ABSTRACT

Title of Document: ASSESSING CELLULOSIC BIOFUEL FEEDSTOCK PRODUCTION ACROSS A GRADIENT OF AGRICULTURAL MANAGEMENT SYSTEMS IN THE US MIDWEST

Ritvik Sahajpal, Doctor of Philosophy, 2014

Directed By: Professor George C. Hurtt
Department of Geographical Sciences

While biofuels are widely considered to be a part of the solution to high oil prices, a comprehensive assessment of the environmental sustainability of existing and future biofuel systems is needed to assess their utility in meeting U.S. energy and food needs without exacerbating environmental harm.

The following questions guide this research:

1. What is the spatial extent and composition of agricultural management systems that exist in the U.S. Midwest?
2. How does sub-grid scale edaphic variation impact our estimation of poplar biomass productivity across a gradient of spatial scales in the U.S. Midwest?
3. How do location and management interactions impact yield gap analysis of cellulosic ethanol production in U.S. Midwest?

In the first chapter, I developed an algorithm to identify representative crop rotations in the U.S. Midwest, using remotely sensed data; and used this information to detect pronounced shifts from grassland to monoculture cultivation in the U.S. Midwest. In the second chapter, a new algorithm is developed to reduce the computational burden of high resolution ecosystem modeling of poplar plantations in U.S. Midwest, with the results from the high resolution modeling being used to estimate the impact of averaging and discretization of soil properties on poplar yield estimates. In the third chapter, a novel yield gap analysis of cellulosic feedstocks on marginal lands in the U.S. Midwest is conducted to determine the management inputs needed to reach their yield potential in a sustainable manner.

The significance of this research lies in providing a spatially explicit regional scale analysis of the tradeoffs between food and fuel production and providing an understanding of which biofuel crops should be grown where to maximize production while mitigating environmental damage.

ASSESSING CELLULOSIC BIOFUEL FEEDSTOCK PRODUCTION ACROSS A GRADIENT OF AGRICULTURAL MANAGEMENT SYSTEMS IN THE US MIDWEST

By

Ritvik Sahajpal

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor George C. Hurtt, Chair
Prof. Anand Patwardhan
Prof. Klaus Hubacek
Prof. Roberto C. Izaurralde
Dr. Xuesong Zhang

Preface

Chapter 2 contains a co-authored paper currently under review. I am the lead author for the paper and am responsible for all aspects of the manuscript and completed the work independently.

R. Sahajpal, X. Zhang, R. C. Izaurralde, I. Gelfand, G. C. Hurtt

Identifying representative crop rotation patterns and grassland loss in the US western corn belt

Comput. Electron. Agr., In review.

Dedication

To all my mentors, family and friends.

Acknowledgements

I have been very fortunate during the course of this dissertation to have been associated with excellent people, both professionally as well as on a personal basis. My deepest thanks to my PhD advisor George Hurtt, who was instrumental in helping me across the finish line. I have learnt a lot about thinking about science from him and hope to continue doing so in the future as well. I have lots of gratitude for César Izaurralde who was not only my supervisor during my research assistant stint at JGCRI, but also played a key role in educating me about the intricacies of ecosystem modeling. Tons of thanks also to my committee members Xuesong Zhang, Klaus Hubacek and Anand Patwardhan for so gracefully sharing their invaluable time and support. My thanks to the entire GEL team as well, in particular, Justin Fisk, Maosheng Zhao, Christine Kang, Katelyn Dolan and Steve Flanagnan for their continuous support and encouragement.

I have been so lucky to have come across wonderful people who became friends during this journey: Jyothy, Praveen, Khaldoun, and Anupam. These are precious life-long associations which I will cherish forever. Thanks guys!

My family has been super supportive and patient throughout this endeavor and made me feel better about the process even when I had doubts. I love them immensely and cannot thank them enough. Heartfelt thanks also go to my amazing wife, Pallavi, without whose rock like support, the end would have been a lot tougher. Thank you everyone!

