable to account for individual crop types, for instance to study the global land use implications of large-scale biofuels (Melillo et al., 2009; Havlik et al., 2011; Hallgren et al., 2013). For our model evaluation, we attempt to compile a database of the most important explanatory variables available at the global scale (table 5.1). Though the list is incomplete, we show that these variables are adequate to reproduce the changes in the hotspots of historical land use change. To use the allocation framework within an IAM, the stationary explanatory factors (soil and terrain conditions) can be retained, whereas data on transient explanatory factors (e.g. climate and socioeconomics) should be replaced with that simulated by the IAMs. This would allow for studying the two-way interactions between land use and the environment (e.g. climate, hydrology). Further, not all transient explanatory variables used in our historical simulation are projected by current IAMs. An alternative is to replace the explanatory variable with other equivalent proxy indicators simulated by IAMs (e.g. Net Primary Productivity). In such cases, the parameter estimation procedure and evaluation should be repeated using the method discussed in section 5.3, following which the evaluated model can directly be used within IAMs to explore future scenarios.

Land use competition

Competition for land in itself does not drive land use, but is an emergent property of other drivers and pressures (Smith et al., 2010). Our approach accounts for land use competition by simultaneously optimizing the area of cropland and pastureland within each grid cell with the aim of maximizing the overall achievable profits. Hence, the approach prevents inconsistency and approximations in the allocation procedure that would otherwise arise when the spatial patterns for each land use type are determined independently (e.g. see Wang, 2008). Incorporating the effects of competition into spatial allocation models is an important feature, given the potential for growth in demand for agricultural land, particularly for pasture to support projected increase in meat intensive diets, especially in developing countries (Stehfest et al., 2009).

Caveats and concluding remarks

Expansion of cropland and pastureland is accompanied by conversion of native vegetation (except when one land use is converted to another). Therefore, one of the important factors that determine the allocation of cropland and pastureland is the type of native vegetation to be replaced. Each native (and managed) vegetation offers different resistance (cost) to conversion depending on the land use type. For example, historically most of the pastureland (managed grassland) has been derived from natural grasslands (with exceptions, notably Latin America where forests are cleared for cattle ranching). However, we chose not to include the type of native vegetation to be converted as a factor in determining the land use allocation patterns. There are three reasons for this choice.

First, our land use allocation is carried out at 0.5°x0.5° lat/lon (~55 km x 55 km) resolution, consistent with most global scale land use models. A mix of native vegetation usually occupies such large grid areas. Therefore, the difference in the resistance offered by native vegetation across grid cells becomes less important compared to an approach in which land use data are downscaled to a much higher spatial resolution.

Second, available global scale reconstructions of native vegetation for the 20th century are highly uncertain, especially before the 1960s when remote-sensing observations were unavailable (Meiyappan and Jain, 2012). Therefore, it is undesirable to constrain and evaluate a land use allocation model based on highly uncertain data.

Third, the representation of natural landscapes is fundamentally different among current generation biophysical models. Therefore, land cover maps produced for one model cannot be implemented directly within other models (Pitman et al., 2009). As a result, even in the most recent Coupled Model Intercomparison Project phase 5 (CMIP5) (Taylor et al., 2012), land use changes and the resulting changes in native vegetation are estimated sequentially. First, maps of land use from historical reconstructions are harmonized to connect smoothly with the land use maps for the future scenarios produced by the IAMs (Hurtt et al., 2011). The climate modeling teams combine the land use information with different techniques (e.g. Hurtt et al., 2011; Meiyappan and Jain, 2012; Lawrence et al., 2012; Pitman et al., 2009) to estimate changes in the area of native vegetation, so as to be consistent with the land surface components of their climate or earth system models. Therefore, our approach is consistent with current approaches.

Despite irrigation exerting a positive influence on agricultural suitability (Lambin et al., 2001, 2003), data limitations precluded us from including irrigation as an explanatory factor in

our analysis. While contemporary maps of irrigated areas are available (e.g. Portmann et al., 2010; Thenkabail et al., 2009; Siebert et al., 2005), we could not find historical maps on irrigation to account for the transient impacts. In principle, the presented framework could factor irrigation into the analysis, provided the data are available (e.g. for exploring scenarios that vary assumptions about irrigation).

We also do not explicitly account for the effect of policies on land use allocation patterns, despite their importance. Policy effects are explicitly accounted for in global land use or IAMs that would provide aggregate regional demands for land that would drive our allocation model. However ideally it would still be useful to reflect policies that operate at smaller spatial scales as well, such as land protection or national planning schemes. Globally and temporally consistent data on policies are not readily available for the 20th century. Although we have not accounted for policy effects explicitly, they are implicitly reflected in historical land use outcomes and therefore in the effects of other explanatory variables that could be considered proxies. In addition, for exploring policy effects in future scenarios, such assumptions can be accounted through grid cell constraints. For example, we can force grid cells within protected areas to not allow any land use allocation.

We predict only the net changes in land use areas within each grid cell because our model relies on land use reconstructions that provide only net change information. Available global scale reconstructions of land use (e.g. RF, HYDE, and Hurtt et al., 2011 data which use cropland and pastureland transitions from HYDE) are estimated based on the difference in (sub-)national statistics between two time steps, and therefore estimate net changes. Recent regional studies have shown that relying on net changes rather than gross changes (all area gains and losses) could lead to severe underestimation of land use change (Fuchs et al., 2014), and consequently have significant implications for environmental assessments, especially on the terrestrial carbon cycle (Wilkenskjeld et al., 2014).

Another important limitation is that our land use allocation model does not provide any information on land use intensity. Since 1960, a tripling of crop production has been achieved mainly through intensification, with only a 14% extensification (Bruinsma, 2009). Intensification is expected to become even more decisive in the future in the light of growing population, biofuel consumption, and mandates to protect world forests (Foley et al., 2011; Tilman et al., 2011; Phalan et al., 2011). Several global scale grid level indicators of agricultural land use

intensification have recently become available to foster modeling efforts (Kuemmerle et al., 2013), although substantial data gaps, uncertainties, and conceptual challenges exist (Keys and McConnell, 2005; Erb et al., 2013; Kuemmerle et al., 2013; Lambin et al., 2000). If land use intensity is measured by yields (i.e. output per unit area of land use activity) then the product term $S_{\mathrm{lg}}^{t} \acute{} \, d_g$ (derived from Eq. (13), Eq. (14) and Eq. (18)) itself is a measure of land use intensity. In principle the capital-related inputs (e.g. fertilizer, irrigation, pesticides, or mechanization) that increase agricultural yields can be accommodated as explanatory variables in the logit functions, assuming data are available. If land use intensity is measured by the frequency of cultivation (multiple cropping), then the total harvested area within a grid cell in a year could exceed the grid cell area. Our land use allocation procedure is versatile to accommodate such datasets (e.g. Ray et al., 2012; Portmann et al., 2010) as well. The RF data used in this study does not account for multiple cropping (i.e. $Y_{\mathrm{lg}}^{t} < GA_g$). Therefore, for feasibility, we have assumed the decreasing returns-to-scale is strong enough to deter full use of grid cell area (i.e. $d_g < 1$; see Eq. (18)). However, this restriction when relaxed can handle data on multiple cropping. Detailed representation of management characteristics such as land use intensity is important for IAMs to better capture human-environment interactions and to further improve our prediction capacity.

In a complementary study, we will extend the analysis to examine the role of different explanatory factors in shaping the 20th century patterns of agriculture. This will be carried out using two methods: (1) examining the values of the standardized 'β' and 'γ' parameters, and (2) simulating the land use allocation model (last sub-section in 5.3) in the absence of historical changes observed for one or more explanatory variables by keeping the explanatory factors of interest constant at initial values, with all other factors varying with time. We will quantify the effects of an explanatory variable by calculating the grid cell level differences between the final predicted map (2005) without changes in this variable, and the map (2005) obtained by varying all variables (as in Figure **5.**8).

Our ability to model land use change on longer time scales is crucial for exploring policy alternatives, especially because adaptation and mitigation of climate change requires long-term commitment. Notwithstanding the aforementioned caveats, the framework presented here and the approach to evaluation provides an example that can be useful to the IAM, land use, and the Earth system modeling communities.

Acknowledgements

The National Aeronautics and Space Administration (NASA) Land Cover and Land Use Change Program and the National Science Foundation (NSF) Regional Earth System Modeling program supported this work.

5.6 Tables

Table 5.1 List of the 46 potential explanatory factors used in the regression analysis. The explanatory factors cover the time period 1901-2005 at annual resolution. The spatial resolution is 0.5° x 0.5° lat/lon. Each seasonally averaged explanatory factor translates into four explanatory variables in our analysis (one for each season: spring, summer, fall and winter).

Broad Category	Explanatory Factor	Unit
Climate	Seasonally averaged temperature	K
	Seasonally averaged precipitation	mm/day
	Seasonally averaged Potential Evapotranspiration (PET)	mm/day
	Squared seasonally averaged temperature	K^2
	Squared seasonally averaged precipitation	mm^2/day^2
	Squared seasonally averaged PET	mm^2/day^2
	Seasonal Temperature Humidity Index (THI)	°C
Climate Variability	Seasonal Palmer Drought Severity Index (PDSI)	(-)
	Heat wave duration index	No of days
	Simple daily precipitation intensity index	mm/day
Soil Characteristics	Rooting Conditions and Nutrient Retention Capacity	(-)
	Nutrient Availability	
	Oxygen Availability	
	Chemical Composition (indicates toxicities, salinity and sodicity)	
	Workability (indicates texture, clay mineralogy and soil bulk-density)	
Terrain Characteristics	Elevation, Altitude and Slope Combined	
Socioeconomic	Built-up/urban land area	Fraction of grid area (m^2/m^2)
	Urban population density	Inhabitants/km^2
	Rural population density	
	Rate of change in rural population density	Inhabitants/km^2/yr
	Rate of change in urban population density	
	Market Influence Index	International dollars/person

Table 5.2 Data sources used to derive the explanatory variables for model evaluation.

Category	Data Variable	Description/Units	Spatial Characteristics	Period of Availability	Source
Climate	Temperature (T_a)	^oC	0.5 degrees (lat/lon)	1901-2012 (monthly)	Harris et al. (2013)
	Daily Average Maximum Temperature (T_{max})				
	Potential Evapotranspiration	Millimeters			
	Precipitation				
	Wet Day Frequency	days			
	Palmer Drought Severity Index (PDSI)	No units	2.5 degrees[a] (lat/lon)	1870-2010 (monthly)	Dai et al. (2011a,b)
Soil Constraints	Rooting Conditions and Nutrient Retention Capacity	Categorical Data classified into 7 gradient classes of land suitability for agriculture	5 minutes[b] (lat/lon)	Constant with time	Fischer et al. (2012)
	Nutrient Availability				
	Oxygen Availability				
	Chemical Composition (indicates toxicities, Salinity and Sodicity)				
	Workability (indicates texture, clay mineralogy and soil bulk-density)				
Terrain Constraints	Elevation, Slope and Inclination Combined	Categorical Data classified into 9 gradient classes			
Socioeconomic Factors	Urban/built-up land	% of grid cell area	5 minutes[b] (lat/lon)	10,000 BC – 2005 AD (decadal)[c]	Goldewijk et al. (2010)
	Urban Population	Inhabitants/km^2			
	Rural Population				

Table 5.2 (Cont.)

	Gross Domestic Product (GDP) per capita	Constant 1990 international (Geary-Khamis) dollars/person	National level	1 AD-2010 (annually between 1800-2010)[d]	Bolt and Van Zanden (2013)
	Market Accessibility	No units	1 km[b] (lat/lon)	~2005	Verburg et al. (2011)

[a] Was linearly interpolated to 0.5°x0.5° spatial resolution.
[b] We aggregated the data to 0.5°x0.5° spatial resolution by area-weighed averaging.
[c] We calculated annual estimates by linear interpolation of decadal data.
[d] Missing values for countries were gap filled using nearest values.

Table 5.3 The adjusted kappa coefficient estimated by comparing the model allocated land use maps from the historical simulation (last sub-section in 5.3) with RF data for different years. The values are given for cropland and pastureland (brackets).

Year	Land use allocation model developed in this study	Proportional downscaling approach	Mechanistic downscaling approach
1920	0.90 (0.81)	0.73 (0.76)	0.76 (0.72)
1940	0.88 (0.81)	0.71 (0.73)	0.73 (0.69)
1960	0.85 (0.80)	0.63 (0.60)	0.69 (0.60)
1980	0.82 (0.80)	0.61 (0.53)	0.64 (0.58)
2005	0.87 (0.83)	0.59 (0.58)	0.66 (0.61)

5.7 Figures

Figure 5.1 The nine aggregate world regions used in this study (indicated by different colors). Of these, three are individual countries (US, China, and India) and the other six are aggregated regions. The spatial data for annual cropland and pastureland from historical reconstruction (RF data), were aggregated to these nine regions and used as regional demand constraint to the land use allocation model. The partitioning of the globe into 127 sub-regions is also shown.

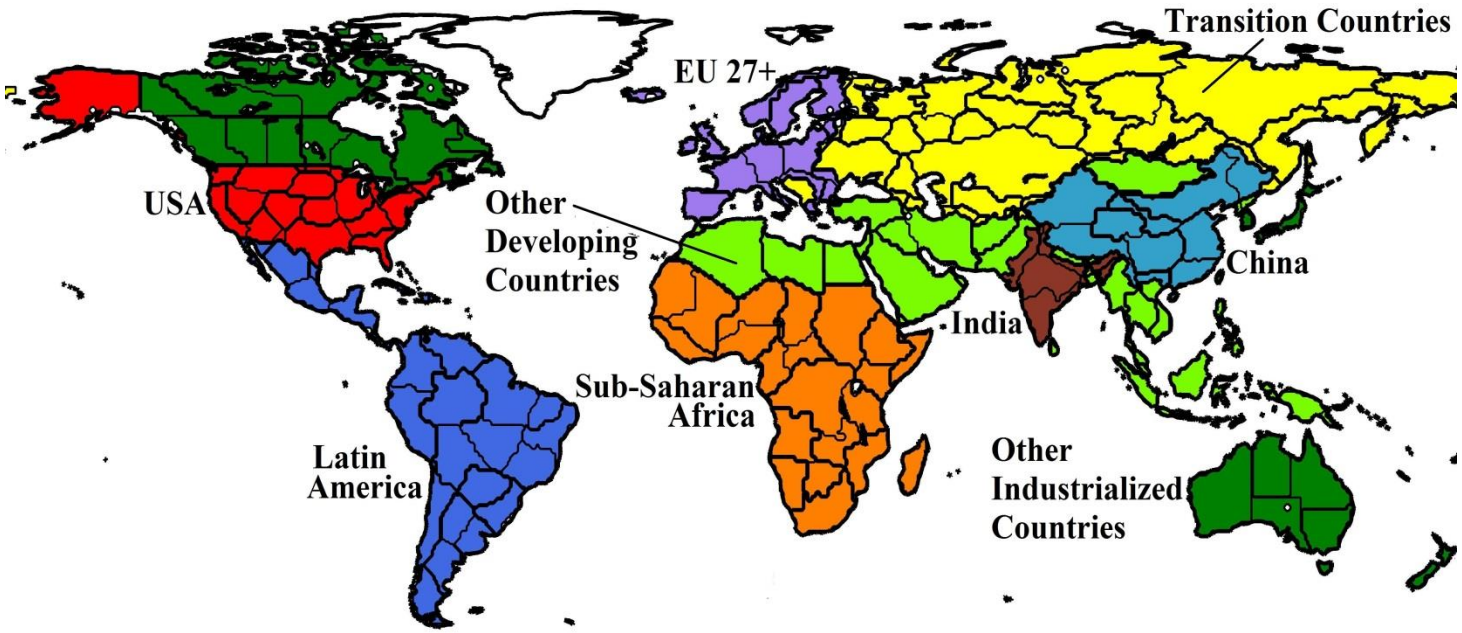


Figure 5.2 Schematic representation of the flow of model evaluation experiment.

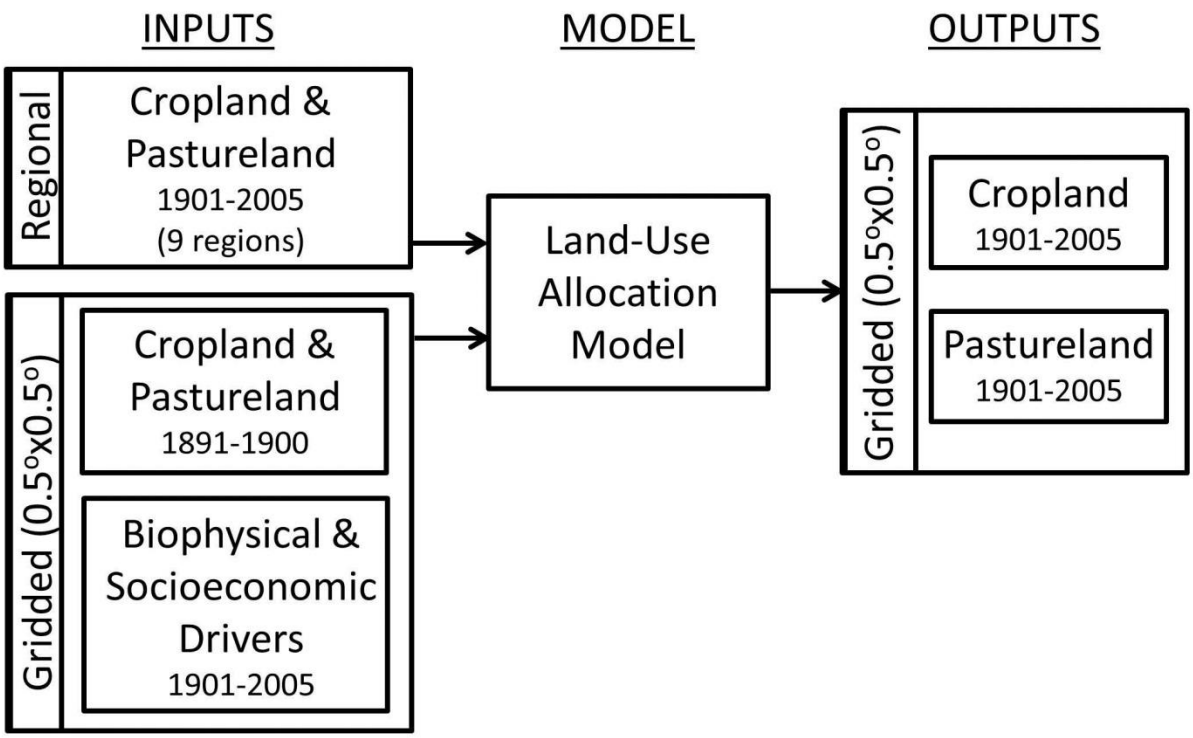


Figure 5.3 The cropland and pastureland maps (RF data) for the year 1900 used to form the lagged land use term in the land use allocation model. Units are in percentage of land area within each grid cell.

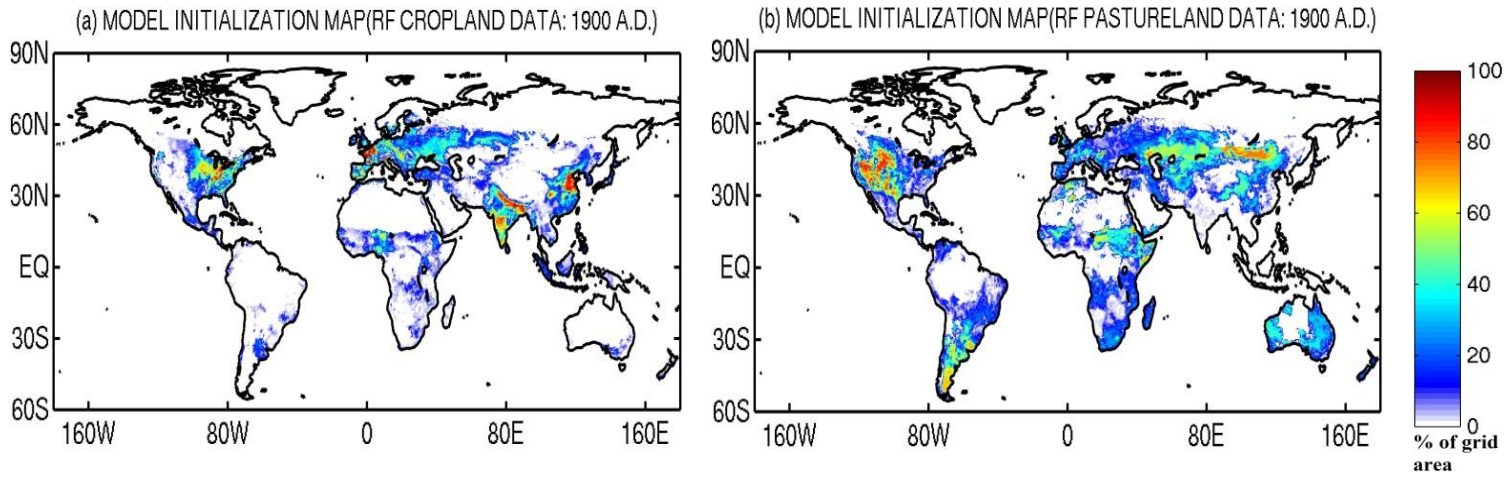


Figure 5.4 Model predicted map for cropland and pastureland (top panels) after 20 years of model simulation (i.e. 1920 A.D.). The RF data for 1920 (bottom panels) is shown for comparison purpose. Units are in percentage of land area within each grid cell.

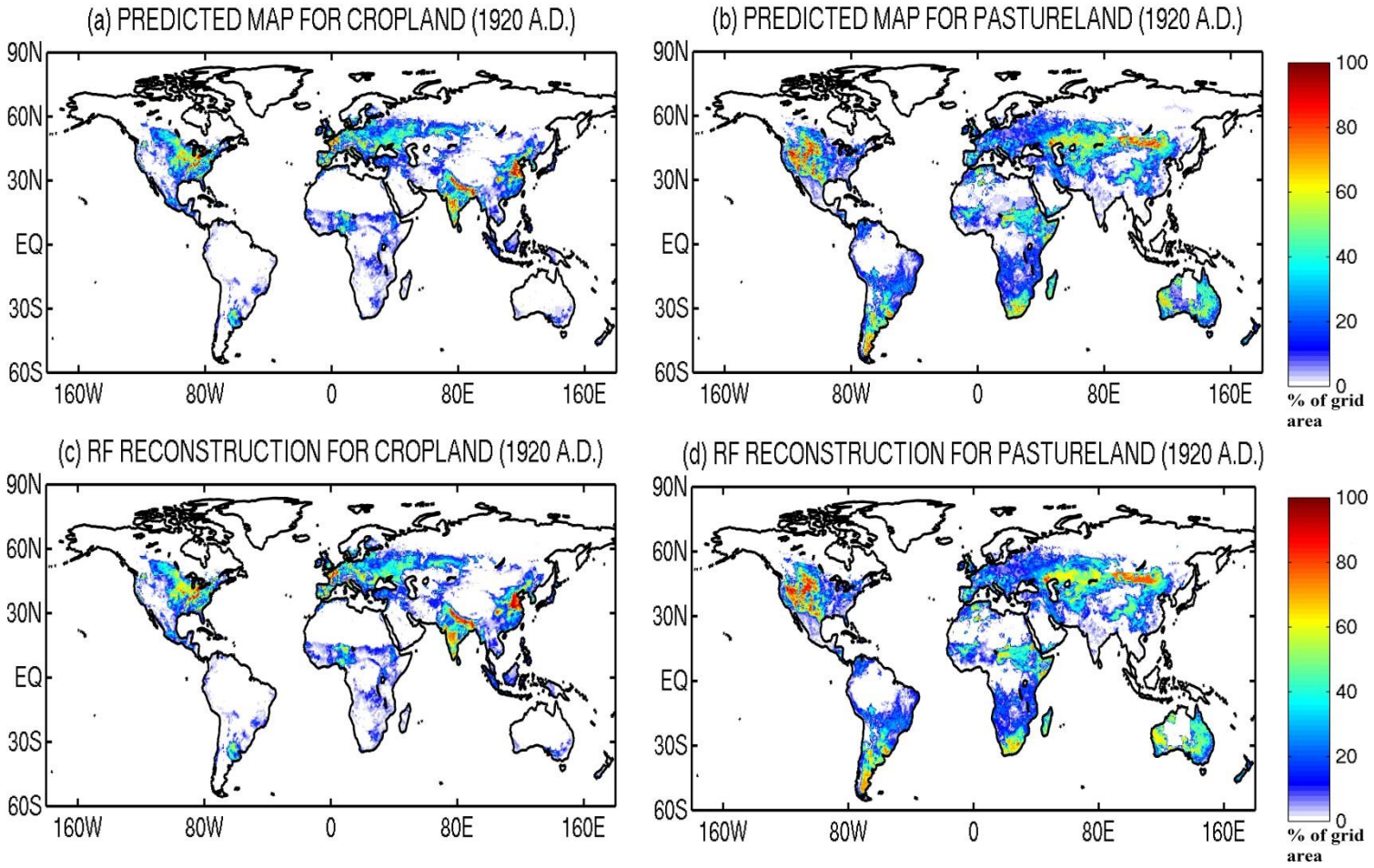


Figure 5.5 Same as Figure 5.4, but for the year 1940.

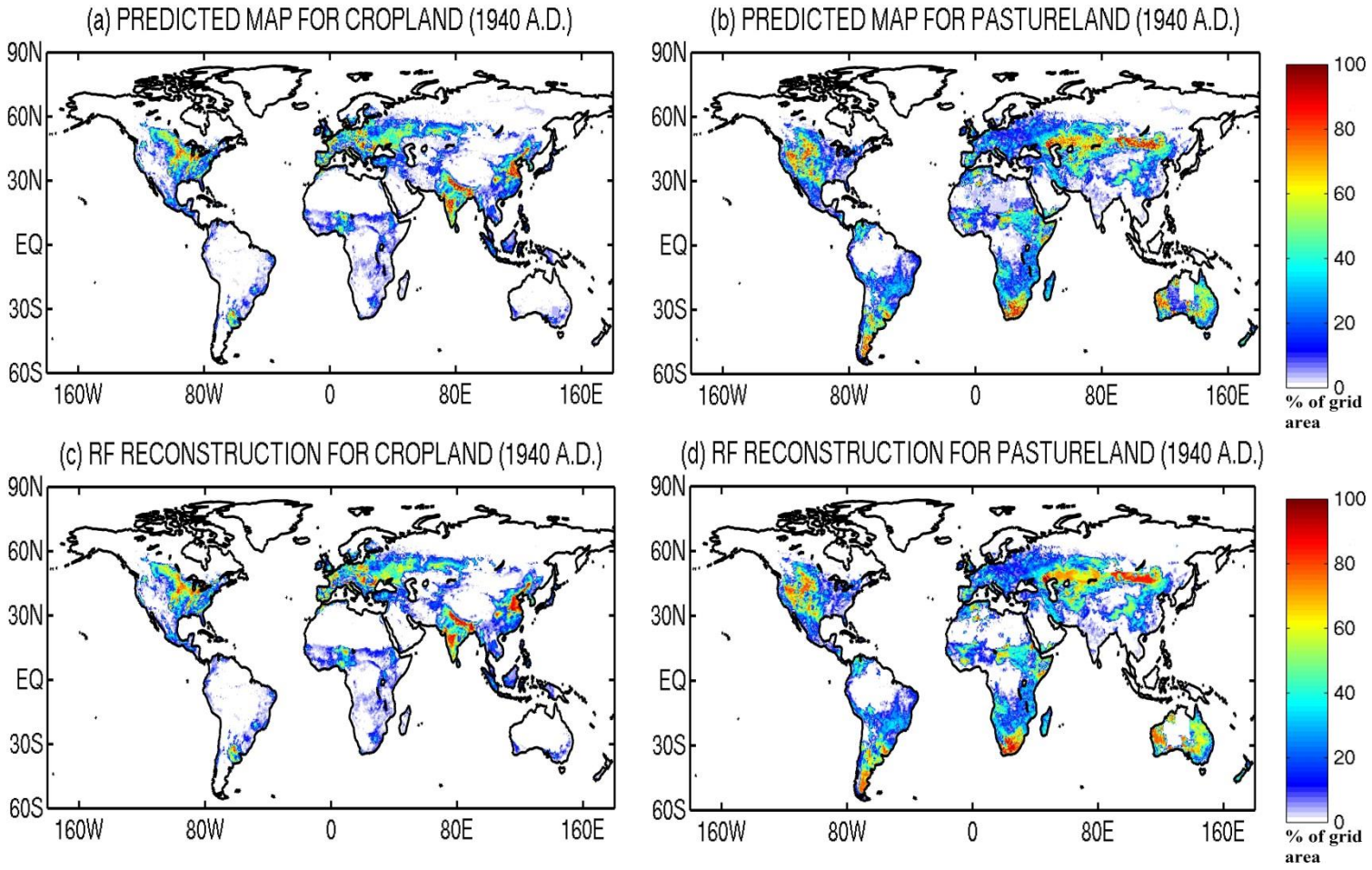


Figure 5.6 Same as Figure 5.4, but for the year 1960.

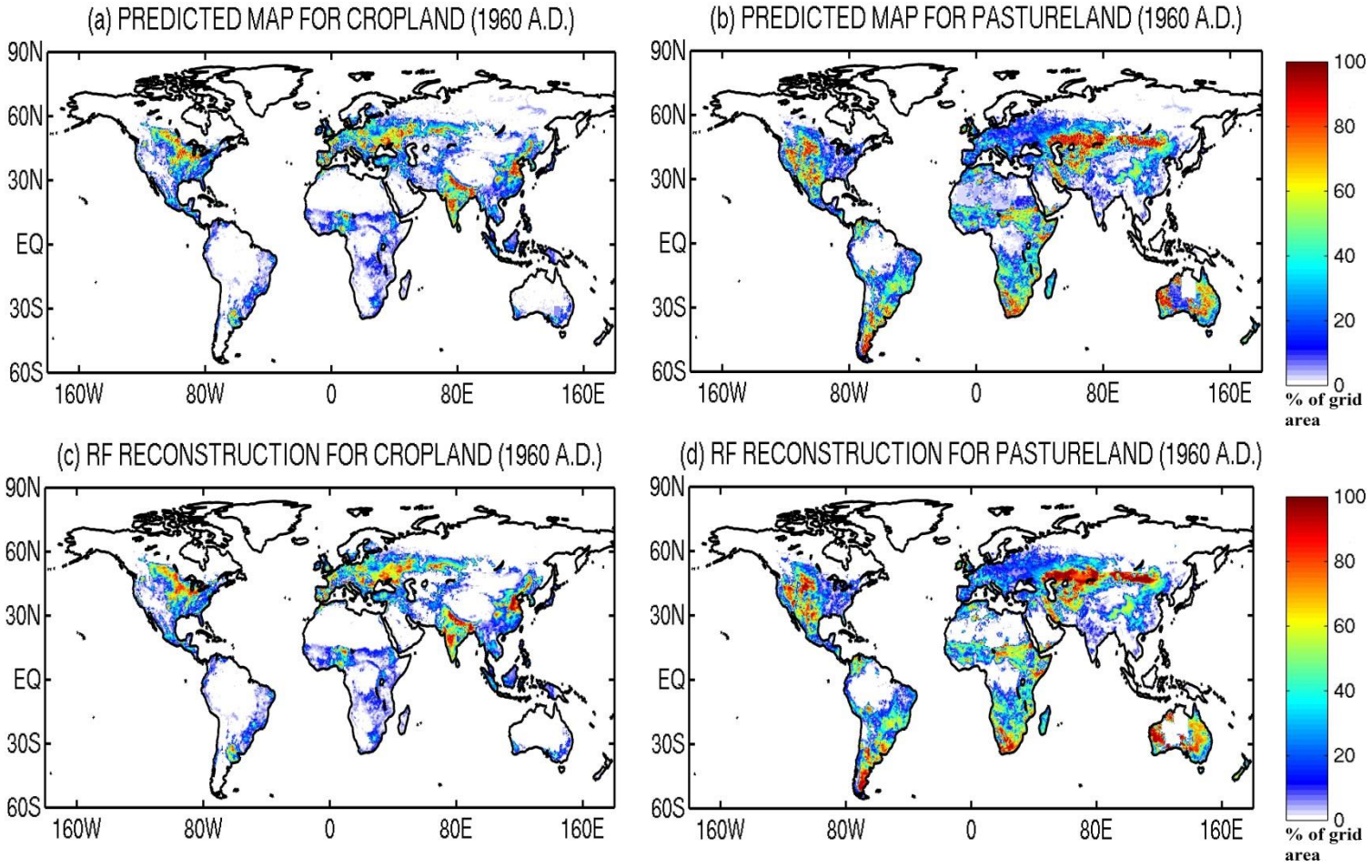


Figure 5.7 Same as Figure 5.4, but for the year 1980.

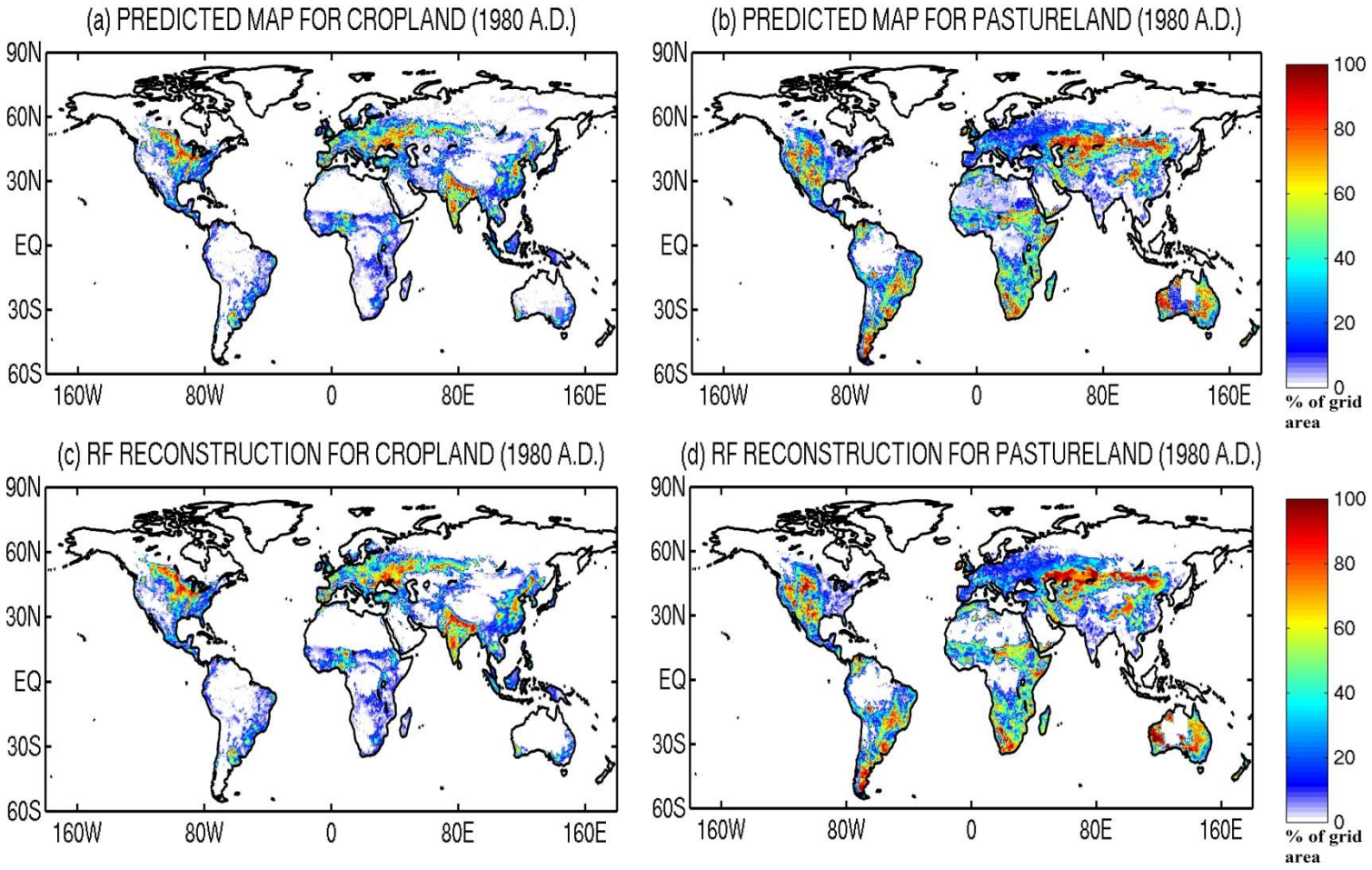


Figure 5.8 Same as Figure 5.4, but for the year 2005.

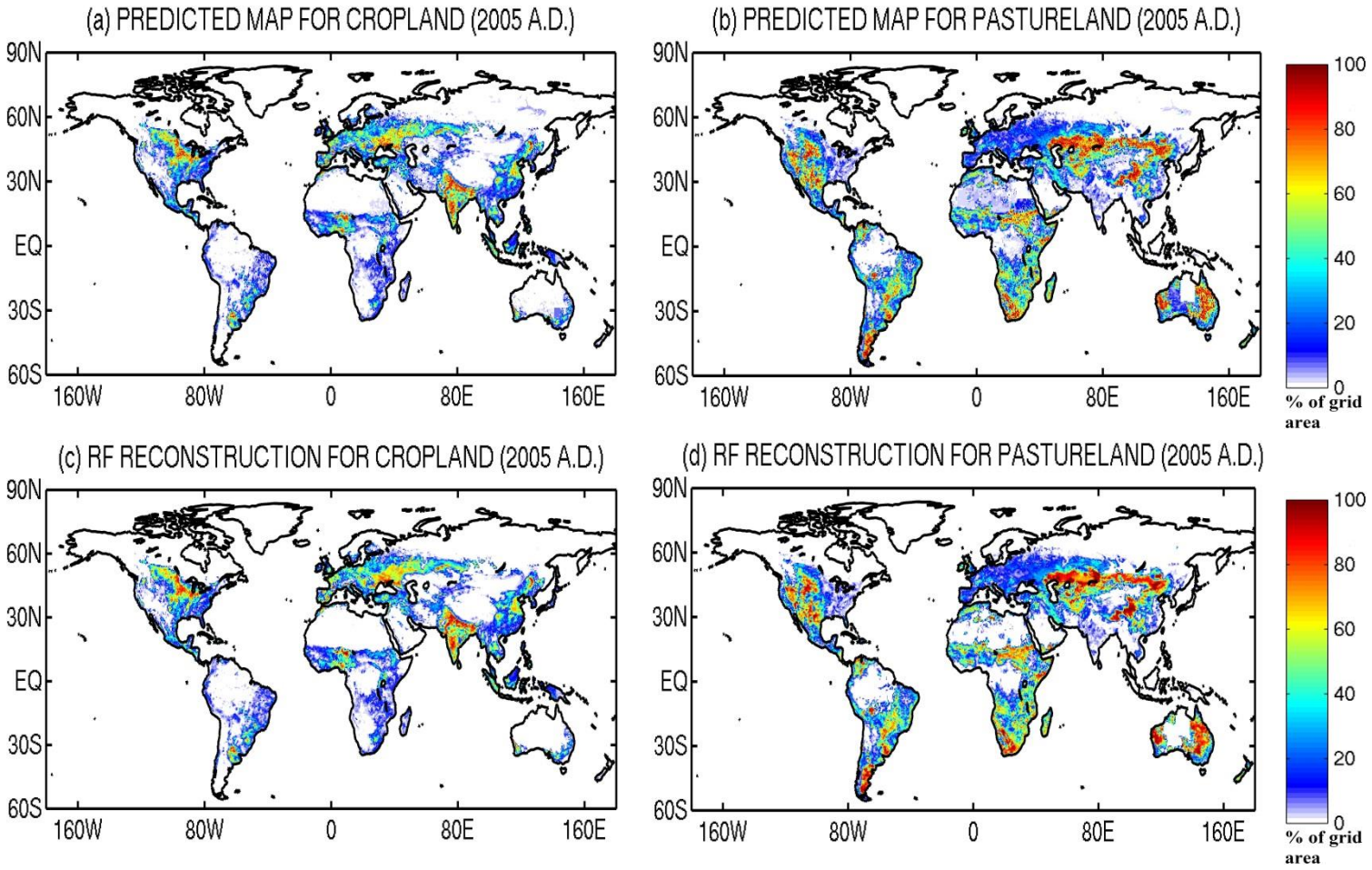


Figure 5.9 Comparison of annual net transitions (1900-2005) for pastureland between two widely used spatial reconstructions: (a) HYDE 3.1 database (Klein Goldewijk et al., 2011), and (b) RF data used in our historical simulation. Net transitions for 1900-2005 are calculated as the difference between 2005 map and the 1900 map divided by the number of years. Positive values indicate a net pastureland expansion over the period 1900-2005, and vice-versa for negative values. Units are in $km^2\ yr^{-1}$.

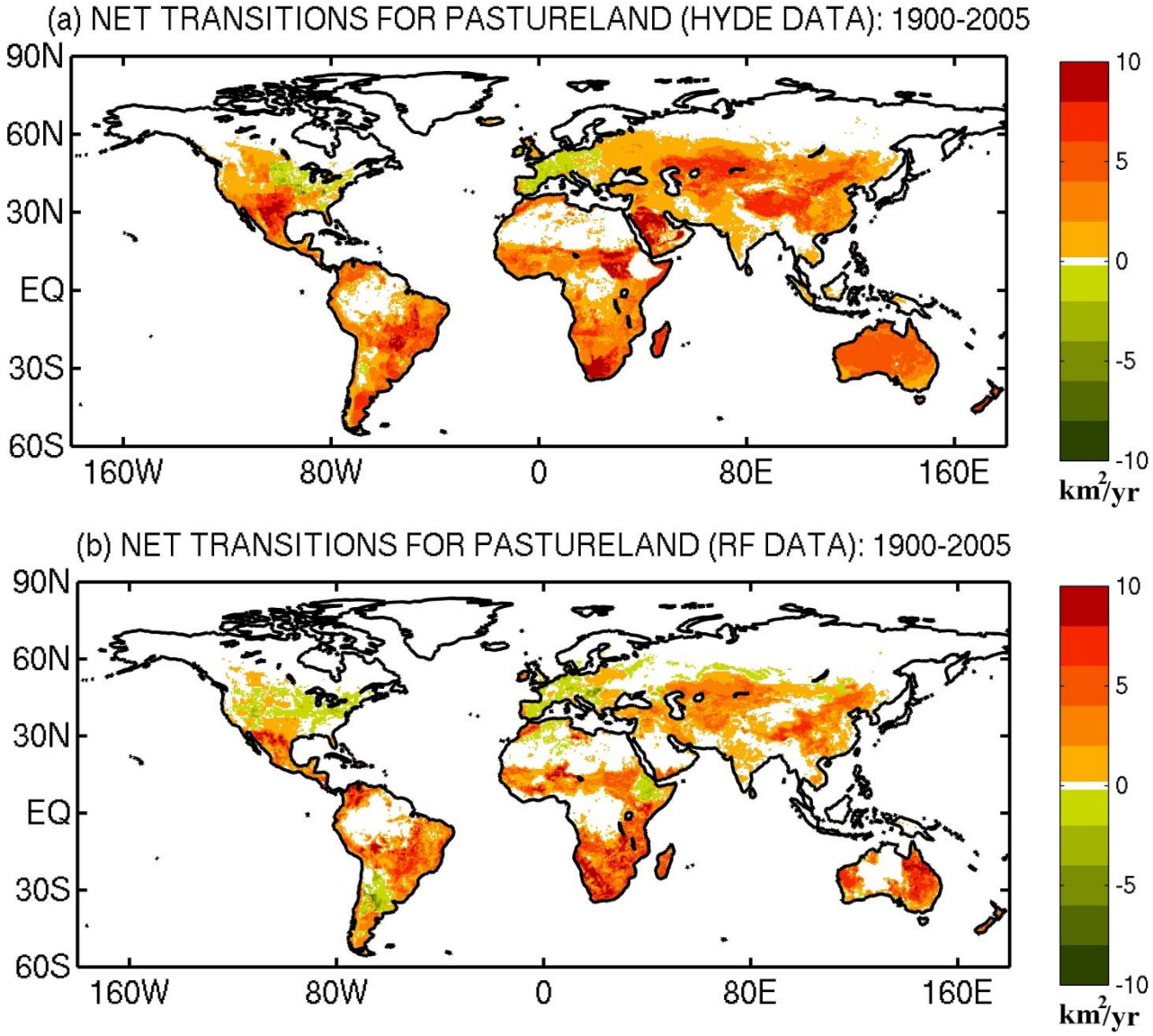


Figure 5.10 Net transitions for cropland over the US, averaged over the period 1900-1960 based on: our land use allocation model (top-left), RF data (top-right), proportional downscaling approach (mid-left), constrained proportional downscaling approach (mid-right), and HYDE 3.1 data (bottom). Net transitions are calculated as explained in Figure 5.9 captions. Units are in km^2 yr^{-1}.

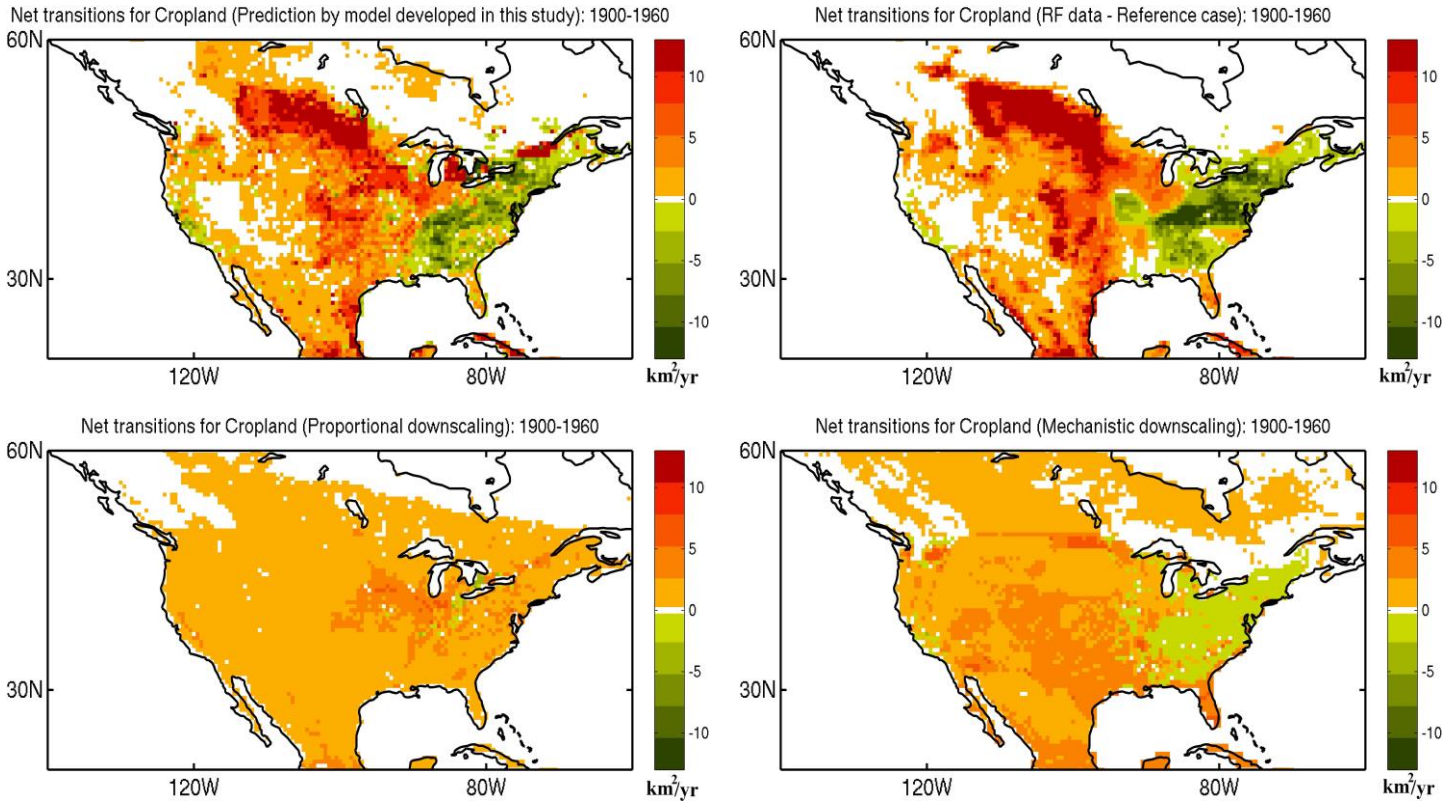


Figure 5.10 (Cont.)

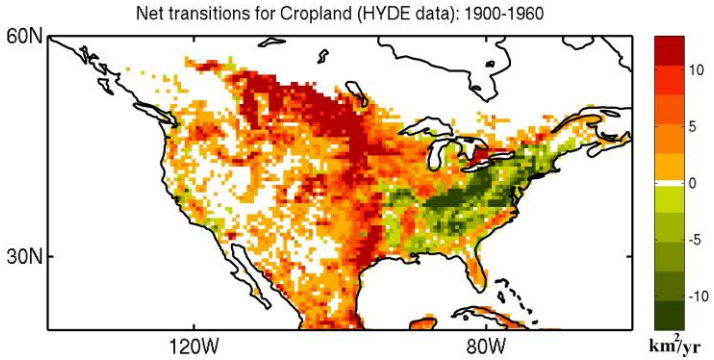


Figure 5.11 Net transitions for cropland as in Figure 5.10, but shown for the European region. Net transitions shown are averaged for the period 1935-1960 (left panels) and 1960-2005 (right panels). Units are in $km^2\ yr^{-1}$.