Table of Contents

List of Tables

List of Figures

Chapter 1: Context of Cellulosic Biofuel Feedstock Production in the US Midwest

1.1 Background

The annual world primary energy consumption is around 500 EJ (BP 2011), which is the energy equivalent of the 2011 earthquake in Japan, *daily*. Fossil fuel (oil, natural gas and coal) combustion provides nearly 87% of our consumption. This releases over 8 Gt C/yr of greenhouse gas (GHG) emissions (Marland et al., 2008), accelerates changes in composition of natural waters and air (Bertine and Goldberg, 1971) and has chronic effects on human health (Cifuentes et al., 2001). While world energy consumption is projected to increase by 53 percent from 2008 to 2035 (Conti et al., 2011), there is an urgent need to curb emissions (Kharecha and Hansen, 2008). The future energy portfolio needs to be reassessed.

As a low-carbon energy source, biofuels are widely considered to be a part of the solution to high oil prices (Tyner, 2008), climate change (Tilman et al., 2009) and carbon dioxide (CO_2) emissions from fossil fuels (Hill et al., 2009) . Although accounting for less than 3 percent of the global transportation fuel supply, biofuel production worldwide has increased six-fold, from 4.8 billion gallons in 2000 to nearly 28 billion gallons in 2010 (Coyle, 2007, OECD-FAO, 2011), and is expected to provide nearly 12 percent of the global transportation fuel consumption by 2050 (Demirbas, 2008). Almost three-quarters of the current global biofuel production or nearly 21 billion gallons comes from first-generation biofuels (made from sugar, starch and vegetable oil) and is concentrated in the US, Europe and Brazil (Coyle, 2007). US ethanol production derives almost exclusively from fermenting corn (*Zea mays*) and has increased eight-fold in the last decade, from 1.6 billion gallons to 13.5 billion gallons of ethanol per year (Conti et al., 2011). This dramatic increase has helped US surpass Brazil, which uses sugarcane as its biofuel feedstock, in ethanol production.

The US biofuel production targets are even more ambitious: 36 billion gallons per year by 2022, of which 20 billion gallons will come from second generation biofuels consisting of non-grain or cellulosic plant material. Similar mandates exist in Europe, where biofuel production (mostly from rapeseed oil) increased fifteen-fold between 1998 and 2010, to 2.5 billion gallons (European Biodiesel Board, 2011) and would further need to increase four-fold to attain a 5.75% market share target for biofuels (Commission of the European Communities, 2006).

A confluence of factors is expected to increase the human and capital resources dedicated to biofuel feedstock production in the future. Favorable market conditions have led to an increase in corn acreage nationally by almost 20 percent from 2006 to 2007 (Landis et al., 2008), increased farmer acceptance for using cellulosic feedstock (Swinton et al., 2010) and additional cropland has become available for cultivation because of the decrease in the maximum acreage enrolled in the Conservation Reserve Program (Westcott, 2010). In the recent past, U.S. government subsidies on corn-based ethanol production have also encouraged farmers to shift from cultivating food crops to corn (Scharlemann and Laurance, 2008).

Agriculture plays a key role in the human domination of the global ecosystem (Vitousek et al., 1997). They are also responsible for several ecosystem services ranging from food (production) to nutrient regulation to habitat functions. By diverting existing land-uses towards more biofuel feedstock production, we could exacerbate environmental harm in several ways. Ironically, growing biofuel feedstock can actually increase GHG emissions, particularly if they replace existing carbon rich vegetation on virgin lands. This incurs a carbon debt which can take anywhere from a few years to several hundred years to repay depending on the land-use change and management inputs for biofuel feedstock production (Searchinger et al., 2008, Fargione et al., 2008).

The simplification of agroecosystems, through expansion of agricultural land supporting a single crop type is an important cause behind the decline in farmland biodiversity (Bianchi et al., 2006). As a result, ecosystem services associated with biodiversity, like nutrient recycling, microclimate regulation and natural pest control, have also deteriorated. Beyond their ecological importance, these ecosystem services provide other tangible benefits. For instance, the suppression of pest populations in crops by natural enemies reduces yield loss and the need for excessive use of pesticides (Gardiner et al., 2009, Landis et al., 2008, Meehan et al., 2011).

First generation biofuels can affect food security negatively. While no significant impact of biofuel production on feedstock prices has been observed currently (Ajanovic, 2011), perhaps because yields have increased concomitantly; estimates on projected price increases range between 2-65 percent for corn, 2-33 percent for wheat and 1-76 percent for soy (Naylor et al., 2007). The variance in the estimates comes from assumptions about the quantity of biofuel feedstock production. One factor in favor of biofuel feedstock production is their potential to promote rural development via wealth transfer to traditionally cash-poor farmers (Rajagopal et al., 2007), thereby potentially at least partially mitigating future food price increases.