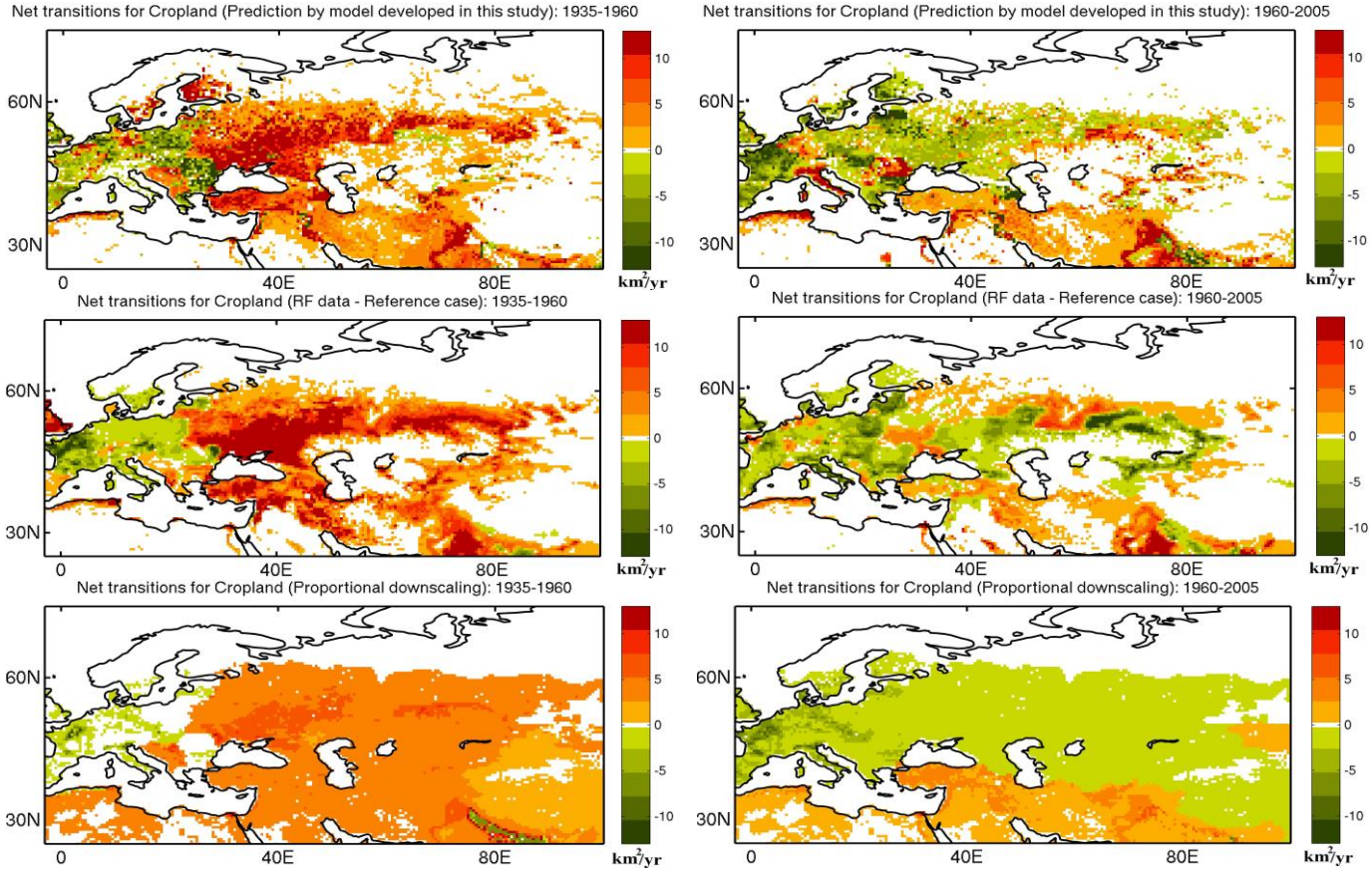


Figure 5.11 (Cont.)

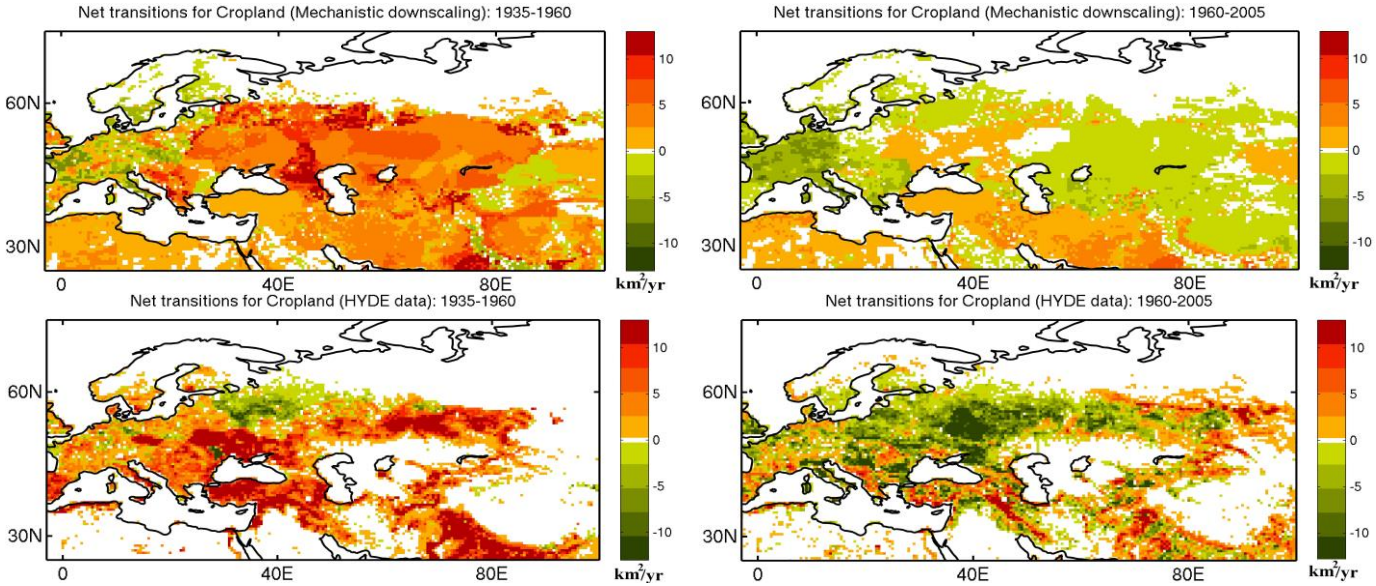


Figure 5.12 Net transitions for cropland as in Figure 5.10, but shown for the tropics and rest of southern latitudes. Net transitions shown are averaged for the period 1920-1980. Units are in km^2 yr^{-1}.

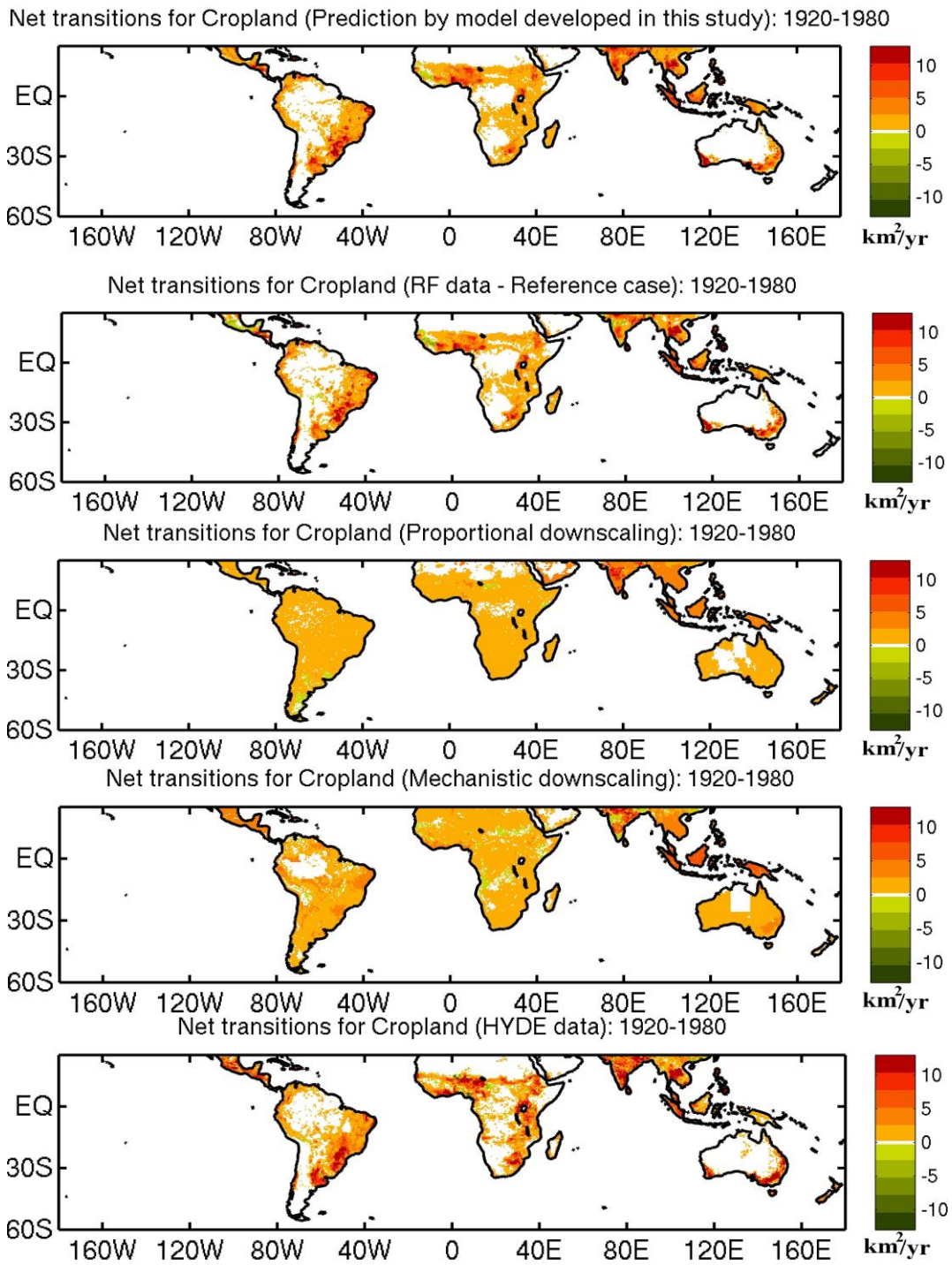


Figure 5.13 Estimated ‘a’ and ‘b’ parameters for each of the 127 sub-regions. The data for both the plots has been independently sorted in ascending order for visualization purpose. The parameters are unit less.

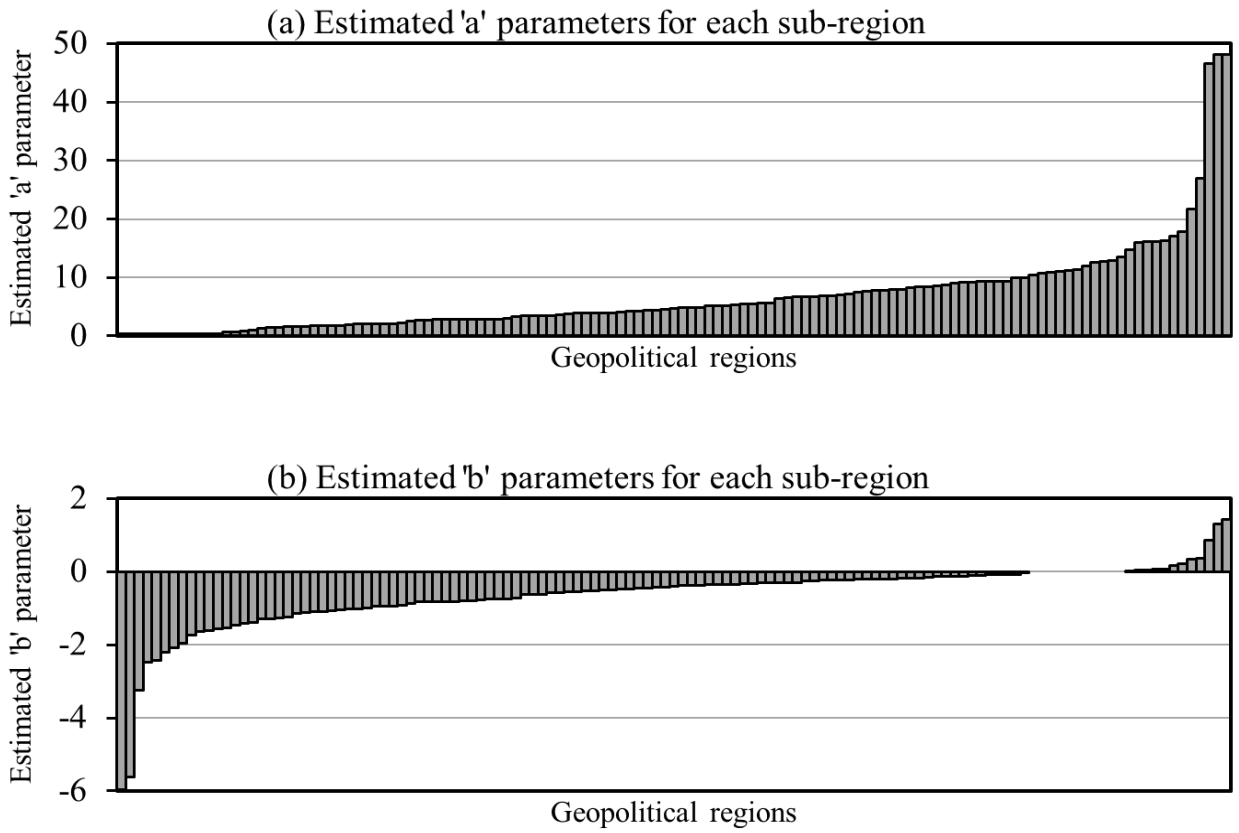


Figure 5.14 Cropland and pastureland maps predicted using proportional downscaling approach at the end of model simulation (2005 A.D). Units are in percentage of land area within each grid cell.

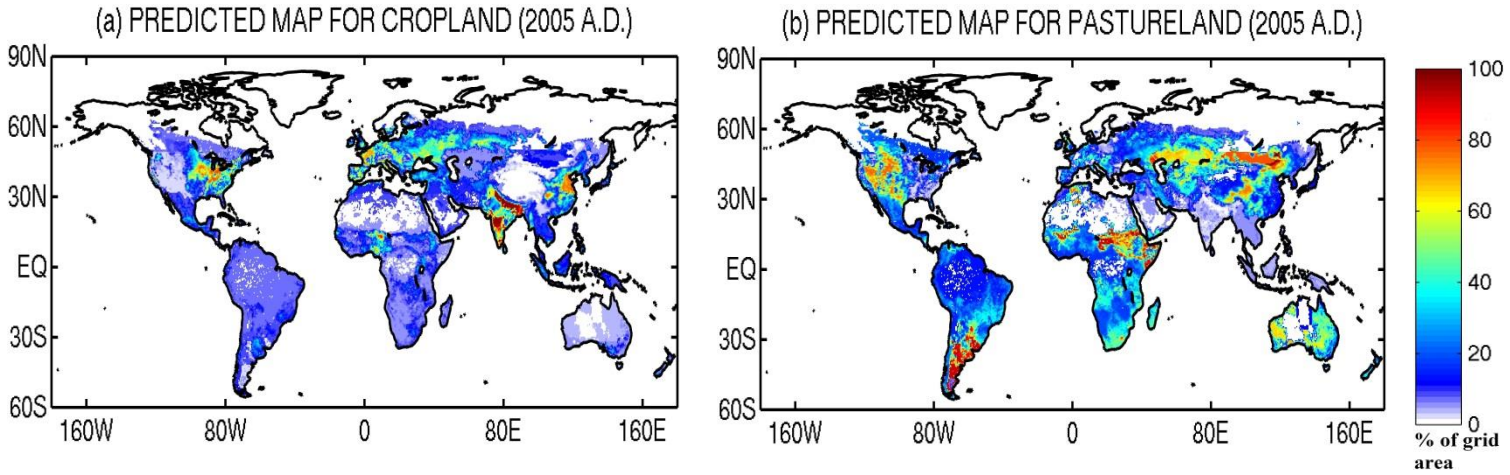


Figure 5.15 Cropland and pastureland maps predicted using constrained proportional downscaling approach at the end of model simulation (2005 A.D). Units are in percentage of land area within each grid cell.

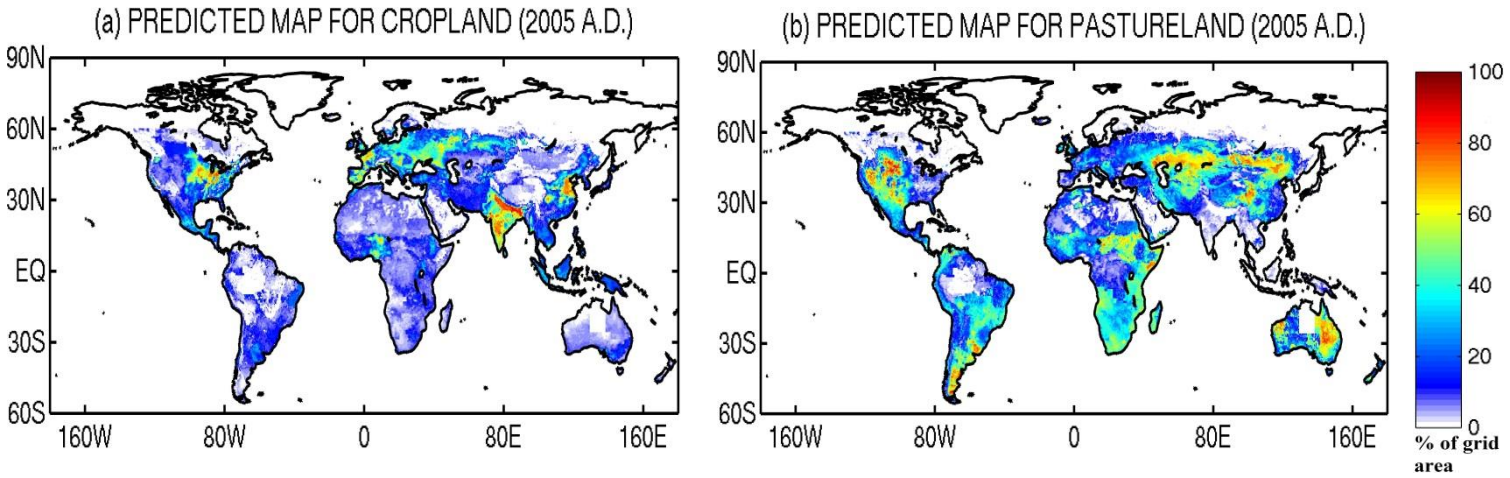
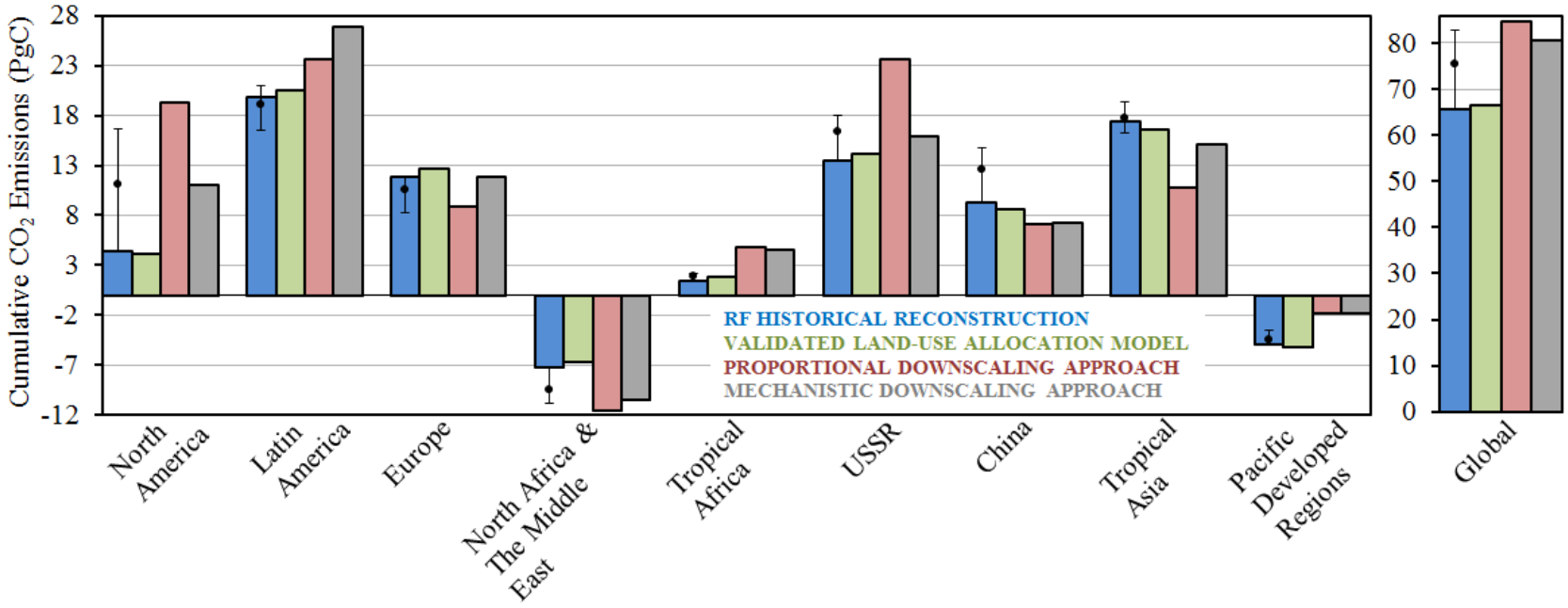


Figure 5.16 Estimated carbon emissions from changes in the cropland and pastureland areas calculated using six different land use change datasets. The data presented are aggregated to regional scales and cumulated over the period 1900-2005. The nine regions shown here are consistent with Jain et al. (2013) and different from the nine aggregate regions used in our model evaluation. The black dots represent the estimates averaged over the three historical reconstructions of land use (RF, HYDE 3.1 and Houghton), and the uncertainty bars indicate the maximum range across the three reconstructions. Units are in PgC (1 PgC = 10^{15} gC).



5.8 References

Agarwal, C., Green, G. M., Grove, J. M., Evans, T. P., & Schweik, C. M. 2002. A review and assessment of land use change models: dynamics of space, time, and human choice (pp. 12-27). Newton Square, PA: US Department of Agriculture, Forest Service, Northeastern Research Station.

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., ... & Vazquez-Aguirre, J. L. 2006. Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research: Atmospheres (1984–2012), 111(D5).

Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M., & ten Brink, B. 2009. GLOBIO3: a framework to investigate options for reducing global terrestrial biodiversity loss. Ecosystems, 12(3), 374-390.

Augustin, N. H., Mugglestone, M. A., & Buckland, S. T. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology, 339-347.

Bennett, E. M., Carpenter, S. R., & Caraco, N. F. 2001. Human impact on erodable phosphorus and eutrophication: a global perspective. BioScience, 51(3), 227-234.

Bolt, J., and van Zanden, J. L. 2013. The Maddison Project: The First Update of the Maddison project; Re-Estimating Growth Before 1820. Maddison-Project Working Paper WP-4 (Available from http://www.ggdc.net/maddison/maddison-project/home.htm).

Bouwman AF, Kram T, Klein Goldewijk K 2006. Integrated modeling of global environmental change. An overview of IMAGE 2.4. Netherlands Environmental Assessment Agency, Bilthoven.

Breen, B. 1996. Regression Models: Censored, Sample Selected or Truncated Data. Thousand Oaks, Ca.: Sage Publications, London.

Briassoulis, H. 2000. Analysis of land use change: theoretical and modeling approaches. In: Loveridge S. (ed.), The Web Book of Regional Science West Virginia University, Morgantown.

Bruinsma, J. 2009. The resource outlook to 2050: By how much do land, water and crop yields need to increase by 2050?. In How to feed the World in 2050. Proceedings of a technical

meeting of experts, Rome, Italy, 24-26 June 2009. (pp. 1-33). Food and Agriculture Organization of the United Nations (FAO).

Dai, A. 2011a. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. Journal of Geophysical Research: Atmospheres (1984–2012), 116(D12).

Dai, A. 2011b. Drought under global warming: a review. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 45-65.

Dormann, C., M McPherson, J., B Araújo, M., Bivand, R., Bolliger, J., Carl, G., ... & Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30(5), 609-628.

Ekström, M., Jones, P. D., Fowler, H. J., Lenderink, G., Buishand, T. A., & Conway, D. (2007). Regional climate model data used within the SWURVE project–1: projected changes in seasonal patterns and estimation of PET. *Hydrology and Earth System Sciences*, *11*(3), 1069-1083.

Erb, K. H., Haberl, H., Jepsen, M. R., Kuemmerle, T., Lindner, M., Müller, D., ... & Reenberg, A. 2013. A conceptual framework for analysing and measuring land use intensity. Current Opinion in Environmental Sustainability. doi: 10.1016/j.cosust.2013.07.010

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Tóth, G., van Velthuizen, H., ... & Wiberg, D. 2012. Global Agro-Ecological Zones (GAEZ v3. 0): Model Documentation. International Institute for Applied systems Analysis (IIASA), Laxenburg. Rome, Italy: Austria and the Food and Agriculture Organization of the United Nations (FAO) (Data accessible from http://www.fao.org/nr/gaez/en/).

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... & Snyder, P. K. 2005. Global consequences of land use. Science, 309(5734), 570-574.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., ... & Zaks, D. P. 2011. Solutions for a cultivated planet. Nature, 478(7369), 337-342.

Friedman, J., Hastie, T., & Tibshirani, R. 2010. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. 2007. Pathwise coordinate optimization. The Annals of Applied Statistics, 1(2), 302-332.

Fuchs, R., Herold, M., Verburg, P., Clevers, J., & Eberle, J. 2014. Gross changes in reconstructions of historic land use for Europe during the 20th century. Global Change Biology (In press).

Geist HJ, McConnell WJ, Lambin EF, Moran E, Alves D, Rudel TK. 2006. Causes and trajectories of land use/cover change. In: Land use and Land-Cover Change. Local Processes and Global Impacts (eds Lambin EF, Geist HJ), pp. 41–70, Springer-Verlag, Berlin Heidelberg.

Goldewijk, K. K., Beusen, A., & Janssen, P. 2010. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. The Holocene, 20(4), 565-573.

Golub, A., Hertel, T., & Sohngen, B. 2008. Land use modeling in recursively-dynamic GTAP framework (No. 2609). GTAP Working paper NO. 48. Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.

Gouel, C., & Hertel, T. W. 2006. Introducing Forest Access Cost Functions into a General Equilibrium Model. GTAP Research Memorandum, 8.

Green, R. E., Cornell, S. J., Scharlemann, J. P., & Balmford, A. 2005. Farming and the fate of wild nature. Science, 307(5709), 550-555.

Hallgren, W., Schlosser, C. A., Monier, E., Kicklighter, D., Sokolov, A., & Melillo, J. 2013. Climate impacts of a large-scale biofuels expansion. Geophysical Research Letters, 40(8), 1624-1630.

Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H. 2013. Updated high-resolution grids of monthly climatic observations. Int. J. Climatology. doi: 10.1002/joc.3711

Havlik, P., Schneider, U. A., Schmid, E., Böttcher, H., Fritz, S., Skalský, R., ... & Obersteiner, M. 2011. Global land use implications of first and second generation biofuel targets. Energy Policy, 39(10), 5690-5702.

Heistermann, M., Müller, C., & Ronneberger, K. 2006. Land in sight?: Achievements, deficits and potentials of continental to global scale land use modeling. Agriculture, Ecosystems & Environment, 114(2), 141-158.

Hibbard, K., Janetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Betts, R., ... & Feddema, J. J. 2010. Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling. International Journal of Climatology, 30(13), 2118-2128.

Houghton RA. 2008. Carbon flux to the atmosphere from land use changes: 1850- 2005. In: TRENDS: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Hubbard, K. G., Stooksbury, D. E., Hahn, G. L., & Mader, T. L. 1999. A climatological perspective on feedlot cattle performance and mortality related to the temperature-humidity index. Journal of production agriculture, 12(4), 650-653.

Hunter, J. E., & Hamilton, M. A. 2002. The advantages of using standardized scores in causal analysis. Human Communication Research, 28(4), 552-561.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., ... & Wang, Y. P. 2011. Harmonization of land use scenarios for the period 1500–2100: 600 years of global gridded annual land use transitions, wood harvest, and resulting secondary lands. Climatic Change, 109(1-2), 117-161.

Irwin, E. G., & Geoghegan, J. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems & Environment, 85(1), 7-24.

Jain, A. K., Meiyappan, P., Song, Y., & House, J. I. 2013. CO2 Emissions from Land-Use Change Affected More by Nitrogen Cycle, than by the Choice of Land Cover Data. Global change biology, 19, 2893-2906.

James, F. C., & McCulloch, C. E. 1990. Multivariate analysis in ecology and systematics: panacea or Pandora's box?. Annual Review of Ecology and Systematics, 21, 129-166.

Keys, E., & McConnell, W. J. 2005. Global change and the intensification of agriculture in the tropics. Global Environmental Change, 15(4), 320-337.

Kindermann, G., Obersteiner, M., Sohngen, B., Sathaye, J., Andrasko, K., Rametsteiner, E., ... & Beach, R. 2008. Global cost estimates of reducing carbon emissions through avoided deforestation. Proceedings of the National Academy of Sciences, 105(30), 10302-10307.

Klein Goldewijk, K., & Verburg, P. H. 2013. Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set. Landscape ecology, 28(5), 861-877.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography, 20(1), 73-86.

Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P. H., Estel, S., ... & Reenberg, A. 2013. Challenges and opportunities in mapping land use intensity globally. Current Opinion in Environmental Sustainability.

Lambin, E. F., & Meyfroidt, P. 2011. Global land use change, economic globalization, and the looming land scarcity. Proceedings of the National Academy of Sciences, 108(9), 3465-3472.

Lambin, E. F., Geist, H. J., & Lepers, E. 2003. Dynamics of land use and land-cover change in tropical regions. Annual review of environment and resources, 28(1), 205-241.

Lambin, E. F., Rounsevell, M. D. A., & Geist, H. J. 2000. Are agricultural land use models able to predict changes in land use intensity?. Agriculture, Ecosystems & Environment, 82(1), 321-331.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., ... & Xu, J. 2001. The causes of land use and land-cover change: moving beyond the myths. Global environmental change, 11(4), 261-269.

Lancaster, T. 2000. The incidental parameter problem since 1948. Journal of econometrics, 95(2), 391-413.

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., ... & Thornton, P. E. 2012. Simulating the Biogeochemical and Biogeophysical Impacts of Transient Land Cover Change and Wood Harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. Journal of Climate, 25(9).

Le Quéré, C., Peters, G. P., Andres, R. J., Andrew, R. M., Boden, T. A., Ciais, P., ... & Zaehle, S. 2014. Global carbon budget 2013. *Earth System Science Data*, *6*(1), 235-263.

Lesschen, J. P., Verburg, P. H., & Staal, S. J. 2005. Statistical methods for analysing the spatial dimension of changes in land use and farming systems. International Livestock Research Institute.

Letourneau, A., Verburg, P. H., & Stehfest, E. 2012. A land use systems approach to represent land use dynamics at continental and global scales. Environmental Modelling & Software, 33, 61-79.

Lotze-Campen, H., Popp, A., Beringer, T., Müller, C., Bondeau, A., Rost, S., & Lucht, W. 2010. Scenarios of global bioenergy production: The trade-offs between agricultural expansion, intensification and trade. Ecological Modelling, 221(18), 2188-2196.

Lubowski, R. N., Plantinga, A. J., & Stavins, R. N. 2008. What drives land use change in the United States? A national analysis of landowner decisions. Land Economics, 84(4), 529-550.

MEA. 2005. Ecosystems and human well-being. Vol. 5. Washington, DC: Island Press, 2005.

Meiyappan, P., & Jain, A. K. 2012. Three distinct global estimates of historical land-cover change and land use conversions for over 200 years. Frontiers of Earth Science, 6(2), 122-139.

Melillo, J. M., Reilly, J. M., Kicklighter, D. W., Gurgel, A. C., Cronin, T. W., Paltsev, S., ... & Schlosser, C. A. 2009. Indirect emissions from biofuels: how important?. Science, 326(5958), 1397-1399.

Melton, J. R., & Arora, V. K. 2014. Sub-grid scale representation of vegetation in global land surface schemes: implications for estimation of the terrestrial carbon sink. Biogeosciences, 11(4), 1021-1036.

Mertens, K. C., Verbeke, L. P. C., Ducheyne, E. I., & De Wulf, R. R. (2003). Using genetic algorithms in sub-pixel mapping. *International Journal of Remote Sensing*, *24*(21), 4241-4247.

Meyfroidt, P., Lambin, E. F., Erb, K. H., & Hertel, T. W. (2013). Globalization of land use: distant drivers of land change and geographic displacement of land use. *Current Opinion in Environmental Sustainability*, *5*(5), 438-444.

Meyfroidt, P. 2012. Environmental cognitions, land change, and social–ecological feedbacks: an overview. Journal of Land Use Science, (ahead-of-print), 1-27.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., ... & Wilbanks, T. J. 2010. The next generation of scenarios for climate change research and assessment. Nature, 463(7282), 747-756.

Mu, J. E., McCarl, B. A., and Wein, A. M. 2013. Adaptation to Climate Change: changes in farmland use and stocking rate in the U.S. Mitigation and Adaptation Strategies for Global Change, 18 (6). Doi: 10.1007/s11027-012-9384-4.

NRC. 2014, Advancing Land Change Modeling: Opportunities and Research Requirements. Washington, DC: National Academy Press.

O'Neill, B. C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., & Zigova, K. 2010. Global demographic trends and future carbon emissions. Proceedings of the National Academy of Sciences, 107(41), 17521-17526.

OECD 2012. OECD Environmental Outlook to 2050: The Consequences of Inaction. Organisation for Economic Co-operation Development, Paris. doi: 10.1787/9789264122246-en

Olesen, J. E., & Bindi, M. 2002. Consequences of climate change for European agricultural productivity, land use and policy. European Journal of Agronomy, 16(4), 239-262.

Overmars, K. P., De Koning, G. H. J., & Veldkamp, A. 2003. Spatial autocorrelation in multi-scale land use models. Ecological Modelling, 164(2), 257-270.

O'Neill, B., and Verburg, P. 2012. Spatial land use modeling: simulating decisions and their consequences for climate, carbon, and water, Global Land Project (GLP) News Letter, Issue No 9, ISSN: 2316-3747.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. 2003. Multi-agent systems for the simulation of land use and land-cover change: a review. Annals of the Association of American Geographers, 93(2), 314-337.

Parry, M., Rosenzweig, C., Iglesias, A., Fischer, G., & Livermore, M. 1999. Climate change and world food security: a new assessment. Global environmental change, 9, S51-S67.

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P., Fernandez-Manjarrés, J. F., ... & Walpole, M. 2010. Scenarios for global biodiversity in the 21st century. Science, 330(6010), 1496-1501.

Phalan, B., Onial, M., Balmford, A., & Green, R. E. 2011. Reconciling food production and biodiversity conservation: land sharing and land sparing compared. Science, 333(6047), 1289-1291.

Pielke, R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., ... & de Noblet, N. 2011. Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdisciplinary Reviews: Climate Change, 2(6), 828-850.

Pitman, A. J., de Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., ... & Voldoire, A. 2009. Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. Geophysical Research Letters, 36(14).

Portmann, F. T., Siebert, S., & Döll, P. 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochemical Cycles, 24(1).

Ramankutty, N. 2012. Global Cropland and Pasture Data from 1700-2007. Available online at [http://www.geog.mcgill.ca/~nramankutty/Datasets/Datasets.html] from the LUGE (Land Use and the Global Environment) Laboratory, Department of Geography, McGill University, Montreal, Quebec, Canada. [Accessed on 25 September 2012].

Ramankutty, N., & Foley, J. A. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), 997-1027.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochemical Cycles, 22(1).

Ramankutty, N., Foley, J. A., Norman, J., & McSweeney, K. 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. Global Ecology and Biogeography, 11(5), 377-392.

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., & Foley, J. A. 2012. Recent patterns of crop yield growth and stagnation. Nature communications, 3, 1293.

Reilly, J., Melillo, J., Cai, Y., Kicklighter, D., Gurgel, A., Paltsev, S., ... & Schlosser, A. 2012. Using land to mitigate climate change: hitting the target, recognizing the trade-offs. Environmental science & technology, 46(11), 5672-5679.

Rokityanskiy, D., Benítez, P. C., Kraxner, F., McCallum, I., Obersteiner, M., Rametsteiner, E., & Yamagata, Y. 2007. Geographically explicit global modeling of land use change, carbon sequestration, and biomass supply. Technological Forecasting and Social Change, 74(7), 1057-1082.

Ronneberger, K., Berrittella, M., Bosello, F., & Tol, R. S. 2009. KLUM@ GTAP: Introducing biophysical aspects of land use decisions into a computable general equilibrium model a coupling experiment. Environmental Modeling & Assessment, 14(2), 149-168.

Ronneberger, K., Tol, R.S.J., Schneider, U.A. 2005. KLUM: A simple model of global agricultural land use a a coupling tool of economy and vegetation. Working Paper FNU-65. Hamburg University, Hamburg, Germany.

Rounsevell, M. D. A., Arneth, A., Alexander, P., Brown, D. G., de Noblet-Ducoudré, N., Ellis, E., Finnigan, J., Galvin, K., Grigg, N., Harman, I., Lennox, J., Magliocca, N., Parker, D., O'Neill, B. C., Verburg, P. H., and Young, O. 2014. Towards decision-based global land use

models for improved understanding of the Earth system, Earth Syst. Dynam., 5, 117-137, doi:10.5194/esd-5-117-2014.

Rounsevell, M. D., & Arneth, A. 2011. Representing human behaviour and decisional processes in land system models as an integral component of the earth system. Global Environmental Change, 21(3), 840-843.

Sarofim, M. C., & Reilly, J. M. 2011. Applications of integrated assessment modeling to climate change. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 27-44.

Schaldach, R., Priess, J. A., & Alcamo, J. 2011. Simulating the impact of biofuel development on country-wide land use change in India. Biomass and Bioenergy, 35(6), 2401-2410.

Schlenker, W., & Roberts, M. J. 2006. Nonlinear effects of weather on corn yields. Applied Economic Perspectives and Policy, 28(3), 391-398.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., and Feick, S. 2005. Development and validation of the global map of irrigation areas, Hydrol. Earth Syst. Sci., 9, 535-547, doi:10.5194/hess-9-535-2005.

Smith, P., Gregory, P. J., Van Vuuren, D., Obersteiner, M., Havlík, P., Rounsevell, M., ... & Bellarby, J. 2010. Competition for land. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 2941-2957.

Souty, F., Brunelle, T., Dumas, P., Dorin, B., Ciais, P., Crassous, R., ... & Bondeau, A. 2012. The Nexus Land-Use model version 1.0, an approach articulating biophysical potentials and economic dynamics to model competition for land-use. *Geoscientific Model Development*, (1), 1297-1322.

Souty, F., Dorin, B., Brunelle, T., Dumas, P., & Ciais, P. 2013. Modelling economic and biophysical drivers of agricultural land use change. Calibration and evaluation of the Nexus Land use model over 1961–2006. Geoscientific Model Development Discussions, 6(4), 6975-7046.

Stehfest, E., Bouwman, L., van Vuuren, D. P., den Elzen, M. G., Eickhout, B., & Kabat, P. 2009. Climate benefits of changing diet. Climatic change, 95(1-2), 83-102.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society, 93(4), 485-498.

Tebaldi, C., Hayhoe, K., Arblaster, J. M., & Meehl, G. A. 2006. Going to the extremes. Climatic Change, 79(3-4), 185-211.

TEEB. 2010. The Economic of Ecosystems and Biodiversity: Mainstreaming the Economics of Nature: a Synthesis of the Approach, Conclusions and Recommendations to the TEEB.

Thenkabail, P. S., Biradar, C. M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., ... & Dutta, R. 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. International Journal of Remote Sensing, 30(14), 3679-3733.

Tilman, D., Balzer, C., Hill, J., & Befort, B. L. 2011. Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences, 108(50), 20260-20264.

U.S. Environmental Protection Agency (EPA). 2000. Projecting land use change: A summary of models for assessing the effects of community growth and change on land use patterns. Office of Research and Development Publication EPA/600/R-00/098. Cincinnati, OH: U.S. Environmental Protection Agency.

UNEP. 2012. Global Environmental Outlook: Environment for the future we want, GEO 5. United Nations Environmental Programme.

van Asselen, S., & Verburg, P. H. 2012. A Land System representation for global assessments and land use modeling. Global Change Biology, 18(10), 3125-3148.

van Asselen, S., & Verburg, P. H. 2013. Land cover change or land use intensification: simulating land system change with a global-scale land change model. Global Change Biology. doi: 10.1111/gcb.12331

van Vuuren, D. P., Bayer, L. B., Chuwah, C., Ganzeveld, L., Hazeleger, W., van den Hurk, B., ... & Strengers, B. J. 2012. A comprehensive view on climate change: coupling of earth system and integrated assessment models. Environmental Research Letters, 7(2), 024012.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., ... & Rose, S. K. 2011. The representative concentration pathways: an overview. Climatic Change, 109(1-2), 5-31.

Verburg, P. H., Ellis, E. C., & Letourneau, A. 2011. A global assessment of market accessibility and market influence for global environmental change studies. Environmental Research Letters, 6(3), 034019.

Verburg, P. H., Schot, P. P., Dijst, M. J., & Veldkamp, A. 2004. Land use change modelling: current practice and research priorities. GeoJournal, 61(4), 309-324.

Verburg, P.H., van Asselen, S., van der Zanden, E.H., Stehfest, E. 2013. The representation of landscapes in global scale assessments of environmental change. Landscape Ecology. doi: 10.1007/s10980-012-9745-0.

Wang, X. 2008. Impacts of Greenhouse Gas Mitigation Policies on Agricultural Land, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA. Available at globalchange.mit.edu/files/document/Wang_PhD_08.pdf (accessed July 28, 2013).

Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T., & Reick, C. (2014). Comparing the influence of net and gross anthropogenic land use and land cover changes on the carbon cycle in the MPI-ESM. *Biogeosciences Discussions*, *11*(4), 5443-5469.

Zou, H., & Hastie, T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

CHAPTER 6

Dynamics and determinants of land change in India: linking remote sensing to village socioeconomics

6.1 Abstract

The pressure on India's land resources for food, feed, and fuel far exceeds the global average. Despite, there are no national scale high-resolution remote-sensing estimates of land change in contemporary India. Here, we present national level estimates of land-cover conversions in India for 1985-1995-2005 (decadal), based on a wall-to-wall analysis of Landsat imageries. Further, we investigated the drivers of land-cover conversions that are most important at national scale be estimating spatial models between land-cover conversions and high-resolution spatial data on biophysical and socioeconomic factors. Our driver analysis is specific to land change process, and we estimated models at both national level and for sub-national hotspots.

Our results indicate massive conversions between cropland and fallow land between 1985 and 2005 indicating the low resilience of cropland in India. Our driver analysis at national scale indicates the high dependency of crop-fallow systems on monsoon and post-monsoon climate, labor migration, and access to market and irrigation facilities. These results emphasize the critical need to extension and better management of common-pool resources (e.g. irrigation in rain-fed areas) and critical support services to reduce under-utilized land.

We also find that India has experienced increased forest loss with time. Major drivers of forest loss at national scale between 1985 and 2005 were manufacturing of wooden furniture's/timber products, cattle/dairy/leather products (due to over-grazing), and manufacturing of wooden agricultural implements, mining/quarrying activities, and industrial development. We also find that colder and wetter regions and regions without electricity were also positively associated with forest loss, indicating over extraction of domestic fuel wood and building materials. Our analysis underscores the crucial need for forest policies to adopt a bottom-up approach by involving local communities and village panchayats to effectively implement afforestation programs, e.g. by planting tree species that benefit the local community.

6.2 Introduction

India occupies ~2.4% of the world's land area, but supports more than a sixth of the world's human and livestock population (Census of India 2011; Livestock Census, 2012). Over the last decade, India added 181.5 million people (Census of India 2011), roughly the population of Brazil. This high population to land ratio, coupled with socioeconomic development (Hubacek et al., 2007; World Bank Group, 2015; United Nations, 2014), has placed tremendous pressure on India's land resources for food, feed, and fuel, causing extensive environmental degradation (Table 6.1).

The pressure on India's land resources is expected to further intensify in the future, in the light of growing population and economy (World Bank Group, 2015; United Nations, 2014, 2015), and climate change (Lobell et al., 2008, 2012; Aggarwal, 2008; Auffhammer et al., 2012; Singh et al., 2002; O'Brien et al., 2004; World Bank, 2015). Therefore, a key challenge for land use planning in India is to maximize both food and livelihood security, and simultaneously minimize environmental degradation from land-use and land-cover change (LULCC). Land is closely tied to both food and livelihood security as over half of India's population is dependent on agriculture and allied sectors for livelihood (Census of India 2011). The vulnerability of Indian agriculture to climate change and climate variability can also impact the quality and sustainable use of land, considering one-fourth of India's population lives in poverty (Singh et al., 2002; O'Brien et al., 2004; World Bank, 2015). On the other hand, India being one of the ten most forest-rich nations of the world, has received increasing attention under the REDD+ mechanism to protect its forests to help mitigate climate change, preserve its rich biodiversity, and support ecosystem services (Agarwal et al., 2011; Ravindranath et al., 2012). For similar reasons, India's national forest policy aims to increase its forest cover from the existing ~21% of its total geographical area to a minimum of 33% (MoEF, 1988; Joshi et al., 2011). Addressing these competing challenges requires an understanding of the dynamics of LULCC at national scale.

In this study, we quantified land-cover conversions (complete replacement of one land cover by another) at national scale using a wall-to-wall analysis of high-resolution (~23.5m) Landsat-MSS/TM imageries at decadal time intervals (1985-95-05). Landsat imageries are crucial to capture the changes within the highly heterogeneous and fragmented landscapes of

cropland and forest typical of India (Figure S1; Reddy et al., 2013; Jain et al., 2013; Roy et al., 2015; Banger al., 2013; Gilbert, 2012; Ravindranath et al., 2014; Puyravaud et al., 2010; Pandit et al., 2007). We further investigated the causes of land-cover conversions that are most significant at national scale, by estimating spatial models between land-cover conversions and hypothesized socioeconomic and biophysical factors. We augment our statistical analysis through synthesis of 102 case studies that incorporates field knowledge (e.g. social surveys, local expertise) on the causes of LULCC. Importantly, our study covers the period of economic liberalization in India (1991 onwards) following which the pressure on land resources intensified.

Our analysis advances existing satellite-based assessments of LULCC in India on three aspects. First, this is the first study to quantify land-cover conversions at national scale. We used consistent data and method to map LULCC over time thereby minimizing errors in change detection (see methods). Our maps have an overall accuracy of 95%, thus proving only accurate and reliable information on LULCC. Earlier high-resolution land cover mapping activities at national scale was mostly one-time effort (Roy et al., 2015)–hence, unavailable at regular time intervals; their project-specific thematic classification, and varying data quality make compilation of consistent time-series imageries difficult. Notably, India monitors forest cover and trees outside forest bi-annually (FSI, 2013), but not forest conversions. The land cover replaced or preceded by forest has varying environmental impacts (Don et al., 2011; Mahmood et al., 2014). India also maps wasteland periodically (NRSC, 2011), but not wasteland conversions. The sources of wasteland and how reclaimed wastelands are used is crucial to plan and evaluate development efforts.

Second, we tracked conversions among 11 land classes that include fallow land (Table S2). As land is scarce in India, understanding the dynamics of under-utilized land (conversions between cropland and non-productive uses, specifically fallow land and wasteland) is vital to land use planning. Identifying fallow land requires analysis of satellite imageries over multiple seasons and years. We are unaware of any high-resolution global or national remote sensing product (e.g. Chen et al., 2015; Roy et al., 2015) that differentiates between cropland and fallow land.

Third, to investigate the causes of LULCC, we compiled the most detailed (~630,000 villages; Figure S2) national level spatial database on over 200 socioeconomic variables for two census years (1991 and 2001). Earlier national level studies for India (Roy et al., 2015) were limited to land cover mapping, hindered by limited and coarse spatial data on socioeconomics at national scale. For example, the best spatial data on population is at sub-district level (~5500 political units) (CIESIN, 2005), and other common variables at either district (~600 units) or state level from scattered sources. Our new socioeconomic data is commensurate with Landsat, and provides a 100-fold and 1000-fold improvement in spatial detail compared to sub-district and district data respectively.

We focused on three broad LULCC: 1. under-utilization of cropland area, 2. deforestation and forest degradation, and 3. increase in forest area. Our focus on these three LULCC is supported by two reasons: 1. as discussed above, they are central to land use planning, and 2. our analysis indicates they are both widespread and most significant at national scale, collectively accounting for 55-60% of all area conversions during both decades.

6.3 Methods and data

Here, we describe our methods briefly. See supplementary text for further details.

Data

Table S11 summarizes key datasets used with references. Our village-level spatial data on socioeconomics is new to this study. We highlight socioeconomic and LULUC data, both of which are central to our analysis.

Our socioeconomic database spans over 200 variables at the highest level of spatial disaggregation (~630,000 villages and towns) from two consecutive censuses (1991 and 2001). We created the spatial database by combining tabular information from the Indian census (each household is surveyed and aggregated to village/town level) with seamless village and town level administrative boundaries of India prepared as a part of this study, sourced from Survey of India toposheets. Apart from high spatial detail (Figure S2), our data provides important village-specific information undecipherable at coarser spatial level. For example, we included village-specific primary occupation(s) which reflect the base of the socio-economic culture prevalent in

rural parts of India. Our analysis can therefore offer key insights on the effects of rural livelihood on LULCC. We also observe high granularity in some socioeconomic variables that can get masked at coarser resolution (Figure S28).

We have a dedicated article describing the methodology and validation of the LULCC database, with basic land cover area statistics (Roy et al., 2015). This study presents detailed land conversion analysis of the LULCC database. The data was prepared using Landsat-MSS/TM imageries, supplemented by IRS 1C-LISS II, and Resourcesat 1-LISS III images. Our data has ~23.5m resolution, with features mapped at 1:50,000 scale. We interpreted the satellite data using on-screen visual interpretation technique for two decades (1985-1995-2005). Manual interpretation of detailed Landsat images is time-consuming. Therefore, studies with large spatial coverage, typically interpret Landsat images on sampling basis, representative of the study region (e.g. Gibbs et al., 2010). In contrast, our analysis is a wall-to-wall mapping effort at national scale. Our data has been extensively validated with over 12600 field points, and has an overall accuracy of 95% (class accuracy between 87% and 100%).

To minimize errors in land change detection between 2005 and 1995, we overlaid 1995 Landsat imagery over 2005 map and traced boundaries where land change had occurred, leaving unchanged boundaries unmodified. In this way, we minimized human-errors that could have otherwise occurred if both 2005 and 1995 maps were visually-interpreted independently and land change were inferred by differencing the two maps. We followed similar approach to detect land change between 1985 and 1995, using 1995 map as reference.

Analysis

We report LULCC estimates at national (Tables S2-S5) and state level (Table S12; Dataset S1), and by agro-ecological zones (AEZs) (Dataset S2) considering their policy relevance to forest and agriculture (see text S2 for rationale). AEZs are regions delineated by similar climatic and soil conditions. We estimated statistical models specific to land-cover conversion, at both national level and for regional hotspots (identified by AEZs). In Indian context, AEZs are the optimal units for macro-level land use planning and efficient transfer of technology, as India's economy is highly dependent on agriculture and allied sectors (including forestry).

Statistical estimation

Our statistical analysis between land-cover conversions patterns (dependent variable) and their concomitant socioeconomic and biophysical factor (or their proxies; independent variable) is carried out at 1km x 1km resolution and for each decade separately. The 1km resolution was mainly a tradeoff between the 23.5m LULCC data, and relatively coarser socioeconomic data (~2km x 2 km per village on average). To minimize loss of information, while aggregating the 23.5m LULUC data, we calculated the fraction of 1km grid-cell undergoing land-cover conversion, as opposed to pixel representation.