To avoid competing with food, currently retired lands like those in the conservation reserve program (CRP) could be brought back into production. However, if used for growing first generation biofuels, this can not only accrue carbon debt of around 50 years but also add to health costs (Hill et al., 2009) and nutrient runoff and eutrophication (Donner and Kucharik, 2008). Thus, even though corn-ethanol reduces GHG emissions by displacing fossil fuel usage, conversion of CRP lands into corn-ethanol production has been discouraged (Piñeiro et al., 2009).

The search for beneficial biofuels should focus on the twin objectives of sustainable biofuel feedstocks that do not compete with food crops and do not induce either direct or indirect land-use change. The sustainable biofuel feedstocks include byproducts of human activities (crop residues, forestry wastes) and purpose-grown perennial mixes and woody bioenergy crops. Due care needs to be taken to avoid excessive harvesting of crop residues as it can intensify soil erosion by tenfold or more, increase GHG emissions and also increase eutrophication due to runoff (Pimentel et al., 2009). To minimize GHG emissions from land-use change, we need to identify lands that are initially not storing large quantities of carbon in soil and vegetation but are capable of producing abundant biomass with limited management inputs (Tilman et al., 2009).

When done right, biofuels can provide a solution to meeting the global environmental, food security and energy challenges (Robertson et al., 2008, V.H. Dale et al., 2010, Tilman et al., 2009, K. Kline et al., 2011). This proposal aims to address that by assessing biofuel feedstock production and ecosystem service tradeoffs across a gradient of agricultural management systems in the US Midwest.

1.2 Objectives

The significance of this research lies in providing a spatially explicit regional scale analysis of the tradeoffs between food and fuel production and providing an understanding of which biofuel crops should be grown where to maximize production while mitigating environmental damage.

Specifically, the project will focus on the following questions:

1. What is the spatial extent and composition of agricultural management systems that exist in the US Midwest?
2. How does sub-grid scale edaphic variation impact our estimation of poplar biomass productivity across a gradient of spatial scales in the US Midwest?
3. How do location and management interactions impact yield gap analysis of cellulosic ethanol production in US Midwest?

1.3 Theoretical framework

The theoretical framework of this research is the concept of sustainability. Biofuels research in particular has emphasized sustainability, leading to creation of national biofuel research centers like the Great Lakes Bioenergy Research Center (GLBRC) (Slater et al., 2010), with a dedicated team examining sustainability of various biofuel feedstock options. Sustainable biofuel feedstock option include native perennial mixes (Tilman et al., 2009), which more fully utilizing limited resources thus attaining greater productivity.

1.4 Scale

The research will focus on the US Midwest (fig. 1-1). This includes the states of North Dakota (ND), South Dakota (SD), Nebraska (NE), Minnesota (MN), Iowa (IA), Kansas (KS), Missouri (MS), Wisconsin (WI), Illinois (IL), Indiana (IN), Michigan (MN) and Ohio (OH). I selected the US Midwest as the study area because it accounts for most of the non-specialty agricultural production in the US and is already the major producer of corn ethanol.

1.5 Scope

The proposed research focuses on ecosystem service impacts of biofuel feedstock production. As such, it does not include economic externalities like impact of market prices and farmer acceptance of growing second generation biofuels.