Our statistical modeling technique draws on our recent work (Meiyappan et al., 2014), and is common to land change modeling studies (text S2). We model the relationship between dependent and independent variables as a "fractional" binomial logit model. The model allows for fractional outcomes in dependent variables, consistent with our LULCC data aggregation technique. As our independent variables have different units and scale, we standardized all continuous variables using Z-score prior to estimation. We use a state of the art method, the elastic-net penalty to account for multicollinearity across independent variables. We used bootstrap resampling with 500 replicates, where we resampled the observations (grid-cells) and we fitted a new model to the data. The bootstrap, in addition to providing confidence intervals, also accounts for spatial autocorrelation typical to gridded LULCC datasets.

Synthesis of case studies

Our synthesis provides a bottom-up analysis on the causes of LULCC in India. We performed a systematic literature search on ISI Web of Science and Google Scholar for studies on LULCC covering India and our study period. We additionally included key (sub-) national reports, not indexed in either literature database. In total, we reviewed 643 articles, of which we discarded 177 as irrelevant (38 of which discussed drivers of LULCC processes not a focus of our study). Of the remaining 466 articles, over three-fourth focused only on land change detection, highlighting the relatively less attention on understanding the causes of change. The 102 articles in our synthesis provide information on the causes of land change typically by combining one or more of: household surveys, field transects, and regional/local expertise of authors. Often, studies also included remote-sensing component. The studies are summarized in

Table S7-10, and the study locations are visualized in Figure S29. Our hypothesized socioeconomic and biophysical factors for statistical estimation were grounded using synthesis literature.

6.4 Results

Under-utilization of Cropland Area

We find massive shifts between cropland and fallow land during both decades (Figure 6.1). About 35% (1985-95) and 46% (1995-05) of all areas that underwent land-cover conversion in India resulted from changes between cropland and fallow land, indicating the low resilience of cropland in India. Furthermore, India has consistently reclaimed ~10% of existing wasteland to cropland during each decade. These development efforts are however void considering the much larger amount of cropland being fallowed concurrently.

A spatial disaggregation (Figure 6.2) indicates that over 70% of shifts from cropland to fallow land and vice-versa are confined to five Agro-Ecological Zones (AEZs): the Western Plain, Kachchh, and part of Kathiawar Peninsula (AEZ2), and the semi-arid zones (AEZ4, 5, 6, and 8). These five zones also enclose over 90% of wasteland reclaimed during both decades, indicating reorganization between cropland and under-utilized land within each AEZ. The confinement of under-utilization of cropland to specific zones indicates efficient alignment of future development efforts is logistically feasible.

Causal analysis at national scale (Figures 6.3a, S4a) indicates monsoon and post-monsoon climate is important in determining under-utilization of cropland, echoing previous studies (e.g. Lobell et al., 2008; Kumar et al., 2004; Guiteras, 2009; Mall et al., 2006). Perhaps most importantly, our study indicates changing labor dynamics is an emerging factor in under-utilization of cropland. Post-liberalization (1991 onwards), we observe wide-spread reduction in main agricultural laborers (especially male) and male marginal cultivators (Figure S5), primarily driven by urbanization and better income opportunities (relatively less strenuous and more stable non-agricultural jobs). During 1995-05, we find areas converted from cropland to under-utilized land had substantially lower male main agricultural labor (AEZ 2) and male marginal cultivators (semi-arid hotspots) compared to counter-factual (buffer villages) (Figure S6b). We also find positive association between under-utilization of cropland and proportion of main female

cultivators, indicating imperfect labor market (i.e. male cultivators are more efficient in resource conservation).

Factors prominent in explaining under-utilization of cropland (Figures 6.3a, S4a, S6-S9), often were also prominent in explaining vice-versa conversion (i.e. land reclamation), but with opposite sign (Figures 6.3b, S4b, S10-S12). Our results broadly indicate that market accessibility (market frequency, proximity to town), knowledge to reclaim land (proportion of literate population, educational facilities, access to information), and ability of invest (income, credit societies) are crucial to reducing under-utilization of cropland.

We find varying regional dynamics with respect to average farm size. In AEZ2 larger plots are prone to under-utilization because resources (capital, agricultural labor, and irrigation) are a limiting factor to fuller utilization of land area (Figure S6). The massive reclamation of under-utilized land to cropland in AEZ2 (1995-05) is primarily from improvements in tube well and well irrigation leading to diversification to water-intensive cash crops (e.g. cotton; see Figures S12-S14) in small areas converting some lands fallow. In semi-arid hotspots, we find smaller plots are prone to under-utilization because of low technical efficiency (smaller plots are uneconomical to mechanize) (Figures S7-S9). These zones show positive association between under-utilization of cropland and lack of irrigation infrastructure, lack of male cultivators, and soil degradation and salinization indicating poor land management (small plots are generally more intensive used).

Deforestation and Forest Degradation

During 1985-95, India lost ~3.5% (~27150 km^2) of the existing forest area in 1985 (~764100 km^2), and the rate increased to ~3.8% during 1995-05 (~28350 km^2 loss of ~745100 km^2 forest in 1995) (Figure 6.1). Cropland was the major source of forest conversion during both decades, contributing to over 47% of forest loss in 1985-95, and 41% during 1995-05. The relative area of forest lost to shrubland increased from ~29% in 1985-95 to ~32% in 1995-05. About 7% of forest loss during both decades is attributable to commercial plantations. These trends are in stark contrast with the 1988 National Forest Policy that regards forest as a national asset, and imposed strict rules to protect them (MoEF, 1988; Joshi et al., 2011).

A regional breakdown (Figure 6.2) indicates that forest loss is widespread across India, and the hotspots change over time. For example, in AEZ19 that enclose the Western Ghats (biodiversity hotspot), 6.8% of the regions forest area in 1985 was converted to other land use by 1995 (35% each to shrubland and plantation, and 23% to cropland). In subsequent decade, the region's forest area loss declined by half. In AEZ5, 4.9% of the regions forest area (in 1985) was converted to other land use by 1995, and the rate increased to 7.9% in subsequent decade. Nonetheless, Eastern Plateau and Eastern Ghats (AEZ12), Central Highlands (AEZ10), and Western Himalayas (AEZ14) emerged as persistent hotspots for both decades. About 59% (1985-95) and 56% (1995-05) of the total forest area diverted to cropland were confined to AEZ5, 10, 12, and 17. About 84% (1985-95) and 80% (1995-05) of the total forest area converted to shrubland were confined to AEZ4, 5, 10, 12 and 19. AEZ12 accounted for 40% (1985-95) and 35% (1995-05) of all forest area converted to shrubland.

National level causal analysis (Figures 6.4a, S15) show strong spatial dependence between forest area loss and village occupation. Villages with following occupation were prominently related to forest loss, compared to counter-factual (buffer villages): wooden furniture's/timber products, cattle/dairy/leather products (due to over-grazing), manufacturing of wooden agricultural implements, mining/quarrying activities, and industrial development. Colder and wetter regions and regions without electricity were also positively associated with forest loss, indicating over extraction of domestic fuel wood and building materials.

Both nationally (Figures 6.4a, S15) and across regional hotspots (Figures S16-S20) we find prominent positive association between forest loss and lack of irrigation, lack of agricultural credit societies, low income, small farms, proportion of marginal agricultural labor, and highly eroded agricultural soils, indicating that low agricultural productivity increases the pressure on adjoining forests. Most diversion of forest to cropland is encroachment, because national forest policy does not favor diversion of forest to non-forest, which requires prior approval from central government. Furthermore, we find the forest area diverted to cropland have not declined with time (Figure 6.1) indicating weak implementation of national forest policy.

We find prominent negative association of forest loss with steep slope (difficult to access), and protected areas. While land protection reduces forest loss, 9% (1985-95) and 7.6% (1995-05) of total forest loss has still occurred within protected areas, and 11.2% (1985-95) and

8.7% (1995-05) within 5km buffer from the perimeter of protected areas (critical to maintain the functionality of protected landscapes), indicating level of protection is important and has improved over time (Figure S21).

A regional analysis indicates that in Western Ghats (1985-95), mining activities, manufacturing of agricultural implements, and villages dependent on coconut and coffee plantations (encroachment) show positive association with forest loss (Figure S19). Across all hotspots in central India (AEZ5, 10, and 12), mining/quarrying activities, industrial development, and factors associated with low agricultural productivity (e.g. high erosion, small farms) show positive association to forest loss (Figures S16-S18). Other factors prominently associated to forest loss are wooden furniture's/timber extraction and cattle over-grazing (AEZ5; Figure S18); villages making bamboo products (AEZ12; Figure S17); villages making forest products (e.g. kendu leafs/beedi, leaf plates, baskets, brooms, match sticks, paper) (AEZ10; Figure S16); colder temperatures (extraction of firewood and building materials), wooden furniture's/timber, and making of woolen blankets (indicating sheep browsing) (AEZ14; Figure S20).

Increase in Forest Area

India recorded a positive trend in forest area gain over time (Figure 6.1). The gain in 1995-05 was 24% higher than the preceding decade, compensating for the increased forest loss during 1995-05. Overall, India experienced a net forest loss of over 18000 km^2 consistently during both decades (see text. S1). Reversion of cropland and shrubland together explain 65% (1985-95) and 78% (1995-05) of forest area gain. AEZ5, 10, and 12 were persistent hotspots of forest area gain in both decades (Figure 6.2); however, the magnitude was much smaller compared to forest area loss in the respective zones. The area of forest gain in AEZ10 more than doubled between decades, the increase mainly sourced from cropland. During 1995-05, substantial area of shrubland was recovered to forest in AEZ4, 5, and 12.

We find positive association between forest area gain and following variables (similar for both national and hotspots; Figures 6.4b, S22-S26): low male cultivators (especially marginal) indicating migration, smaller farms, lack of critical support services (irrigation and capital), and poor soils (characterized by one or more of: degradation, salinization, shallow depth, and low

cation exchange capacity). These relationships indicate abandonment of marginally productive cropland, followed by natural regeneration of pioneer forest tree species.

We also find positive association between forest area gain and protected areas, proportion of tribal population, and sacred groves. Tribes are culturally linked to forests and they are typically motivated by state forest department to jointly manage forest through protection, restoration of degraded forest, and enrichment plantations (notable exception of North-East India where tribes practice shifting cultivation). Sacred groves are protected by local community due to cultural/religious beliefs.

Across the three regional hotspots (AEZ5, 10, and 12), forest area gains were positively associated with state dummies, mined-out areas, density of forestry workers, and density of community workers. The states identified through state dummies typically have larger amount of forest inundated to water bodies (irrigation projects), and forest diverted to built-up land (e.g. roads, industries) (Figure S27). Both state dummies and greening of mined-out areas indicate compensatory afforestation by respective state governments to partly compensate for forest loss. Forestry workers are employed by forest department, and are a proxy for level of protection and control. These workers are typically involved in forest maintenance, wildlife protection, fire observations, and interface with tourism, among others. Community workers include presence of non-governmental organization that helps with restoration efforts in collaboration with forest department and local communities. Restoration typically includes greening firewood, fodder, and planting species that benefit local communities.

6.5 Discussion

Evaluation of Results and Caveats

A general understanding of the dynamics of LULCC over larger regions of India is limited, hindering effective national level planning and policy-making. While evidence-based research (e.g. social surveys) offers field-level insights, data collection is expensive and typically covers small regions. It is hard to generalize the dynamics of LULCC by studying few villages in a country of over 600,000 villages with diverse agro-ecological and socio-cultural environment. To this end, our analysis provides a comprehensive coverage by linking LULCC information from space, with rich and uniform socioeconomic data collected from each village and town in

the country. To evaluate and reinforce our model results, we synthesized 102 ground-level studies to identify accumulated effects that are statistical stronger than any individual study due to increased sample size and greater diversity (Table S7-S10).

Our synthesis indicates that all three LULCC are driven by different combinations of multiple factors. Nonetheless, the accumulated effects show commonalities that are generally congruent with our statistical analysis (Figure 6.5). Under-utilization of cropland is mainly driven by: new income opportunities, labor shortage, and lack of infrastructure and capital. Reclamation of under-utilized land depends mainly on critical support services, education, and infrastructure. Illegal forest encroachment, domestic use, human constructions, and industrial exploitation are common causes of forest loss. Forest increases mainly occur through government efforts in harmony with local community, and with increased level of protection.

Causal factors uncommon at national scale can be most important regionally. For example, all studies from North-East India report shifting cultivation as the main cause of forest loss (Table S9). Some factors also behave differently in individual cases. For example, studies stemming from same AEZ report opposing effects on how education affects under-utilization of cropland (Tables S7, S8). Education causes a shift to off-farm jobs, thus increasing under-utilized land. In contrast, with education farmers perceive higher returns to investment on land, invest more on resource conservation, and have better access to information leading to fuller land utilization. Such heterogeneity is concurrent and important to recognize; in such cases, our statistical analysis covering the entire region helps identify the dominant effect.

Two caveats are in order. First, as we estimated LULCC from decadal Landsat imageries, they capture only the decadal changes in LULCC, and can mask within-decade variations including intermediary land uses. Especially, inter-annual climate variability causes fluctuations in fallow land. However, the conversions between cropland and fallow inferred between decadal end points will reflect only the climate-effect of end point (see Text S2). We also do not capture land fallowed as a part of multiple cropping system used to restore and maintain soil fertility. Excluding cropland-fallow system, other land-cover conversions (e.g. forest to cropland) tend to be unidirectional at decadal time scale, due to high cost of land reversion (Gibbs et al., 2010; Pandey and Seto, 2015).

Second, both forest degradation and regrowth are gradual and cause subtle modifications to land cover. However, our data detects changes only when the magnitude of modification is large enough to cause shift from one land cover category to another. The resulting bias is likely minimal because: (1) persistent modification of forest would likely manifest as a change in land cover within a decade, and (2) our statistical estimation weighs each observation (grid cell) by the magnitude of land change; thus, small changes have less influence in our model.

Implications for Land Use Planning

India is amongst the world's fastest growing economy and population. Therefore, India needs to produce more food, and also protect its land resources. An optimal pathway to this challenge is to increase the area under forest, built-up land, and pastureland, and decrease the area under under-utilized land and cropland. Built-up land will increase with economic development (e.g. infrastructure for "Make in India" national program). Pastureland should increase to: (1) reduce animal pressure on forest, (2) meet increased diary demands with increasing disposable income and population, and (3) reduce protein deficiency. Under-utilized land should reduce to produce more food, and for afforestation (on wasteland). Currently, about 46% of India's land area is cropland (net area sown), of which only ~40% is cropped over once a year (DES, 2015). For comparison, China's net area sown is ~43% lower compared to India, but their total grain production is more than double of India (Nath et al., 2015), despite lower bio-physical potential. India therefore has higher potential to produce more food on existing cropland. Further, marginal cropland that increases soil erosion should be diverted for afforestation and development, while retaining fertile cropland. In comparison, our analysis shows that India is currently following a sub-optimal pathway (see Table S12 for state-wise summary).

Both public and private investments (e.g. through Integrated Wasteland Development Program) have resulted in consistent reclamation of wasteland to cropland. Concurrently, farmers have fallowed much larger areas of cropland, representing an undesired tradeoff to wasteland reclamation. Numerable social surveys have shown that Indian farmers invest more on protecting fertile cropland (Shiferaw et al., 2006; Maikjuri et al., 1997; Wani et al., 2011; Kuppannan and Devarajulu, 2009; Nüsser et al., 2012) than restoring degraded land. Therefore,

better orientation of investment portfolios with farmer's attitude can effectively help reduce under-utilized land.

Our results also emphasize the critical need to extension and better management of common-pool resources (e.g. irrigation in rain-fed areas) and critical support services to reduce under-utilized land. This will likely deter distress out-migration of farmers, and suppress a potential positive feedback because higher non-farm income will reduce the incentive to invest (capital and labor) on soil-water conservation furthering land degradation and migration.

Land fragmentation is an emerging challenge to reducing under-utilization of cropland in semi-arid zones of India. Small farms have low technical efficiency, and have increased risk of soil degradation (see Table S13 for AEZ-wise analysis). The problem is likely compounded in the future with increasing population, and further sub-division of households with two expected outcomes: (1) smaller land holdings will make farming increasingly uneconomical, and (2) increasing urban wage rates (economic development) will pull small and marginal farmers as urban laborers. Both outcomes are undesirable to reduce under-utilized land. There are two potential avenues to make agriculture a more economically viable exercise for farmers. First, India's land ceiling act of 1976 limits the maximum size of land holding that an individual/family can own to 4 ha to help reduce inequality in land ownership. However, the limit needs update to suit the current situation. Second, the 4 ha limit need not necessarily be a consolidated holding. Therefore, effective strategies to consolidate farmer's fragmented holdings as an operational unit will ensure higher productivity and economic returns.

Forestry in India should balance three needs: subsistence (fuel wood, forest produce, etc.), commerce (industry, mining, timber, paper etc.), and environmental protection. While there are no clear priorities assigned by Indian government to balance these needs, our analysis indicates that forest protection (viz. national forest policy) is likely subordinate to commerce. Land use planning should carefully evaluate the tradeoffs between the value of goods produced through exploiting forest land, and the long-term benefits of protecting forest ecosystems.

Currently, the livelihoods of ~173,000 villages in India are tightly linked to forests (Nayak et al., 2012). In the absence of other alternatives, the economic dependence of village communities on forests has increased forest depletion across many regions. The ongoing and

future planned privatization of afforestation programs in India tends to maximize corporate profits, with no space for community involvement. Our analysis underscores the crucial need for forest policies to adopt a bottom-up approach by involving local communities and village panchayats to effectively implement afforestation programs, e.g. by planting tree species that benefit the local community. There already exist best practices on forest management tested at community level in India (e.g. Nagendra, 2009; Kumar, 2015; Liser, 2000; Bhattacharya et al., 2010). However, these models need to be up-scaled, ingrained as policy, and integrated with implementation system through capacity building and technology upgrades. Decentralization coupled with policies aimed at increasing non-forest income would further reduce the pressure on forests.

Acknowledgements

This work contributes to the NASA Land-Cover/Land-Use Change Program (No: NNX414AD94G).

6.6 Tables

Table 6.1 Comparison by numbers: The role of land-use and land-cover change (LULCC) on key environmental problems compared between World and India for present/past period. See Table S1 for details.

Environmental Problem	Role of LULCC	
	World	**India**
Human land use	55% of land area	83% of land area
Climate change	20-24% of GHG emissions	25-30% of GHG emissions
Biodiversity loss	14% of species richness	22% of species richness
Land degradation	8-41% of land area	~57% of land area
Water use for agriculture	70% of withdrawal	91% of withdrawal
Nutrient excess in crops (Water pollution)	56% of nitrogen; 48% of phosphorous	74% of nitrogen; 71% of phosphorous

6.7 Figures

Figure 6.1 Gains, losses, and net changes in land use and land cover areas at aggregate national scale for the two decades (km^2/decade): 1985-95 and 1995-05. "Water bodies" include water bodies, aqua culture and permanent wetlands. "Others" include Salt Pan, Snow and Ice. Data from this figure is provided in Table S3-S5.

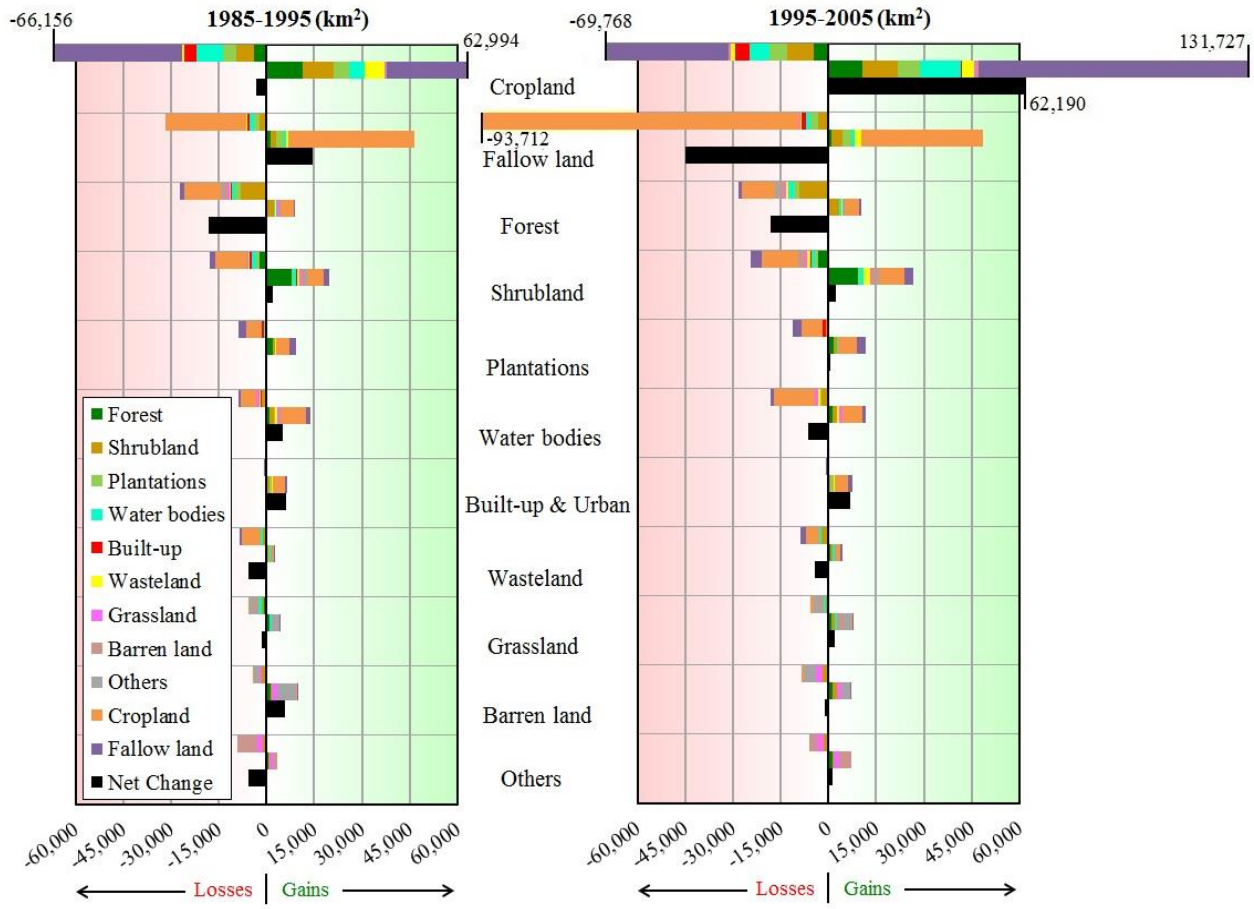


Figure 6.2 Spatial breakdown of the three major land-cover conversions: conversions between cropland and under-utilized land (UTL), forest gain, and forest loss. The size of circles is proportional to the magnitude of change within each 0.5º x 0.5º lat/long grid cells. The inset bar plot show the percent contribution by region to the national total (shown besides bar; units in x1000 km^2/decade and rounded to nearest integer). The regions are based on Agro-Ecological Zones (AEZs) of India (Table S6). See Figure S3 for a more detailed breakdown by AEZ.

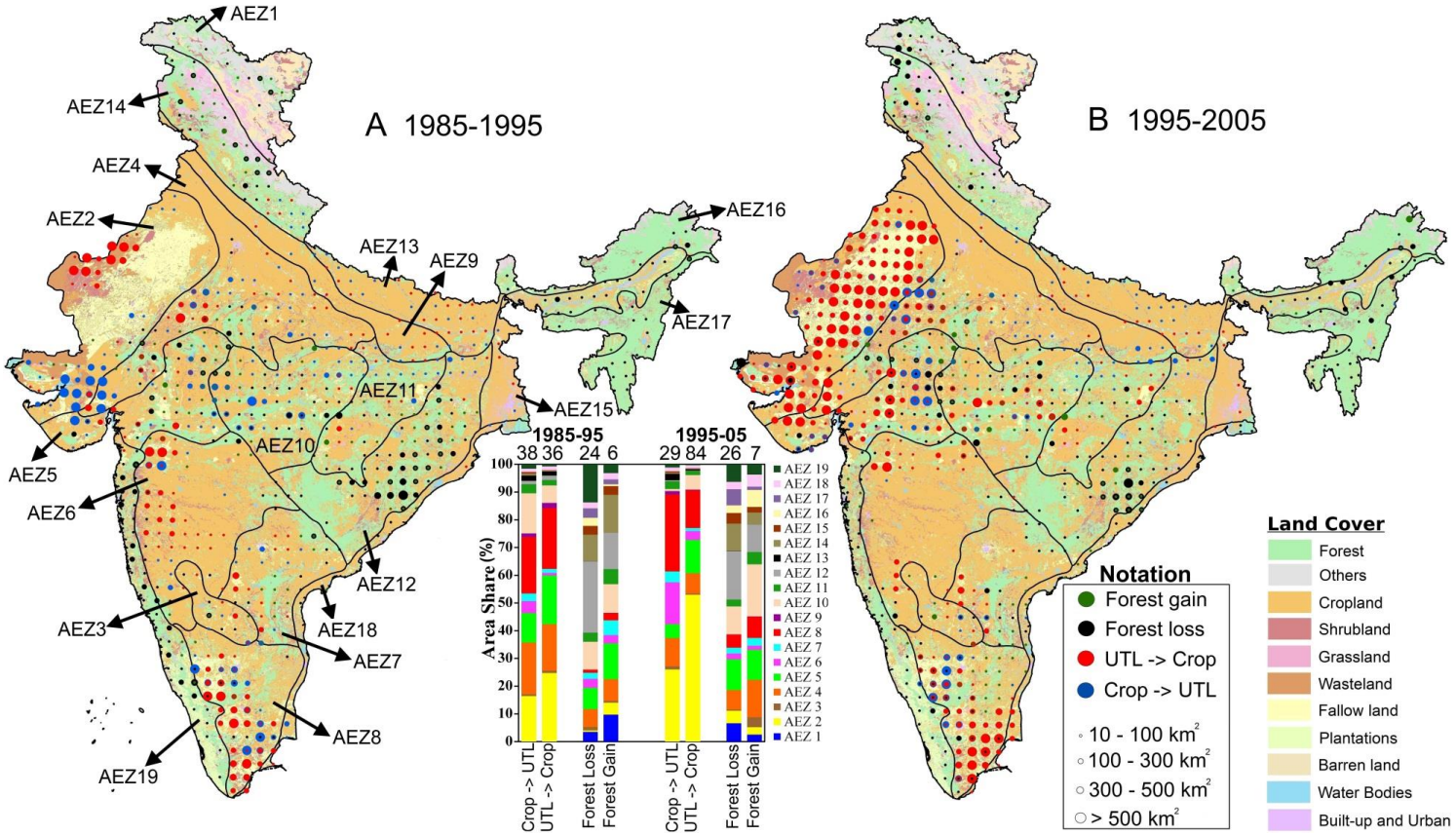


Figure 6.3 Factors most prominent in explaining: (a) conversion of cropland to under-utilized land at national scale (1995-05), and (b) vice-versa conversion i.e. conversion of under-utilized land to cropland at national scale (1995-05). The plots show the standardized regression coefficients of the ten most important variables (largest absolute mean estimates across coefficients) estimated using the spatial land change model (see methods). Results from bootstrap resampling with 500 replicates: central red line show mean estimate; error bars (blue) show 5% to 95% confidence interval; whiskers show 25% to 75% confidence interval. See Figure S4 for national-scale estimates corresponding to 1985-95.

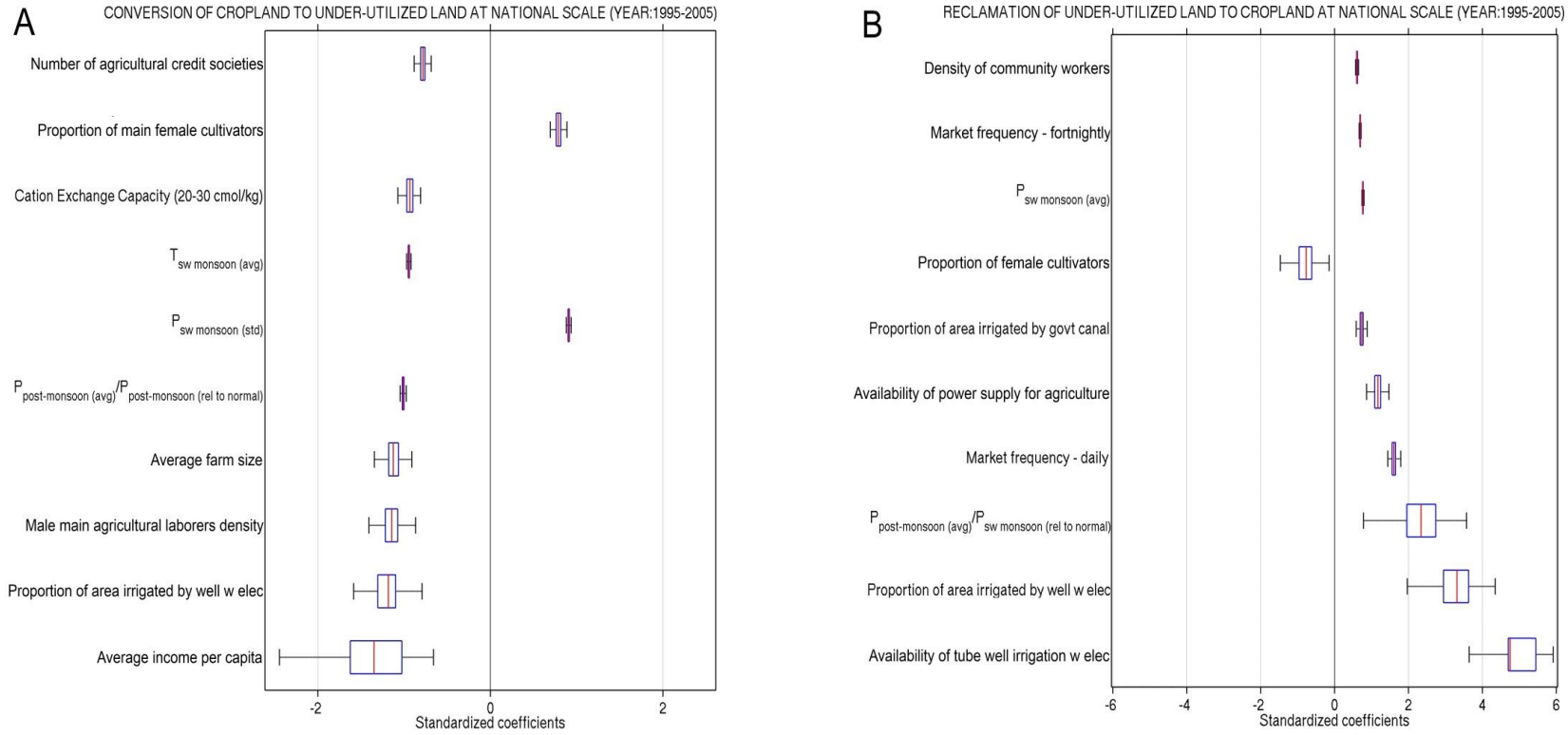
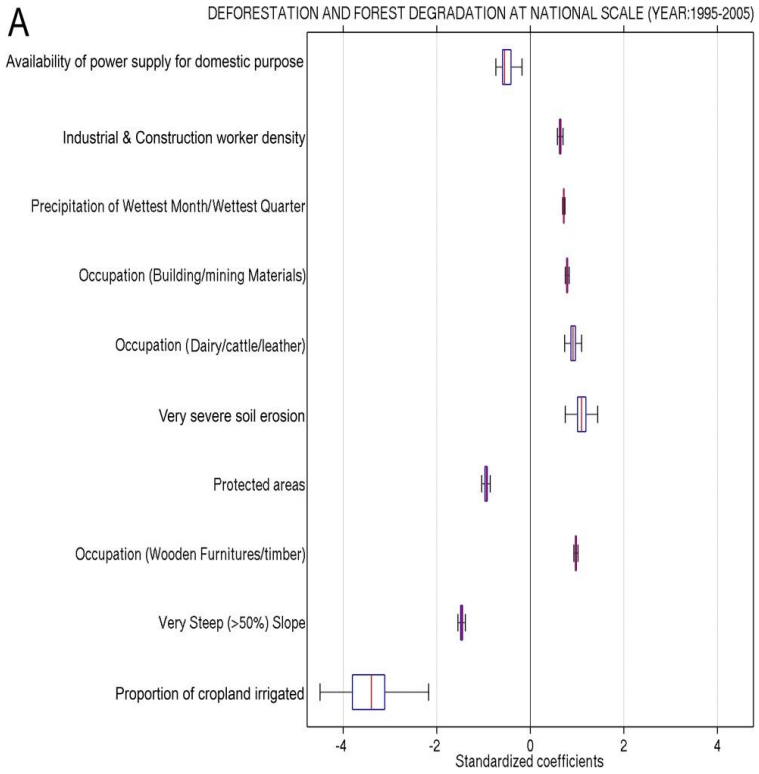


Figure 6.4 Similar to Figure 6.3, but for: (a) forest area loss at national scale (1995-05), and (b) forest area gain at national scale (1995-05). See Figure S13 for national-scale estimates corresponding to 1985-95.



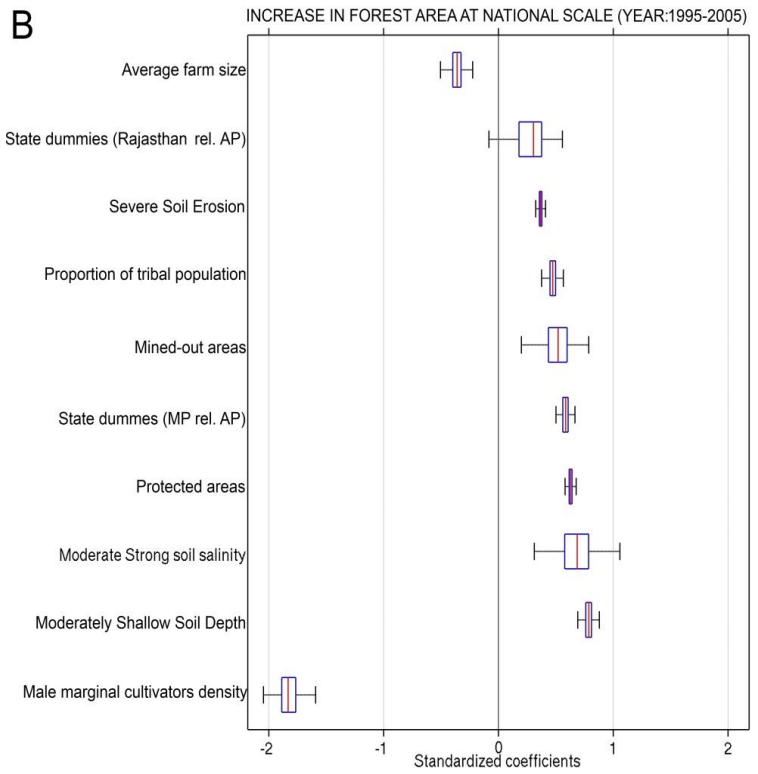


Figure 6.5 Frequency distribution of the causal factors identified from the synthesis of 92 case studies. (a) Conversions from cropland to under-utilized land and vice-versa, and (b) forest area loss and gains.

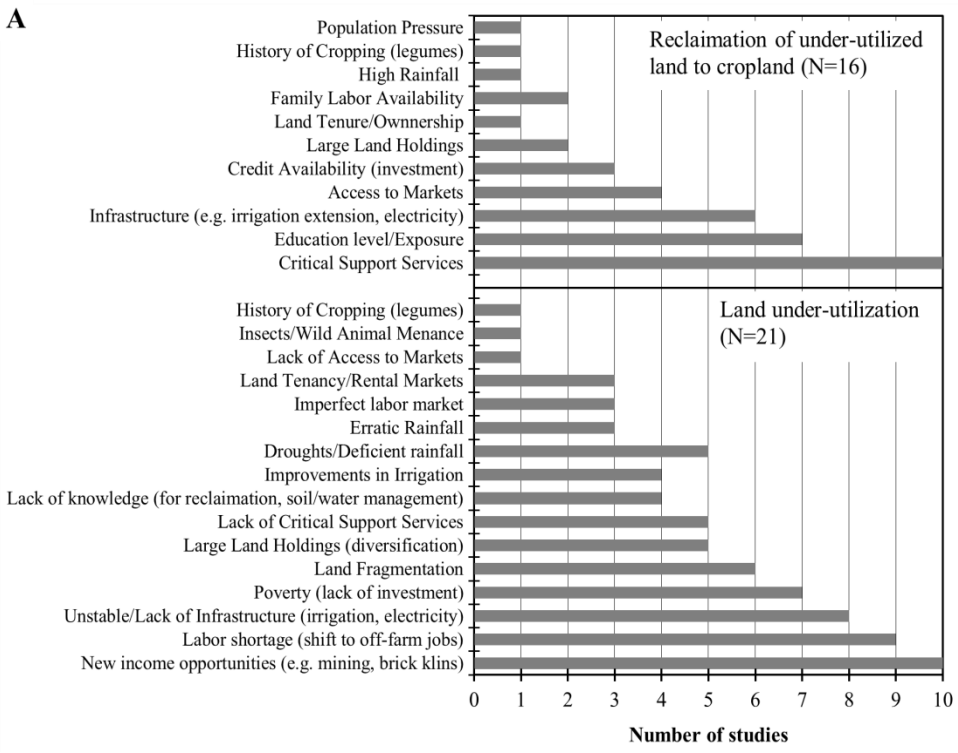
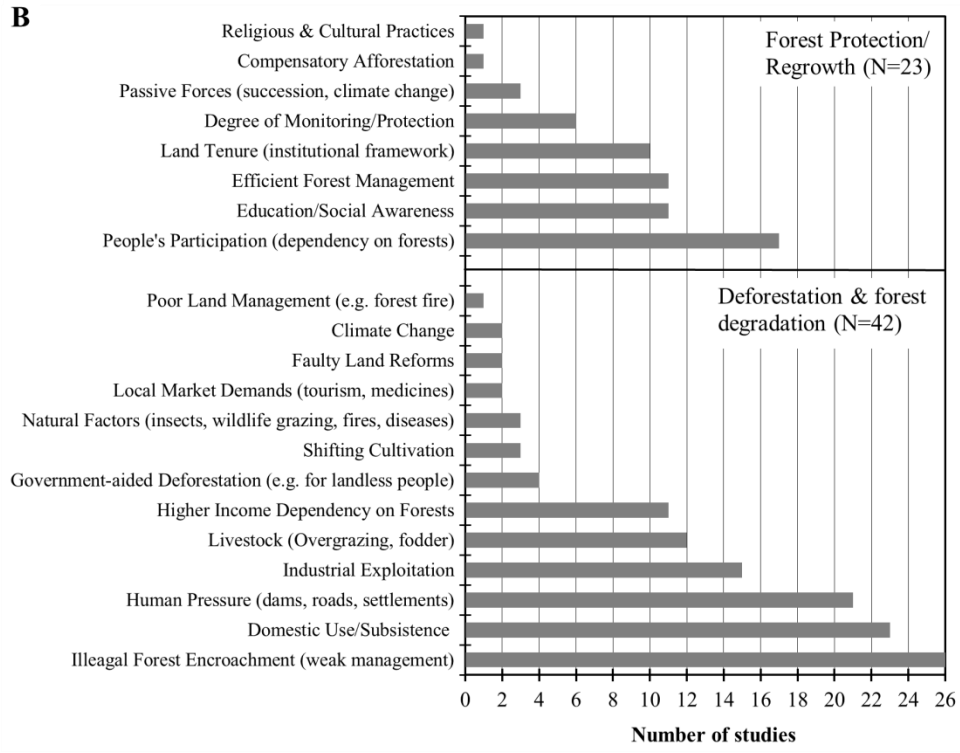


Figure. 6.5 (cont.)



6.8 References

Aggarwal, P. K. (2008). Global climate change and Indian agriculture: impacts, adaptation and mitigation. Indian Journal of Agricultural Sciences, 78(11), 911.

Agrawal, A., Nepstad, D., & Chhatre, A. (2011). Reducing emissions from deforestation and forest degradation. Annual Review of Environment and Resources, 36, 373-396.

Auffhammer, M., Ramanathan, V., & Vincent, J. R. (2012). Climate change, the monsoon, and rice yield in India. Climatic Change, 111(2), 411-424.

Banger, K., Tian, H., & Tao, B. (2013). Contemporary land cover and land use patterns in India estimated by different regional and global data sets. Journal of Land Use Science, (ahead-of-print), 1-13.

Bhattacharya, P., Pradhan, L., & Yadav, G. (2010). Joint forest management in India: Experiences of two decades. Resources, Conservation and Recycling, 54(8), 469-480.

Census of India, (2011). Census provisional population totals. Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, New Delhi, India. Accessible from http://censusindia.gov.in/ (Retrieved on 14 February 2015).

Center for International Earth Science Information Network (CIESIN), and Centro Internacional de Agricultura Tropical (CIAT). (2005). Gridded Population of the World, Version 3 (GPWv3), Palisades, NY: CIESIN, Columbia University.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., ... & Mills, J. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27.

Dilip Kumar, P. J. (2015). Between Panchayat, Community and the State: The Case for Joint Institutions for Managing Forests in India. Administrative Culture, 16(1), 4-23.

Directorate of Economics and Statistics (various years). Ministry Of Agriculture and Farmers Welfare, Government of India, New Delhi. http://eands.dacnet.nic.in (Accessed: May 13, 2015) .

Don, A., Schumacher, J., & Freibauer, A. (2011). Impact of tropical land-use change on soil organic carbon stocks–a meta-analysis. Global Change Biology, 17(4), 1658-1670.

FSI (2013). India State of Forest Report. Forestry Survey of India (Ministry of Environment and Forest). Dehradun, India.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proceedings of the National Academy of Sciences, 107(38), 16732-16737.

Gilbert, N. (2012). India's forest area in doubt. Nature, 489(7414), 14-15.

Guiteras, R. (2009). The impact of climate change on Indian agriculture. Manuscript, Department of Economics, University of Maryland, College Park, Maryland.

Hubacek, K., Guan, D., & Barua, A. (2007). Changing lifestyles and consumption patterns in developing countries: A scenario analysis for China and India. Futures, 39(9), 1084-1096.

Jain, M., Mondal, P., DeFries, R. S., Small, C., & Galford, G. L. (2013). Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. Remote Sensing of Environment, 134, 210-223.

Joshi, A. K., Pant, P., Kumar, P., Giriraj, A., & Joshi, P. K. (2011). National forest policy in India: Critique of targets and implementation. Small-Scale Forestry, 10(1), 83-96.

Krishna Kumar, K., Kumar, R. K., Ashrit, R. G., Deshpande, N. R., & Hansen, J. W. (2004). Climate impacts on Indian agriculture. International Journal of Climatology, 24(11), 1375-1393.

Kuppannan, P., & S. K. Devarajulu. (2009). Impacts of Watershed Development Programmes: Experiences & Evidences from Tamil Nadu. Munich Research Papers in Personal EconomicsPersonal Archive. http://mpra.ub.uni-muenchen.de/18653/

Lise, W. (2000). Factors influencing people's participation in forest management in India. Ecological Economics, 34(3), 379-392.

Livestock census, (2012). 19th Livestock Census-2012 All India report. Ministry of Agriculture, Department of Animal Husbandry, Dairying, and Fisheries. Government of India, New Delhi, India. Accessible from http://dahd.nic.in/

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. Science, 319 (5863), 607-610.

Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. Science, 319 (5863), 607-610.

Lobell, D. B., Sibley, A., & Ortiz-Monasterio, J. I. (2012). Extreme heat effects on wheat senescence in India. Nature Climate Change, 2(3), 186-189.

Mahmood, R., Pielke, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., ... & Fall, S. (2014). Land cover changes and their biogeophysical effects on climate. International journal of climatology, 34(4), 929-953.

Maikhuri, R. K., Semwal, R. L., Rao, K. S., & Saxena, K. G. (1997). Agroforestry for rehabilitation of degraded community lands: a case study in the Garhwal Himalaya, India. International Tree Crops Journal, 9(2), 91-101.

Mall, R. K., Singh, R., Gupta, A., Srinivasan, G., & Rathore, L. S. (2006). Impact of climate change on Indian agriculture: a review. Climatic Change, 78(2-4), 445-478.

Meiyappan, P., Dalton, M., O'Neill, B. C., & Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. Ecological Modelling, 291, 152-174.

Ministry of Environment and Forests (MoEF) of India. (1988). National Forest Policy (1988). Resolution No. 3-1/86-FP. New Delhi, India, Department of Environment, Forests and Wildlife.

Nagendra, H. (2009). Drivers of regrowth in South Asia's human impacted forests. Current Science, 97(11), 1586-1592.

Nath, R., Luan, Y., Yang, W., Yang, C., Chen, W., Li, Q., & Cui, X. (2015). Changes in Arable Land Demand for Food in India and China: A Potential Threat to Food Security. Sustainability, 7(5), 5371-5397.

National Remote Sensing Centre (2011) Wastelands Atlas of India 2011. Delhi: Deparment of Land Resources, Ministry of Rural Development.

Nayak, B. P., Kohli, P., & Sharma, J. V. (2012). Livelihood of local communities and forest degradation in India: Issues for REDD+. Ministry of Environment and Forests Government of India New Delhi India.

Nüsser, M., Schmidt, S., & Dame, J. (2012). Irrigation & development in the Upper Indus Basin: characteristics & recent changes of a socio-hydrological system in Central Ladakh, India. Mountain Research & Development, 32(1), 51-61.

O'Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., ... & West, J. (2004). Mapping vulnerability to multiple stressors: climate change and globalization in India. Global environmental change, 14(4), 303-313.

Pandey, B., & Seto, K. C. (2015). Urbanization and agricultural land loss in India: Comparing satellite estimates with census data. Journal of environmental management, 148, 53-66.

Pandit, M. K., Sodhi, N. S., Koh, L. P., Bhaskar, A., & Brook, B. W. (2007). Unreported yet massive deforestation driving loss of endemic biodiversity in Indian Himalaya. Biodiversity and Conservation, 16(1), 153-163.

Prasad, R., & Kant, S. (2003). Institutions, forest management, and sustainable human development–experiences from India. Environment, Development and Sustainability, 5(3-4), 353-367.

Puyravaud, J. P., Davidar, P. and Laurance, W. F. (2010). Cryptic loss of India's forests. Conserv. Lett., 3, 390–394.

Ravindranath, N. H., Murthy, I. K., Priya, J., Upgupta, S., Mehra, S., & Nalin, S. (2014). Forest area estimation and reporting: implications for conservation, management and REDD. Current Science, 106(9), 1201.

Ravindranath, N. H., Srivastava, N., Murthy, I. K., Malaviya, S., Munsi, M., & Sharma, N. (2012). Deforestation and forest degradation in India- implications for REDD+. Current Science102(8), 1117-1125.

Reddy, C. S., Sreelekshmi, S., Jha, C. S., & Dadhwal, V. K. (2013). National assessment of forest fragmentation in India: Landscape indices as measures of the effects of fragmentation and forest cover change. Ecological Engineering, 60, 453-464.

Roy, P. S., Behera, M. D., Murthy, M. S. R., Roy, A., Singh, S., Kushwaha, S. P. S., ... & Tripathi, S. K. (2015). New vegetation type map of India prepared using satellite remote sensing: Comparison with global vegetation maps and utilities. International Journal of Applied Earth Observation and Geoinformation, 39, 142-159.

Roy, P. S., Roy, A., Joshi, P. K., Kale, M. P., Srivastava, V. K., Srivastava, S. K., ... & Kushwaha, D. (2015). Development of Decadal (1985–1995–2005) Land Use and Land Cover Database for India. Remote Sensing, 7(3), 2401-2430.

Shiferaw, B. A., Reddy, V. R., Wani, S. P., & Rao, G. N. (2006). Watershed management & farmer conservation investments in the semi-arid tropics of India: analysis of determinants of resource use decisions & land productivity benefits. Journal of SAT Agricultural Research, 2(1), 1-25.

Singh, R. B., Kumar, P., Woodhead, T. (2002) Smallholder Farmers in India: Food Security and Agricultural Policy. FAO Regional Office for Asia and the Pacific, Bangkok, Thailand.

United Nations, Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352).

United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241.

Wani, S. P., Anantha, K. H., Sreedevi, T. K., Sudi, R., Singh, S. N., & D'Souza, M. (2011). Assessing the environmental benefits of watershed development: Evidence from the Indian semi-arid tropics. Journal of Sustainable Watershed Science & Management, 1(1), 10-20.

World Bank Group. (2015). Global Economic Prospects, June 2015: The Global Economy in Transition. Washington, DC: World Bank. doi: 10.1596/978-1-4648-0483-0. License: Creative Commons Attribution CC BY 3.0 IGO.

World Bank. (2015). A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals. Policy Research Report. Washington, DC: World Bank. doi:10.1596/978-1-4648-0361-1.

CHAPTER 7

Summary and future work recommendations

7.1 Summary

In this chapter, I provide an overall summary of the research carried out in my dissertation.

In Chapter 2, I characterized the historical land-use and land-cover change (LULCC) using annual maps of cropland, pastureland, wood harvest, and urban land as inputs. Due to uncertainties associated with estimates of historical land-use activities, I used three different data sets on agricultural extent to derive three different estimates, consistently using the same rule-based method of prioritizing and converting vegetation. I used information from remote sensing data to constrain and modify the rule-based method to implicitly account for land-cover changes due to indirect anthropogenic or natural causes. The differences among the three estimates produced in this study can be largely explained by the spatial and temporal differences in estimates of cropland and pastureland areas among the three data sets. Therefore these data sets offer a wide range of plausible regional estimates of uncertainty and the extent to which different ecosystems have undergone changes historically.

The biggest source of uncertainty in the global carbon budget remains emissions due to LULCC. Several multi-model comparison experiments have been performed to determine the uncertainty of LULCC in the global carbon budget. The LUCC uncertainty experiments involve using a common land-use data set (e.g. HYDE or RF) in each of the models and comparing the land-use fluxes. However, due to differences in the structure of each model, the method adopted to implement the common land-use data differs significantly between each model. As a result, it is impossible to attribute the estimated uncertainty to model-related uncertainty and uncertainties arising due to differences in the method of implementing land-use data between different models. However, driving the same model with multiple LULCC data sets derived consistently using same method (as presented in Chapter 2), opens a new avenue for studying LULCC data-related uncertainty by eliminating the model-related uncertainty, which is a focus of Chapter 3.

In Chapter 3, I forced a land-surface model, the ISAM with the three LULCC reconstructions (ISAM-HYDE, ISAM_RF, and ISAM-HH) to estimate the sensitivity of

different LULCC datasets to simulated CO_2 emissions from LULCC. The estimated cumulative LULUC emissions over the period 1900 - 2010 based on ISAM-HYDE data are ~180 GtC, which are ~33% of total carbon emissions (345 GtC from burning fossil fuels for the same period). The contribution of LULCC to global anthropogenic carbon emissions (land-use plus fossil fuel) in 1990's and 2000's were ~18 - 22% and 14 - 17% respectively (using fossil fuel emissions presented in the Global Carbon Budget) for our modeled results across three underlying data sets and including the nitrogen cycle.