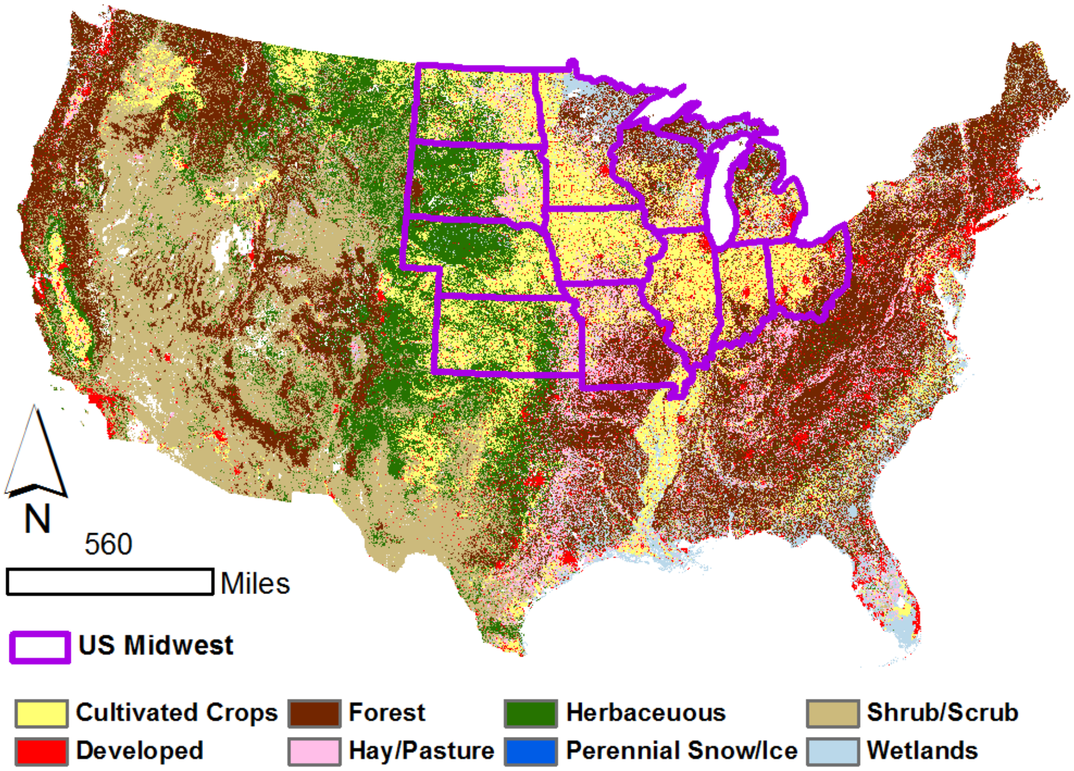

Fig. 1-1. Land-cover across coterminous U.S.

Chapter 2: Identifying Representative Crop Rotation Patterns and Grassland Loss in the US Western Corn Belt

1. *Introduction*

1.1 Background

As the dominant land-use type on Earth, agro-ecosystems cover more than a third of ice-free land surface (Ramankutty et al., 2008). They have a profound impact on the environment which is manifested through global fluxes of greenhouse gases (McCarl and Schneider, 2001), soil carbon dynamics (Lal, 2004), increased surface temperature and drought conditions (Hertel et al., 2010), and provision of ecosystem services (Foley et al., 2011). Human management of these agro-ecosystems, based on economic realities and ecological conditions, can influence both the magnitude and the nature of impact on ecosystem services (Robertson at al., 2000).

A key management activity performed by farmers is the development of crop rotation plans based on economic opportunities and adapted to environmental conditions. Crop rotations have been practiced for thousands of years but crop rotations practiced today are much simpler than those practiced in the past (Bullock, 1992 and Plourde et al., 2013). Compared to a monoculture cropping system supplied with optimum nutrient levels, the practice of crop rotations usually leads to higher yields, which are mainly attributed to improved soil fertility and tilth (Hesterman et al., 1987 and Pierce and Rice, 1988), as well as enhanced pest, disease and weed control (Liebman and Dyck, 1993 and Tilman et al., 2002). When practiced together with a low-intensity tillage regime, crop rotations can potentially reduce the global warming potential of agro-ecosystems (West and Post, 2002). Conversely, the simplification of agro-ecosystems, through expansion of agricultural land supporting a single crop type is an important cause behind the decline in farmland biodiversity (Bianchi et al., 2006). Consequently, ecosystem services associated with diversified crop rotations, like nutrient recycling, addition of organic matter and microclimate regulation have also deteriorated. Beyond their ecological importance, these ecosystem services provide other tangible benefits. For instance, the suppression of pest populations in crops by natural enemies can reduce yield loss and the need for increased use of pesticides (Landis et al., 2008, Gardiner et al., 2009 and Meehan et al., 2011), although there is uncertainty about the linkage between landscape simplification and pesticide use (Larsen, 2013).

With the advent of synthetic fertilizers and pesticides in the 1950s, use of crop rotations declined (Bullock, 1992). However, the increased intensification did not boost yields in comparison to a judicious crop rotation scheme (Mannering and Griffin, 1981). Amidst mounting concerns over the impacts of increased chemical inputs on surface and groundwater quality (Turner and Rabalais, 2003), there has been a renewed interest in crop rotations over the last couple of decades. A confluence of factors is expected to influence the crop rotation patterns and the acreage dedicated to crop production in the future as well. Favorable market conditions have led to an increase in corn (*Zea mays*) acreage nationally by almost 20% from 2006 to 2007 (Landis et al., 2008 and USDA, 2008) reaching a peak in 2010 (Plourde et al., 2013 and USDA, 2011a). A parallel development has been the availability of additional cropland for cultivation because of the decrease in the maximum acreage enrolled in the Conservation Reserve Program (Westcott, 2007). Farmer acceptance for using cellulosic feedstock has also grown (James at al., 2010), bringing new energy crops into the mix.

1.2 Mapping cropland areas and crop rotation patterns

The United States Department of Agriculture (USDA) tracks agricultural activity in the US at scale of individual counties as part of the census of agriculture (USDA, 2009). The analysis of these data provides valuable information on US farms, ranches and feedlots and the farmers who operate them. Due to the expensive and time consuming nature of this activity, new data are made available only once every five years. The coarse spatial resolution and sparse temporal availability precludes the derivation of fine resolution (< 1 ha) annual crop production and rotation maps. The other major source of agricultural activity monitoring is the Acreage (USDA, 2011b), which aggregates crop production information from surveying nearly 3 million ha of agricultural land. While produced annually, this information is provided at the scale of US states, and is therefore unsuitable for fine resolution mapping.

Remotely-sensed data provide a way to mitigate the temporal frequency and spatial resolution limitations of ground based surveys. National scale corn and soybean (*Glycine max*) acreage estimates obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) data at 250m and 500m resolution compare favorably with the USDA's National Agricultural Statistics Service (NASS) survey based acreage estimates (Change et al., 2007 and Wardlow and Egbert, 2008). While the MODIS satellite follows a rapid 16-day repeat cycle and the average farm size of 175 ha (Dimitri et al., 2005) far exceeds MODIS resolution, its utility in determining crop rotations is limited because crop rotations for individual fields in a farm can vary at sub-MODIS resolutions. This limitation is even more evident when using the 1-km resolution Advanced Very High Resolution Radiometer (AVHRR) data to classify crop cover types (Jakubauskas et al., 2002). A capable alternative is the USDA NASS Cropland Data Layer (CDL) which classifies more than 100 crop types in coterminous US at a resolution of 30/56m (Boryan et al., 2011). The oldest CDL product dates back to 1997 for North Dakota. While initially focused on the major crop producing states, the program has expanded to cover the conterminous US from 2008 onwards. The CDL has been used widely in land-cover change detection (Wright and Wimberly, 2013), watershed runoff modeling (Srinivasan et al., 2010), habitat monitoring (Meehan et al., 2010) and in process-based models for biofuel feedstock production analysis (Gelfand et al., 2013).

The simulation of biomass yields, evapotranspiration, runoff and related outcomes in a process-based model, are affected by soil properties, crop types and climatic conditions (Izaurralde et al., 2007). Therefore, it is critical to get both the spatial extent and temporal coverage of crop rotation patterns right for land-cover change analysis and ecosystem modeling. A variety of approaches have been developed to construct crop rotation patterns. These approaches fall into three categories: (a) Methods based on a mathematical framework: Crop rotations are modeled either as a transition matrix with implicit path dependency (Castellazzi et al., 2008) or are based on linear programming models (Detlefsen and Jensen, 2007).