The estimated net global emissions from LULCC (mean and range) across three data sets are 1.88 (1.7 to 2.21) GtC/yr for the 1980's, 1.66 (1.48 to 1.83) GtC/yr for the 1990's, and 1.44 (1.22 to 1.65) for the 2000's. Our estimates are higher than other published estimates that range from 0.80 to 1.5 GtC/yr for the 1990's and 1.1 GtC/yr for the 2000's. These results are higher than other published estimates because they include the effects of nitrogen limitation on regrowth of forests following wood harvest and agricultural abandonment. This effect is particularly noticeable in the cooler non-tropics where nitrogen removal through harvest or burning is not compensated by nitrogen deposition or nitrogen mineralization. The estimated LULUC emissions for the tropics are 0.79±0.25 for the 1980's, 0.78±0.29 for the 1990's and 0.71±0.33 GtC/yr for the 2000's, and for the non-tropics regions are 1.08±0.52, 0.90±0.19 and 0.69±0.12 GtC/yr for the three decades. Our model results indicate that failing to account for the nitrogen cycle underestimates LULCC emissions by about 40% globally (0.66 GtC/yr), 10% in the tropics (0.07 GtC/yr) and 70% in the non-tropics (0.59 GtC/yr). If LULCC emissions are higher than assessed, it means fossil fuel emissions would have to be even lower to meet the same mitigation target.

In Chapter 4, I extended the work presented in Chapters 2 and 3, to investigate the uncertainties in estimating future CO_2 emissions from LULCC driven by from various factors (e.g. LULCC estimates, environmental changes, and model structural and parameter uncertainty). The study involved sensitivity experiments by forcing a single land-surface model, ISAM with different forcing datasets on LULCC and environmental factors; with and without including nitrogen cycle; and, by varying key model parameters within their uncertainty space. Here, I highlight two key conclusions. First, nitrogen limitation of CO_2 uptake is substantial and sensitive to nitrogen inputs. In our model, excluding nitrogen limitation underestimated global total LULUC emissions by 34-52 PgC (~21-29%) during the 20^{th} century and by 128-187 PgC

(90-150%) during the 21st century (Table 7.8). The difference increases with time because nitrogen limitation will progressively down-regulate the magnitude of CO_2 fertilization effect on regrowing forests, due to decreasing supply of plant-usable mineral nitrogen. Second, historically, the indirect effects of anthropogenic activity through environmental changes in land experiencing LULCC (indirect emissions) are small compared to direct effects of anthropogenic LULCC activity (direct emissions). As a result, including or excluding indirect emissions had a minor influence on the estimated total LULUC emissions historically. In contrast, the indirect LULCC emissions for the 21st century are a much larger source to the atmosphere, in simulations with nitrogen limitation. This is because of the gradual weakening of the photosynthetic response to elevated (CO_2) caused by nitrogen limitation. Therefore, what fluxes are including in LULCC emissions across different models is a crucial source of uncertainty in future LULCC emissions estimates.

Using one land-surface model is potentially a limiting factor of the study because it does not represent a broad range of model physics response, especially given that there are significant uncertainties in modeling nitrogen and carbon cycles, LULCC activities considered, and even the method of implementing a given LULCC dataset across biosphere models. Conversely, using a single land-surface model is more appropriate for our analysis because we can consistently isolate the effects on LULCC emissions due to different LULCC activities, LULCC flux definitions, historical LULCC forcing, and future climate forcing. The above effects cannot be consistently isolated using multi-model comparisons because model-based differences (e.g. different land cover representations) make attribution difficult. This study underscores the crucial need for terrestrial biosphere models to consider nitrogen limitation in estimates of the strength of the future land carbon sink, especially on regrowing forests.

In chapter 5, I present a statistical model for land use allocation with an econometric interpretation of land suitability that is based on profit maximization (or cost minimization). The approach integrates economic theory, observed land use, and data on both socioeconomic and biophysical determinants of land use change. It is global in scope and is estimated using long-term historical data, thereby making it suitable for long-term projections, such as in Integrated Assessment Models. The method accounts for spatial heterogeneity in the nature of driving factors across geographic regions. The allocation is modified by autonomous development

(previous and neighboring land use patterns, thereby accounting for temporal and spatial autocorrelation), competition between land use types, and exogenous drivers that are treated as explanatory variables. The spatial and temporal resolution at which the model operates is flexible.

An added value of the statistical model is that the model parameters can be estimated against any given historical LULCC reconstruction, and the estimated parameters can be used to predict future LULCC patterns under any given scenario. This allows for studying the sensitivity of different historical LULCC dataset to simulating future spatial patterns of LULCC.

Typically, global scale land use allocation models are not evaluated in their ability to simulate long-term spatial patterns of historical land use change. They are developed and applied directly to predict future land use patterns (including GLM model used in IPCC CMIP5 and forthcoming CMIP6 that supply gridded LULCC datasets to participating Earth System Models). In contrast to previous approaches, we show that our model can reproduce the broad spatial features of the past 100 years of evolution of cropland and pastureland patterns. We also show that land use allocation approaches based solely on previous land use history (but disregarding the impact of driving factors e.g. GLM), or those based on mechanistically fitting models for the spatial processes of land use change do not reproduce well long-term historical land use patterns. With an example application to the terrestrial carbon cycle, we show that such inaccuracies in land use allocation can translate into significant implications for global environmental assessments. Therefore, our modeling approach and its evaluation provide an example that can be useful to the land use, Integrated Assessment, and the Earth system modeling communities.

In chapter 6, I took a more-detailed (spatially), but regional perspective to understand the dynamics and drivers of LULCC in India. I focused on India because over 80% of India's land surface is under human disturbance, making India as hotspot of LULCC. Moreover, the global LULCC datasets show high discrepancy in the spatial patterns of LULCC even for contemporary period. Therefore, there remain opportunities to significantly improve our understanding of LULCC in India. For this study, I used wall-to-wall analysis of high-resolution Landsat imageries covering the period 1985 to 2005 to quantify land-cover conversions in India. Further, I investigated the drivers of land-cover conversions that are most important at national scale be estimating spatial models between land-cover conversions and high-resolution spatial data on biophysical and socioeconomic factors. The results indicated massive conversions between

cropland and fallow land between 1985 and 2005 indicating the low resilience of cropland in India. The driver analysis at national scale indicates the high dependency of crop-fallow systems on monsoon and post-monsoon climate, labor migration, and access to market and irrigation facilities. I also find that India has experienced increased forest loss with time. Major drivers of forest loss at national scale between 1985 and 2005 were manufacturing of wooden furniture's/timber products, cattle/dairy/leather products (due to over-grazing), and manufacturing of wooden agricultural implements, mining/quarrying activities, and industrial development. Furthermore, colder and wetter regions and regions without electricity were also positively associated with forest loss, indicating over extraction of domestic fuel wood and building materials.

This study advances existing satellite based assessments of LULCC in India on three aspects. First, this is the first study to quantify land-cover conversions at national scale. Earlier high-resolution land cover mapping activities at national scale was mostly one-time effort-hence, unavailable at regular time intervals; their project-specific thematic classification, and varying data quality make compilation of consistent time-series imageries difficult. Second, we tracked conversions among 11 land classes that include fallow land. As land is scarce in India, understanding the dynamics of under-utilized land (conversions between cropland and non-productive uses, specifically fallow land and wasteland) is vital to land use planning. Identifying fallow land requires analysis of satellite imageries over multiple seasons and years. We are unaware of any high-resolution global or national remote sensing product that differentiates between cropland and fallow land. Third, this is also the first modeling study at national scale to investigate the causes of LULCC. To carry out the causal analysis, we compiled the most detailed (~630,000 villages) national level spatial database on over 200 socioeconomic variables for two census years. Earlier national level studies for India were limited to land cover mapping, hindered by limited and coarse spatial data on socioeconomics at national scale.

In the following two sections, I propose two specific research questions that I would like to specifically address after my PhD. The proposed work connects with the modeling tools and datasets I developed during my PhD.

7.2 Future Work: Research Topic 1: Spatial land use in Integrated Assessment: Impacts of modeling uncertainties on simulating decisions and their consequences for terrestrial carbon fluxes

This sub-section is based on a proposal I developed for the NCAR ASP Postdoctoral Fellowship. The proposed work builds on the spatial land use change model presented in Chapter 5.

Summary: Land-use change is both a source and consequence of climate change (Rounsevell et al., 2014). For example, land-use change affects climate through alterations of carbon, energy, and moisture fluxes to the atmosphere (Lawrence et al., 2012; Bonan, 2008). Conversely, agro-climatic studies suggest that cropping areas have gradually expanded to higher latitudes and altitudes over the last two decades due to improved thermal conditions, an adaptation response to climate change (Zhang et al., 2013; Kumar et al., 2012). Additionally, climate variability (causing droughts, heat waves, etc.) can lead to cropland losses, often resulting in famines in food insecure regions (Hansen et al., 2011; Rosenweig et al., 2002). Modeling the feedbacks between land-use change and climate change and variability is therefore crucial for exploring the influence of alternative socio-economic development paths on future greenhouse gas emissions, and climate change mitigation, impacts, and adaptation in integrated assessments.

Given the central role of land use modeling in simulating decisions, current land use allocation approaches can lead to widely divergent outcomes even for a given scenario, highlighting our current uncertainty in modeling land change (e.g. Von Lampe et al., 2014; Schmitz et al., 2014). Moreover, the consequences of these uncertainties for both climate, and terrestrial processes are not well understood (Rounsevell et al., 2014; Hibbard et al., 2010). To underscore, both the CMIP5 and the forthcoming CMIP6 do not consider the impacts of uncertainties in historical land use (and their drivers) on future land use patterns (Hurtt et al., 2011). The proposed work aims to address this important gap by systematically testing the implications of historical data limitations on projected land use patterns, and their consequences for simulated terrestrial processes and economic decisions about land use. The proposed work is

of broad relevance to the land use, Integrated Assessment, and the Earth system modeling communities.

Background: A key priority of NCAR's Integrated Assessment Modeling (IAM) group is to address the above question by developing and applying integrated (primarily through land use) socio-economic (iPETS model) and biophysical models of the climate system (CLM/CESM). Additionally, understanding the biophysical interactions between land-use change and climate is a key priority for the TSS/CLM group.

The iPETS model disaggregates the world into 9 regions and land use decisions are made at this regional level. However, the CLM (the land-surface component within CESM) typically requires land use information at 0.5ºx0.5º resolution. To provide this information, a spatial land use allocation approach that I developed with the IAM group (Meiyappan et al., 2014) is used to downscale the aggregate regional land demands (for cropland and pastureland) to individual grid cells within that region. The land allocation model is based on hypothetical grid cell agents that maximize profits based on concomitant changes in driving factors (e.g. climate change and climate variability, socioeconomic factors including population and market locations). The model estimates statistical relationships between spatial land-use patterns and its driving factors from historical data. The estimated relationships are then used for spatially allocating future land use in iPETS-CLM scenarios (Figure 7.1; note: land-use feedbacks on climate are currently not included).

Proposed research and its significance: My previous work on the development of spatial land allocation model indicated that the choice of historical land use and climate (driver) data could significantly impact the estimated model relationships, potentially affecting the spatial land use projections. For my ASP fellowship, I propose to investigate the implications of uncertainty in historical land use/climate and its consequences for integrated assessments, including both economic decisions about land use and their consequences for terrestrial carbon fluxes. Specifically, the following three questions I propose to investigate will improve the impact and mitigation assessments in IAMs by better accounting for land-related uncertainty, and further help prioritize future research by identifying factors that are most important to determining outcomes in IAMs.

1. **Implications of uncertainties in historical land use reconstruction for modeling spatial land use in IAMs**: Historical land use data are important for spatial land use modeling in IAMs for two reasons. First, the parameters of spatial land use models are estimated based on historical land use data and their driving factors. Second, such reconstructions are used by IAMs as base maps representing current land use as a basis for future projections. Both model parameters and base year distributions in turn affect projected future land use.

However, existing land use reconstructions are not pure observations, but are modeled estimates that draw on national/sub-national statistics to the extent possible. These reconstructions have significant uncertainties due to uncertainties in both modeling (Goldewijk and Verburg, 2014), and inventory datasets (Meiyappan and Jain, 2012). For example, comparison of two well-known land use reconstructions for year 2005 show that the global pastureland area estimated by the HYDE reconstruction (used in CMIP5, and forthcoming CMIP6; Hurtt et al., 2011) is 26% higher than the 26.3 million km^2 estimated by RF data (updated version of Ramankutty and Foley, 1999 data). The relative difference between the two reconstructions is compounded both regionally (Figure 7.2), and as we go farther back in time from 2005.

I propose to examine the impacts of using alternative land use reconstructions for modeling spatial land use in IAMs. For this task, I will estimate two land use allocation models (Esti_RF and Esti_HYDE) using the two land use reconstructions (RF and HYDE), but keeping the set of historical driving factors the same across both estimation. Both the land use allocation models will then be applied for a given iPETS scenario to project spatial land use patterns (each using two different base year maps of current land use from HYDE and RF resulting in four different spatial projections – Esti_RF-Baseyr_RF, Esti_RF-Baseyr_HYDE, Esti_HYDE-Baseyr_RF, Esti_HYDE-Baseyr_HYDE). The difference in the future land use projections between Esti_RF-Baseyr_RF and Esti_HYDE-Baseyr_RF, and between Esti_RF-Baseyr_HYDE and Esti_HYDE-Baseyr_HYDE provides a quantitative understanding of the impacts of uncertainties due to historical model estimation for predicting future land use patterns. Similarly, the difference in the future land use projections between Esti_RF-Baseyr_RF and Esti_RF-Baseyr_HYDE, and between Esti_HYDE-Baseyr_RF and Esti_HYDE-Baseyr_HYDE will isolate the effects of using different base year land use maps on projected land use patterns. The

same set of simulations will be extended to multiple iPETS scenarios to understand the implications of historical land-use uncertainties for predicting spatial land use under alternative land-use trajectories (e.g. cropland abandonment vs. expansion scenarios).

2. **Implications of uncertainties in historical climate for modeling spatial land use in IAMs:** Similar to land use, historical gridded climate datasets are not "pure" observations. They are obtained from "reanalysis" that provides best approximation of the atmosphere using both observations and dynamic models (Trenberth et al., 2008). These reanalyzed products often have unintended trends/biases due to (but not limited to) changing mix of observations (e.g. introduction of new satellites), errors in bias correction methods, imperfect models, and data sparsity (Figure 7.3). For example, a single radiosonde observation can influence the regional precipitation for several 100km in reanalysis due to data sparsity (Bosilovich et al. 2011). This is an important source of uncertainty because precipitation can vary rapidly, especially in complex topographies. To give another example, two state-of-the-art reanalysis, NASA-MERRA and ERA-Interim systematically overestimate small and medium precipitation amounts, and underestimate high amounts, due to errors in convective parameterization scheme in the host forecast model among others (Pfeifroth et al., 2012). Therefore, the choice of reanalysis used for estimating the spatial land model can also affect estimated model parameters, and hence influence the future land use patterns. Here, I propose to run a similar set of experiments as for question 1 (but by keeping historical land-use data same across estimation, but varying climate driver datasets) to understand the sensitivity of climate history to predicted land use patterns.

Further, for projection studies, meteorology calculated prognostically by the climate model (here, CCSM; Figure 7.1) embedded with the CLM is used to drive the land allocation model (Figure 7.1). Therefore, for consistency in climate driver data used between historical model estimation and future projections, one could use the CCSM simulated climate history for model estimation, instead of reanalysis data. However, climate models, in general do not reproduce the climate history accurately, and they are far more divergent than reanalysis products (Figure 7.3). Here, I will also test the implications of using climate history simulated by CCSM (different ensembles) vs. reanalysis products for modeling future spatial land use.

3. Effects of spatial land use on carbon emissions and simulated economic decisions: There are significant gaps in our understanding of how sensitive environmental outcomes are to land use, especially at the regional scale (e.g. Brovkin et al., 2013; Houghton, 2013). For instance, the four different spatial land use projections for a given scenario (based on alternative historical land use as described in question 1) will imply four different future land use emissions pathways from CLM. The first aspect of this task is to understand how different will be the simulated future emission pathways, given the uncertainties in historical land use data.

The second aspect of this task is to understand the effects of spatial land use (and emissions) on mitigation assuming that the same forcing target is maintained for the scenario. Mitigation cost is an example of simulated economic decision affected by land use. For example, higher (lower) land-use emissions imply that more (less) CO_2 emissions from energy would have to be mitigated if the same forcing target for the scenario is to be maintained. Therefore, uncertainty in spatial land use projections (hence, emissions) would translate as uncertainty in mitigation cost assessment in IAMs. The goal of this task is to understand: 1. the effects of land use emissions on mitigation costs, and 2. the effects of uncertainties in spatial land use (therefore, emissions) in simulating land use decisions in IAMs (again using mitigation cost as an example). For this task, given a scenario (e.g. SSP5), I would assume the iPETS scenario would have to meet the SSP5 forcing pathway under any circumstance. Then based on the four different land use emissions pathway for the scenario (as described in the above paragraph), I would use iPETS to estimate the mitigation cost of CO_2 emissions from energy (and their uncertainty range) required to maintain the same forcing pathway. This task would involve me developing codes to link land-use emissions from CLM to the iPETS, which currently is absent (see Figure 7.1; currently only yields are exchanged). I am especially suited for this task, given my familiarity with both iPETS and CLM codes. Further, on the long-term, this linkage will be useful beyond the scope of this task for the NCAR IAM-CLM modeling framework to investigate a variety of land use related questions relevant to policy responses.

7.3 Future Work: Research Topic 2: Estimating India's land change from 1950 to 2011 at 1km resolution

This section is a work that I am currently carrying out, and is an extension to the work presented in Chapter 6. Here, I briefly present the motivation and objectives of the work.

Motivation: In Figure 7.4, I compare the cropland patterns for year 2000 for India between two widely used global land use datasets (HYDE and RF; also presented in Chapter 2). The data show striking differences in cropland patterns even for contemporary period, highlighting the uncertainty over study region. The results are striking because the comparison is for satellite-constrained era; the uncertainties get compounded as we move backward in time (not shown), because the results become increasingly model dependent. Therefore, reducing uncertainties in the historical LULCC patterns over India is important for environmental impact assessment as well as land use planning.

Objectives: The aim of this task is to reconstruct historic LULCC for India annually for the period 1950 to 2011 with more thematic detail (land use/cover classes) and higher spatial resolution (1km lat/long) compared to previous studies. The study will make better use of detailed and reliable historic land use/cover data sources for India compared to previous studies.

The proposed work improves previous LULCC reconstruction efforts on multiple fronts (see Table 7.1 for a summary): (1) I will integrate inventory datasets at district level over the historic period, as opposed to state level in previous studies (roughly a 20-fold improvement in spatial resolution). The increased detail of historic inventory statistics will significantly improve the accuracy of modeled LULCC reconstruction. (2) I will integrate some of the best high-resolution land cover imageries acquired from satellites starting from 1972. High-resolution land cover imageries are crucial to capture the highly heterogeneous (Figure 7.5) and fragmented landscapes (Figure 7.6) of India. (3) The historical land cover patterns prior to satellite era are constrained through hand-digitized topographic maps. This improves spatial constraint on the modeled LULCC patterns historically, thus increasing the accuracy. (4) Both the 1950 and current LULCC patterns will be validated using ground observations using stratified random sampling. The validation exercise will provide better assessment of uncertainty in the modeled product, a key component existing LULCC reconstructions (both global and over India) typically lack.

7.4 Tables

Table 7.1 Summary of input datasets to be used in the study.

Data	Spatial Resolution	Spatial Coverage	Temporal Coverage/Resolution	Remarks
Inventory (Tabular)				
Land use statistics	District level	National	1950-2011 (Yearly)	DES 1998-2011: Digital http://lus.dacnet.nic.in/ 1950-1997: This study (digitized from hard-copy books).
Land use statistics	Village level (~630,000 units)	National	Two census years (1991, 2001)	Tabular data available online (http://censusindia.gov.in/) This study ties the tabular data to village-level administrative boundaries.
Urban area	District level	National	Decadal corresponding to census years (1951, 61, 71, 91, 2001, 11). We could not acquire 1981 data.	Census of India Digital data: 2001 & 11 (http://censusindia.gov.in/) This study digitized data from hard-copy books for 1991 and prior.
Remote Sensing				
Landsat MSS/TM derived land cover	30m (1:50k scale)	National	1985, 1995, 2005	Roy et al. (2015)
Resourcesat-1/AWiFS derived land cover	56m (1:250k scale)	National	2005-2011 (Yearly)	NRSA (2005)
Landsat MSS derived land cover	60m (1:1M scale)	Covering ~20% of India	1972, 1982	This study
Topographic Maps				
Survey of India land cover	Sampled to 30m in GIS (1:50k scale)	Ganga basin covering ~30% of India	1975	Survey of India

Table 7.1 (Cont.)

National Forest Map by Survey of India	Sampled to 1km in GIS (1:4M scale)	National	Circa 1942	Scanned map: http://nla.gov.au/nla.map-vn6196447 Digital version: This study
US Army Maps (Series U502)	Sampled to 1km in GIS (1:250k scale)	National	Circa 1950	Scanned Maps: http://www.lib.utexas.edu/maps/ams/india/ Digital Version: This study Used as independent data for validation of our reconstructed maps.
Ground data				
Field data on land cover selected through stratified random sampling	~16000 field points	National	Circa 2005	Biodiversity Mapping Project (Roy et al., 2015) For validation of current maps.
Auxiliary data				
District-level administrative boundaries of India	Sampled to 1km in GIS (1:250k scale)	National	Decadal corresponding to census years (1951, 61, 71, 81, 91, 11)	Paper Maps: Administrative Atlas of India Digital Version: This study
District-level administrative boundaries of India	Sampled to 1km in GIS (1:250k scale)	National	2001	Digital version: Survey of India
Agro-Ecological Zones	Sampled to 1km in GIS (1:250k scale)	National	Static map representative of contemporary conditions (1950-present)	Gajbhiye and Mandal (2000)

Table 7.2 Key improvements in this study compared to earlier studies. The comparison applies to India for the post-1950 period.

Aspect	This study	Tian et al. (2014)	Global studies[1, 2, 3]
Input data			
Land use statistics	Annual district level statistics for 1950-2011. Village level statistics for 1991 and 2001.	District level statistics at decadal intervals and annual statistics from 1998. Has a tendency to mask out within-decade variations prior 2000.	Annual national level statistics only[1], or combined with state level statistics collected at roughly decadal intervals[2].
Historical built-up area	Constrained by district level inventory data at decadal intervals.	Estimated using population as proxy.	Estimated using population as proxy[3].
Incorporation of remote sensing products	Incorporates high-to-medium resolution remote sensing imageries from 1972 onwards (see Table 7.1).	Incorporates Resourcesat-1/AWiFS derived land cover for 2005 and 2009.	Incorporates coarse-resolution (> 250m) global product(s) circa 1990s and/or 2000s.
Incorporation of historical land map(s)	Spatial patterns of forest prior to satellite era are constrained through historical forest map.	None.	None.
Methods			
Apportioning inventory data to account for changing boundaries and divisions of districts	Decadal apportioning of district data to minimize loss of spatial information.	All data after 1950 are apportioned to 1950 district boundaries resulting in ~ 50% loss of inventory information.	National boundaries remain unchanged. Like districts, states also undergo boundary changes and divisions, thus requires apportioning[2]. However, note the number of states is one order less than the number of districts.

Table 7.2 (Cont.)

Spatial allocation of land use/cover	(1) All land types are allocated simultaneously, consistent with reality (2) Uses available time-series satellite imagery and historic maps to account for changes in land-use patterns over time.	(1) Spatial allocation of each land type is sequential. (2) Assumes current land use patterns mimic historical patterns.	(1) Spatial allocation of each land type is sequential. (2) Assumes current land use patterns mimic historical patterns./driver assumption (not realistic).
Validation	Both historic and current maps extensively validated using toposheets/ground observations.	None.	None.
Output (Final Reconstruction Product)			
Spatial resolution	1 km x 1km	~10 km x 10 km	~10 km x 10 km or coarser
National boundaries	Consistent with official Survey of India maps[*].	Excludes disputed territories of India.	Not applicable (global product).
Thematic coverage	12 land-use/cover classes.	5 land-use/cover classes: cropland, forest, wasteland, built-up & all other lands combined.	Focus on land use: agriculture (crops and/or pastures)[1, 2], or built-up land[3].

1 Klein Goldewijk et al. (2011) (based on DES statistics for 1961-present at national scale as compiled by FAO)
2 Ramankutty and Foley (1999) (see their Appendix A11)
3 Klein Goldewijk et al. (2010)
[*] http://www.surveyofindia.gov.in/files/Criminal_Law_Amend_Act_1990_2.pdf

7.5 Figures

Figure 7.1 Current integrated modeling framework for iPETS-CLM scenarios (2005-2100) showing data exchanged between different model components.

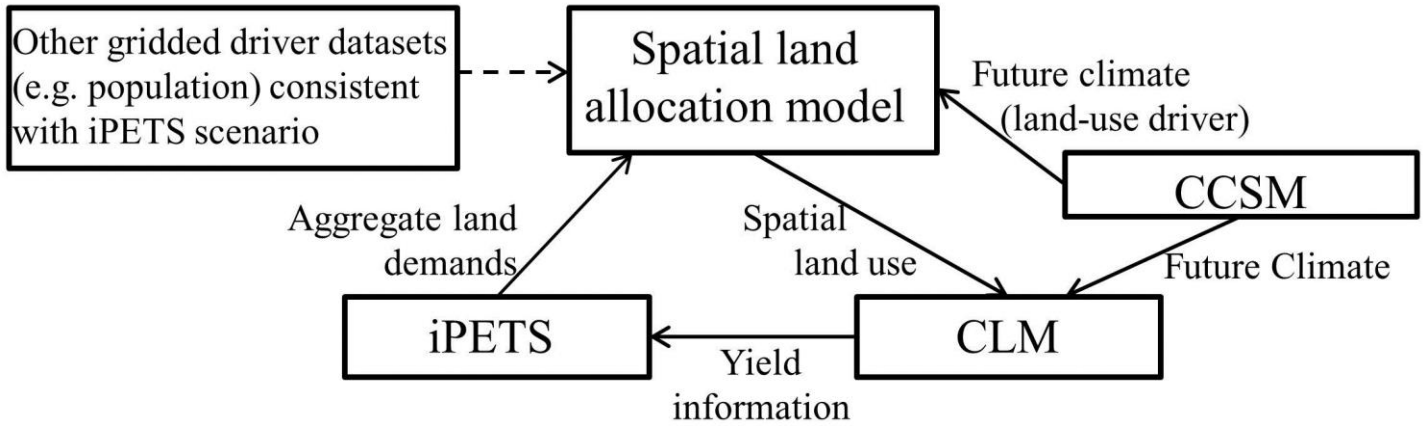


Figure 7.2 Comparison of pastureland area for 2005. Focus on regions highlighted in red where significant differences are found.

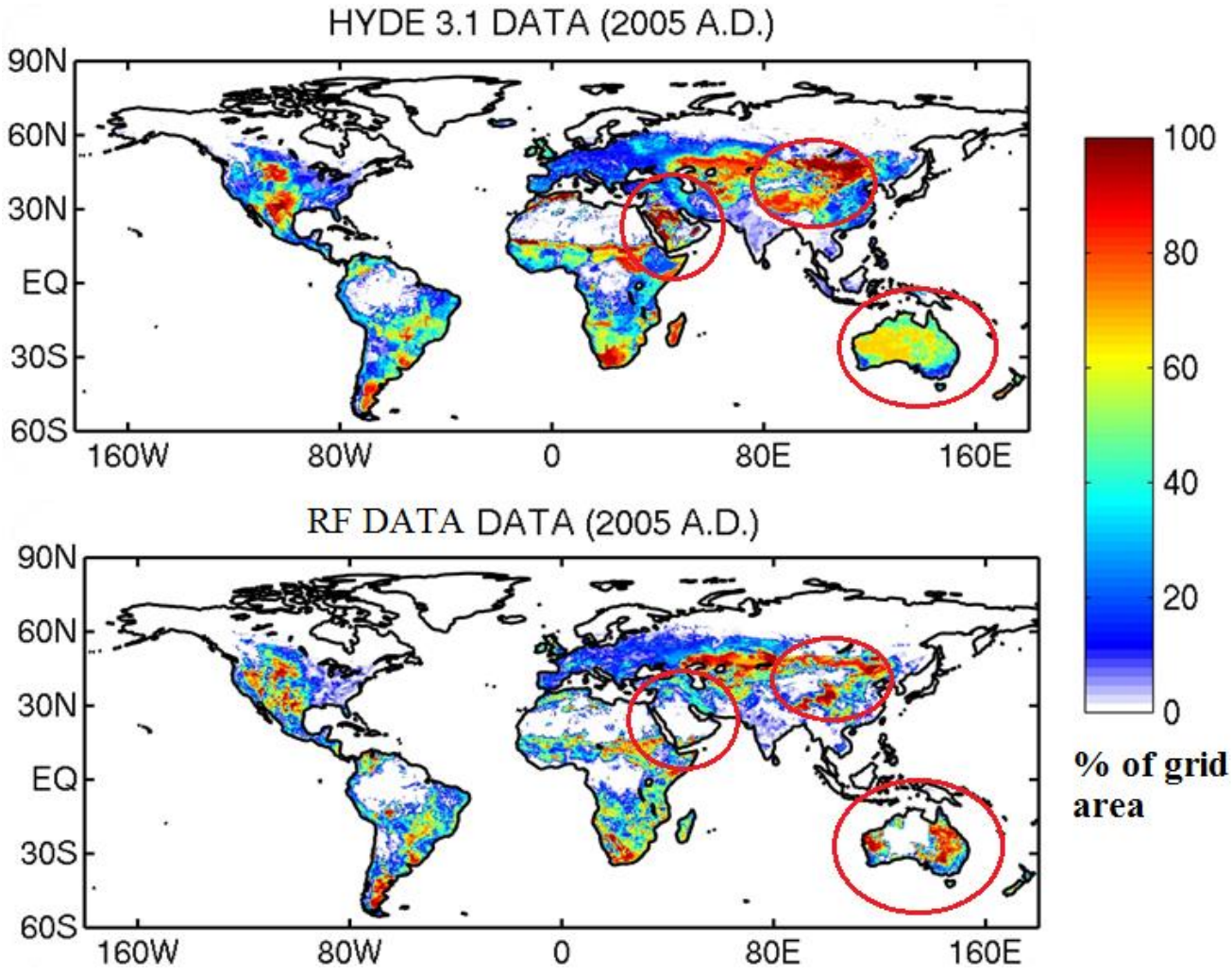


Figure 7.3 Comparison of grid-level (half-deg) maximum differences in mean annual meteorological variables between three 3rd generation reanalyses datasets (CFSR, MERRA, ERA-Intermin), amongst a larger number of reanalysis datasets (as in NCAR Climate Data Guide), and amongst ensemble of 28 CMIP5 models. Computations are based on a data from 2000-2004 (5 year average). (a) Average temperature; (b) Total precipitation; and (c) Incoming short-wave radiation.

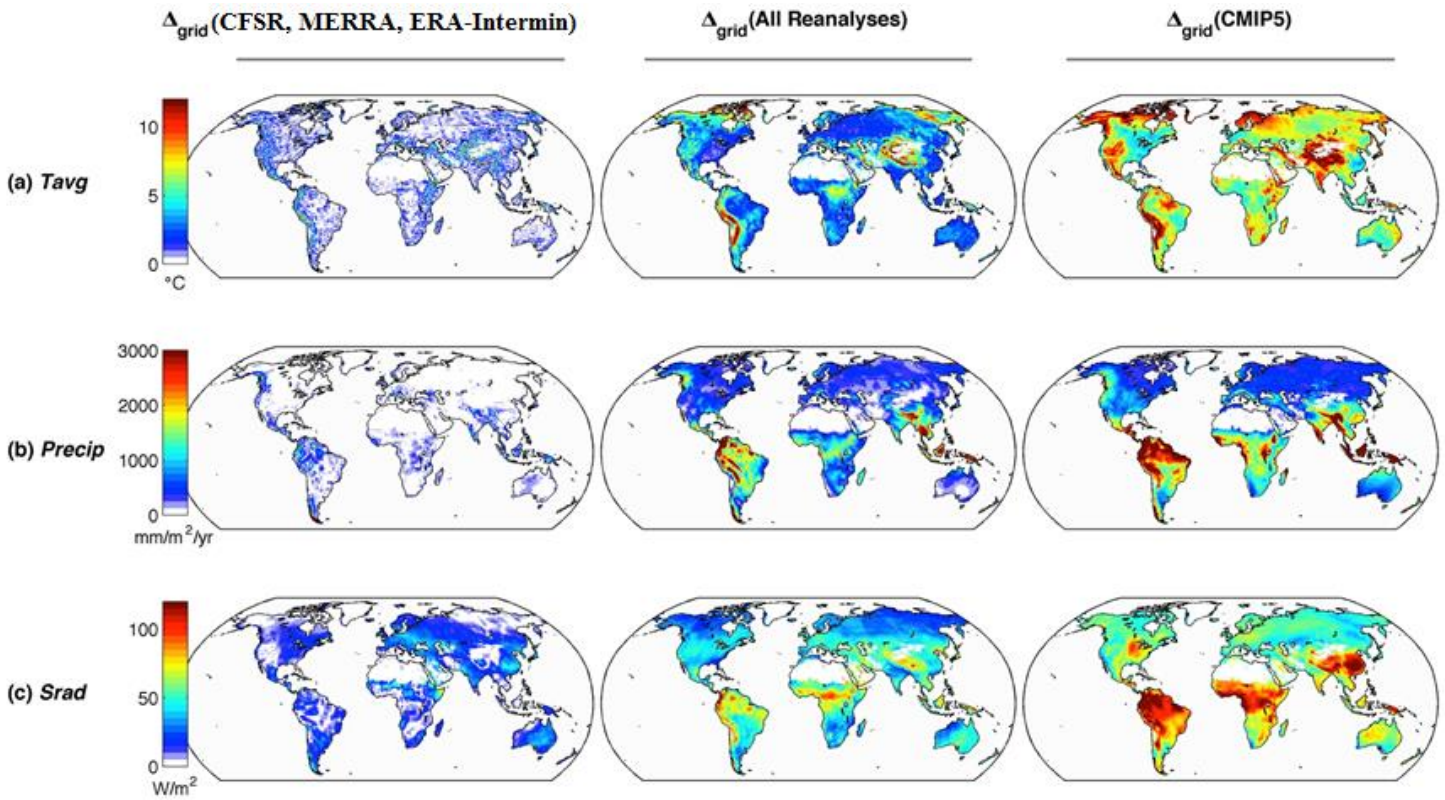


Figure 7.4 Difference in cropland estimated between two global "modeled" datasets zoomed over the Indian region: HYDE3.1 (Klein Goldewijk et al. 2011) minus Ramankutty et al. (2008). Comparison is for the year 2000 at ~10km x 10km spatial resolution.

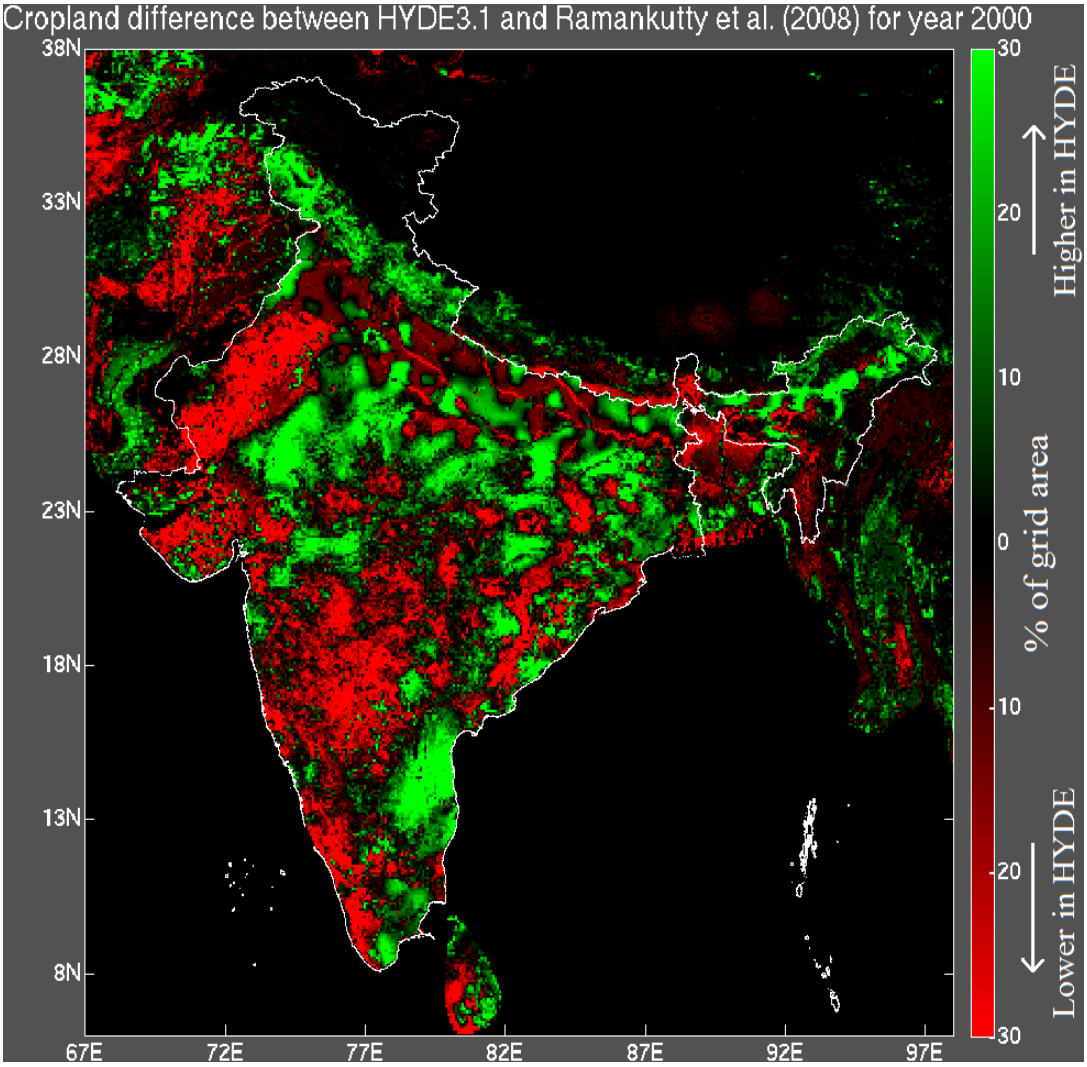


Figure 7.5 Heterogeneity of India's landscapes estimated using Landsat data (year 2005). For estimation, I used a moving window approach with 1 km x 1km cells and calculated the number of unique land cover types, each occupying at least 10% of the cell area.

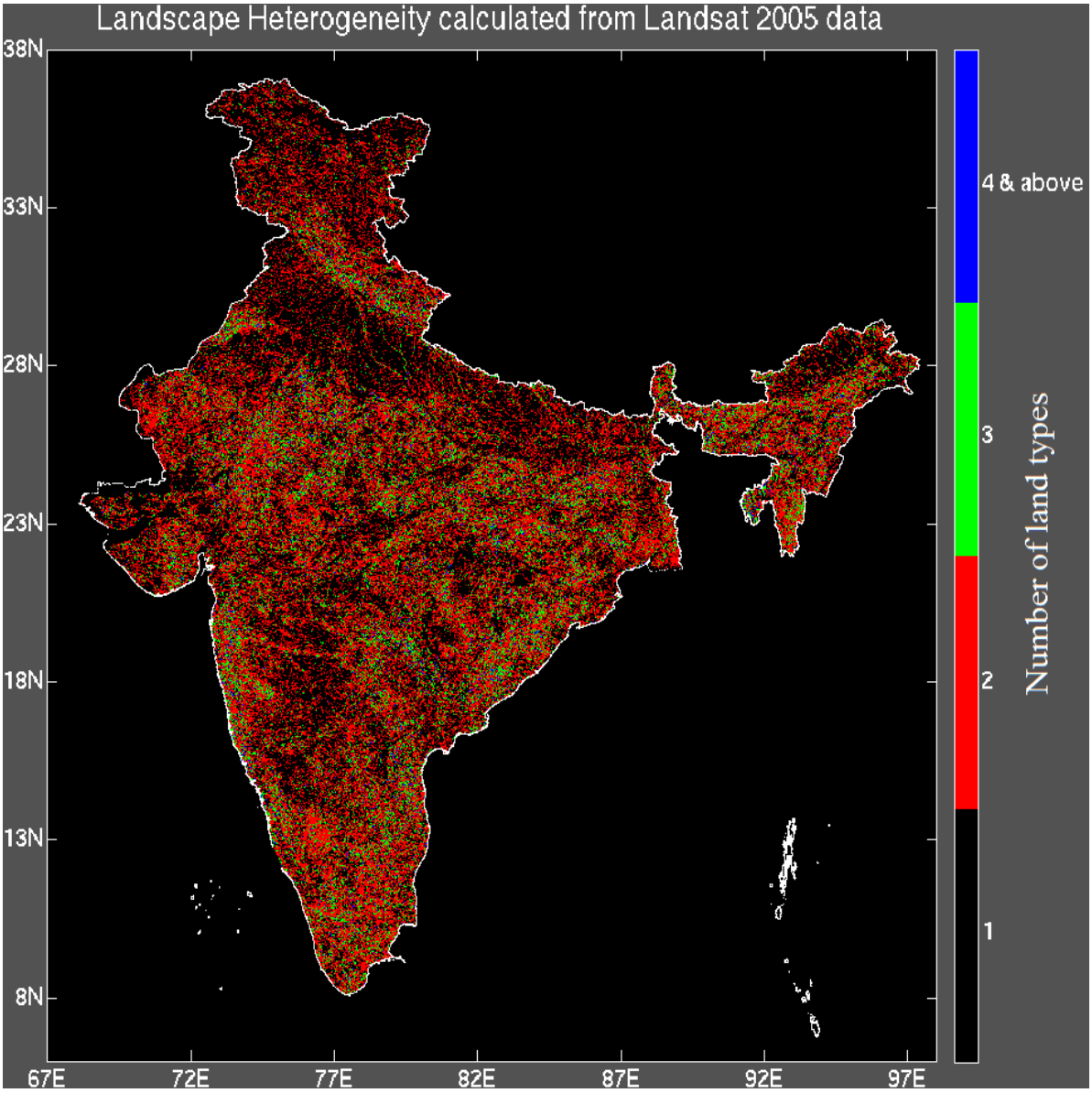
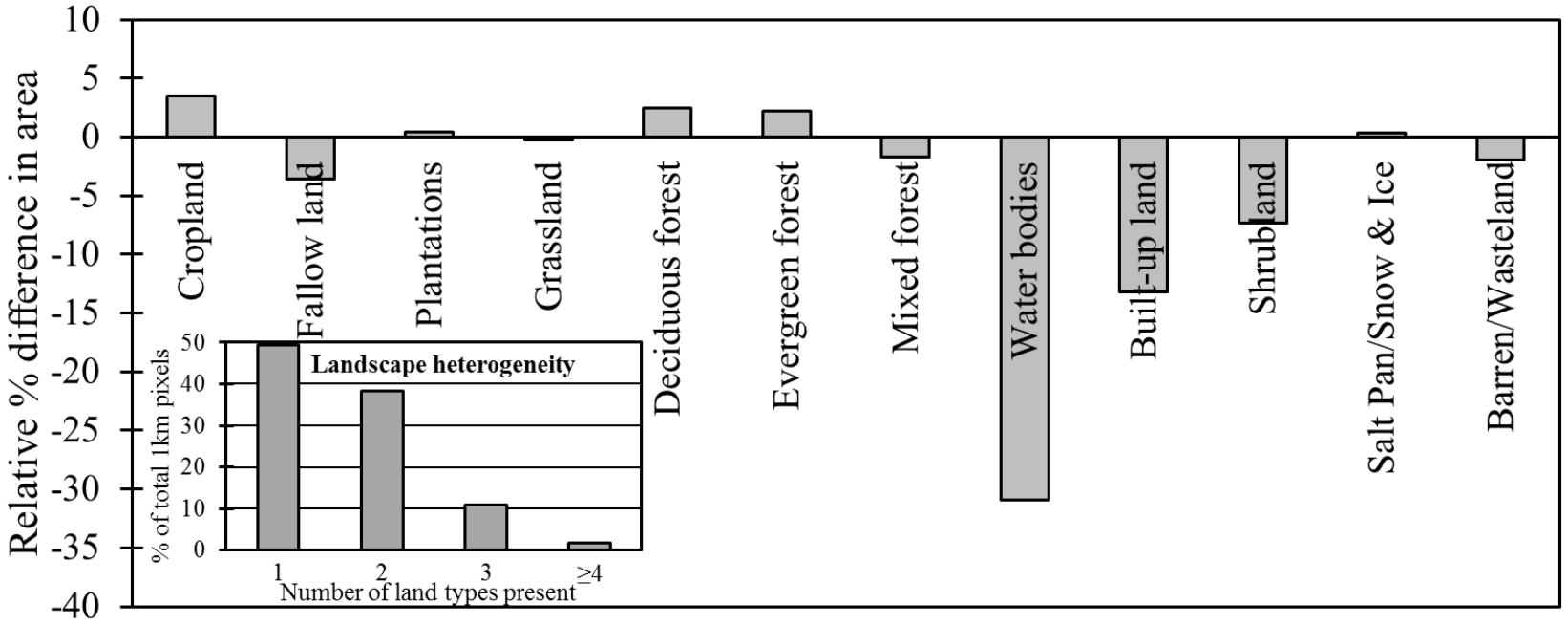


Figure 7.6 Landscape heterogeneity in India's land cover calculated from Landsat data (30m resolution) at national scale. The figure shows the effect of resampling the 30m land cover data to 1km. The approximation leads to significant bias of certain land cover classes (e.g. water bodies) even when aggregated because the fragmented landscapes are eliminated when the data is approximated to coarser resolution.



7.6 References

Bonan, G. B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. science, 320(5882), 1444-1449.

Bosilovich, M. G., Robertson, F. R., & Chen, J. (2011). Global energy and water budgets in MERRA. Journal of Climate, 24(22), 5721-5739.

Brovkin, V., Boysen, L., Arora, V. K., Boisier, J. P., Cadule, P., Chini, L., ... & Weiss, M. (2013). Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century. Journal of Climate, 26(18), 6859-6881.

Gajbhiye, K. S., & Mandal, C. (2000). Agro-ecological zones, their soil resource and cropping systems. Status of Farm Mechanization in India, Cropping Systems, Status of Farm Mechanization in India, 1-32.

Goldewijk, K. K., & Verburg, P. H. (2013). Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set.Landscape ecology, 28(5), 861-877.

Goldewijk, K. K., Beusen, A., & Janssen, P. (2010). Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. The Holocene.

Hansen, J. W., Mason, S. J., Sun, L., & Tall, A. (2011). Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture, 47(02), 205-240.

Hibbard, K., Janetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Betts, R., ... & Feddema, J. J. (2010). Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling.International Journal of Climatology, 30(13), 2118-2128.

Houghton, R. A. (2013). Keeping management effects separate from environmental effects in terrestrial carbon accounting. Global change biology,19(9), 2609-2612.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., ... & Wang, Y. P. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Climatic Change, 109(1-2), 117-161.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., & De Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global Ecology and Biogeography, 20(1), 73-86.

Kumar, S., Merwade, V., Rao, P. S. C., & Pijanowski, B. C. (2013). Characterizing long-term land use/cover change in the United States from 1850 to 2000 using a nonlinear bi-analytical model. Ambio, 42(3), 285-297.

Lampe, M., Willenbockel, D., Ahammad, H., Blanc, E., Cai, Y., Calvin, K., ... & Meijl, H. (2014). Why do global long-term scenarios for agriculture differ? An overview of the AgMIP Global Economic Model Intercomparison. Agricultural economics, 45(1), 3-20.

Lawrence, P. J., Feddema, J. J., Bonan, G. B., Meehl, G. A., O'Neill, B. C., Oleson, K. W., ... & Thornton, P. E. (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. Journal of Climate, 25(9), 3071-3095.

Meiyappan, P., & Jain, A. K. (2012). Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years.Frontiers of Earth Science, 6(2), 122-139.

Meiyappan, P., Dalton, M., O'Neill, B. C., & Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. Ecological Modelling,291, 152-174.

NRSA (2005). Manual National land use and land cover mapping using multi-temporal AWiFS data (LULC-AWiFS). Under Natural Resource Census Programme, Remote Sensing and GIS Application Area. National Remote Sensing Agency, Department of Space Government of India, Hyderabad.

Pfeifroth, U., Mueller, R., & Ahrens, B. (2013). Evaluation of satellite-based and reanalysis precipitation data in the tropical Pacific. Journal of Applied Meteorology and Climatology, 52(3), 634-644.

Ramankutty, N., & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global biogeochemical cycles, 13(4), 997-1027.

Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000.Global Biogeochemical Cycles, 22(1).

Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E., & Bloomfield, J. (2002). Increased crop damage in the US from excess precipitation under climate change. Global Environmental Change, 12(3), 197-202.

Rounsevell, M. D. A., Arneth, A., Alexander, P., Brown, D. G., de Noblet-Ducoudré, N., Ellis, E., ... & Young, O. (2014). Towards decision-based global land use models for improved understanding of the Earth system. Earth System Dynamics, 5(1), 117-137.

Roy, P. S., Behera, M. D., Murthy, M. S. R., Roy, A., Singh, S., Kushwaha, S. P. S., ... & Tripathi, S. K. (2015). New vegetation type map of India prepared using satellite remote sensing: Comparison with global vegetation maps and utilities. International Journal of Applied Earth Observation and Geoinformation,39, 142-159.

Schmitz, C., van Meijl, H., Kyle, P., Nelson, G. C., Fujimori, S., Gurgel, A., ... & Valin, H. (2014). Land-use change trajectories up to 2050: insights from a global agro-economic model comparison. Agricultural economics, 45(1), 69-84.

Tian, H., Banger, K., Bo, T., & Dadhwal, V. K. (2014). History of land use in India during 1880–2010: Large-scale land transformations reconstructed from satellite data and historical archives. Global and Planetary Change, 121, 78-88.

Trenberth, K. E., Koike, T., & Onogi, K. (2008). Progress and prospects for reanalysis for weather and climate. Eos, Transactions American Geophysical Union, 89(26), 234-235.

Zhang, G., Dong, J., Zhou, C., Xu, X., Wang, M., Ouyang, H., & Xiao, X. (2013). Increasing cropping intensity in response to climate warming in Tibetan Plateau, China. Field Crops Research, 142, 36-46.