While these methods attempt to incorporate expert knowledge on suitable crop rotation patterns, they usually do not include the market drivers that influence farmer's decisions on an annual basis. Instead, they are more useful for exploratory modeling studies (Dogliotti et al., 2003). (b) Crop rotation patterns determined through consultation with field experts, USDA extension agents or from NASS survey data (Arabi et al., 2008): Unless the study area is a small farm, it is difficult to obtain spatially explicit information on land-cover and land-cover change through this approach. Therefore, approaches that rely on expert knowledge can be difficult to scale without prior information on what crop rotations are practiced in a region (Xiao et al., 2014), and are susceptible to biases or gaps in that information. (c) Crop rotation patterns determined through remotely-sensed data like CDL: While spatially explicit and generally accurate for major production crops in the US, combining multiple years of CDL to get crop rotation information results in a large number of factorial crop combinations (Stern et al., 2012). The large number of crop rotations are combined with a wide range of other input choices for chemicals and irrigation to generate management scenarios. Modeling frameworks face computational bottlenecks in efficiently simulating a large number of management scenarios. Some modeling studies try to bypass this limitation by selecting far fewer crop rotations based on their acreage (Srinivasan et al., 2010 and Muth et al., 2013). These studies only use a subset of all existing crop rotations and do not quantify the error introduced in the output as a result. Others greatly simplify each crop rotation pattern to focus on general trends i.e. whether the crop rotation is a monoculture or alternating (biannual, triennial, and quadrennial) in nature (Stern et al., 2008, Mehaffey et al., 2011, Secchi et al., 2011 and Plourde et al., 2013). Further, no attempt has been made in existing literature to produce a parsimonious selection of representative crop rotations for a region with acreage estimates comparable to NASS data.

1.3 Objectives and preview

The objectives of this research are to: (a) quantify the diversity of crop rotation patterns and (b) to examine land-cover transitions occurring in the agronomically productive and ecologically sensitive parts of the WCB. In order to achieve these objectives, we developed a novel two-parameter algorithm for identifying crop rotation patterns in the WCB during a three year period (2010 to 2012). We examined the tradeoff between number of representative crop rotations and accuracy by comparing estimated acreage against CDL and NASS crop acreage data. A sensitivity analysis of the algorithm parameters was conducted to assess its performance across the study region. Finally, we used the crop rotation product to estimate grassland conversion to crop cultivation in a wetland dominated part of the WCB. In summary, we examined the effectiveness of the algorithm in reducing the number of crop rotation patterns required to map and model each state by several orders of magnitude, while adequately capturing crop acreage and land-cover change trends.

2. *Materials and Methods*

2.1 Study area

Annual crop production is the dominant land-use in our study area comprising the WCB states of North Dakota (ND), South Dakota (SD), Nebraska (NE), Minnesota (MN) and Iowa (IA). While accounting for more than two-fifths of the corn and soybean acreage planted nationally in 2012 (http://quickstats.nass.usda.gov/), these states also overlap the Prairie Pothole Region (PPR), an ecologically sensitive wetland landscape (Johnson et al., 2010). CDL data for all five states in the WCB has been available since 2006, providing seven years of land-use/cover information for analysis. Starting in 2007, CDL data are provided statewide with cloud-free coverage. For years prior to 2007, CDL data includes metadata files with information on level and extent of cloud contamination.

2.2 Cropland Data Layer

USDA classifies crop cover types in conterminous US using satellite imagery from a variety of sources including MODIS, Advanced Wide Field Sensor (AWiFS) and Landsat Thematic Mapper (TM). To produce the annual CDL product, ground truth information on land-use, acreage and field boundaries is obtained from the USDA farm service agency's Common Land Unit (CLU) dataset. Subsequently, a statewide land-cover classification is produced by performing a supervised classification of the geo-referenced and orthorectified satellite imagery, using the digitized field segments as training samples. By using a comprehensive national ground truth dataset, CDL data achieves producer and user accuracies between 80% - 95% for major agricultural crops that are traded on the US commodity markets, including: corn, soybean, winter wheat (*Triticum aestivum*), cotton (*Gossypium hirsutum*), spring wheat (*Triticum aestivum*) and durum wheat (*Triticum durum*). Crop acreage estimates obtained by counting pixels in the CDL data compare favorably with survey based estimates from NASS (Plourde et al., 2013). The information from the CLU dataset pertains only to agricultural land-cover and is supplemented by the National Elevation Dataset (NED), National Land Cover Dataset (NLCD) percent tree cover and percent impervious products for non-agricultural land-cover classes.

2.3 Data processing

We downloaded the CDL raster data for the WCB states from http://nassgeodata.gmu.edu/CropScape/ (Han et al., 2012). While CDL data are available for the conterminous US starting from 2008, we chose to present our algorithm results and analysis for a three year period (2010 to 2012). Most crop rotations in the US Midwest do not exceed this duration (Plourde et al., 2013). However, the algorithm can be run for any number of years pending availability of CDL data. The resolution of CDL between 2006 and 2009 is 56m, and can be resampled to 30m resolution. The rasters were acquired in Albers conical equal area projection, one file for each year for each state. Each CDL raster is thematic with a unique identifier representing discrete features of crop cover type or in the absence of cropping, the land-cover. A generically defined NonAg land-cover class was created from all non-agricultural land-cover classes in CDL. The CDL classes considered to be non-agricultural included: pasture/hay (CDL code 181), shrubland (code 64), grassland herbaceous (code 171), sod/grassseed (code 59), clovers/wildflowers (code 58), other hay (code 37) and pasture/grass (code 62). This was done because of low confidence in the accuracy of CDL to distinguish between various grassland categories (grass hay, grass pasture, hay/pasture and native grassland). The CDL documentation also suggested using NLCD for all non-agricultural land-cover change analysis[1].

Of the several thousand unique crop rotations that exist for each state (Table 2-1), we wanted to select a subset of representative crop rotations which while much fewer in number, provide acreage estimates comparable to CDL. We present the **Re**presentative **C**rop **R**otations Using Ed**i**t Dis**t**ance (RECRUIT) algorithm to derive representative crop rotations using the CDL. The algorithm was coded in Python and used the ArcPy API from ArcGIS® to perform geospatial analysis. The fully automated algorithm operates as follows (Fig. 2-1):

1. For each state, CDL pixels with the same spatial co-ordinates, but different annual time stamps (2010, 2011 or 2012) were combined in ArcGIS®, producing a unique output value for each unique combination of input values. Each of the output values denotes a distinct crop rotation. Since there are nearly 100 crop types in the CDL database and we analyzed three years of data, the number of possible crop rotations is nearly 1 million (100^3). However, when calculated using multi-year CDL data, there are 13,588 crop rotations ranging from 0.09 ha (1 pixel) to 7.5 million ha (83,278,895 pixels) in size (Table 2-1)

in the WCB. The number of unique crop rotations in the WCB is not the sum of its constituent states because the same crop rotation can exist in different states. The pixels corresponding to a particular crop rotation need not be spatially contiguous and can have NonAg as a land-cover type.

2. For each crop rotation, we computed its acreage by multiplying the cell size of the raster (30m x 30m or 0.09 ha) with the number of pixels corresponding to that rotation. We also computed the percentage of the total cultivated acreage occupied by that crop rotation and then rearranged the crop rotations by descending order of acreage.

3. To determine the representative crop rotations, we selected crop rotations starting from the one with the largest acreage, till either one of two threshold parameters was exceeded. The first parameter is the *cumulative acreage threshold* (α), defined as the percentage of the total acreage occupied by the rotations selected thus far. A value of 0% for this parameter implies that no crop rotation will be selected, while all crop rotations will be selected at 100%. The second parameter is the *marginal increase threshold* (β), defined as the marginal increase in acreage by selecting the current rotation. To assess parameter sensitivity, the algorithm was run for ten cumulative acreage threshold values (0%, 15%, 30%, 45%, 60%, 75%, 80%, 90% and 100%), and eight marginal increase threshold values (0%, 0.5%, 1%, 5%, 10%, 20%, 60% and 100%).

4. After selecting the representative rotations, a large number of rotations remained unaccounted for. Therefore, we used the reclassify command in ArcGIS® to convert each of them to a rotation from the representative list they were most similar to. The edit distance metric is used to quantify the similarity by counting the minimum number of insertions, deletions and substitutions required to transform one crop rotation into another (Ristad and Yianilos, 1998). For example, consider the following three year crop rotations: corn-soybean-corn and a continuous corn rotation. These differ in the middle year, and can be made the same by converting from soybean to corn or vice versa. Therefore the edit distance for this example would be one. This step increased the accuracy of the representative crop rotations and is unique to our algorithm. However, it was also the most time consuming part of RECRUIT, since its performance is directly proportional to the square of the number of years of CDL data used as input. In case of a tie, we reclassified the rotation to the representative rotation with the larger acreage. While we considered crop rotations that are mirror images of one another: e.g. corn-soybean-corn and soybean-corn-soybean to be unique, the RECRUIT tool can optionally merge them.

5. After reclassifying all the remaining rotations, RECRUIT can create a wall-to-wall product by merging the crop rotation raster with non-agricultural land-cover pixels (forests, grasslands, wetlands and urban areas) from the CDL rasters.

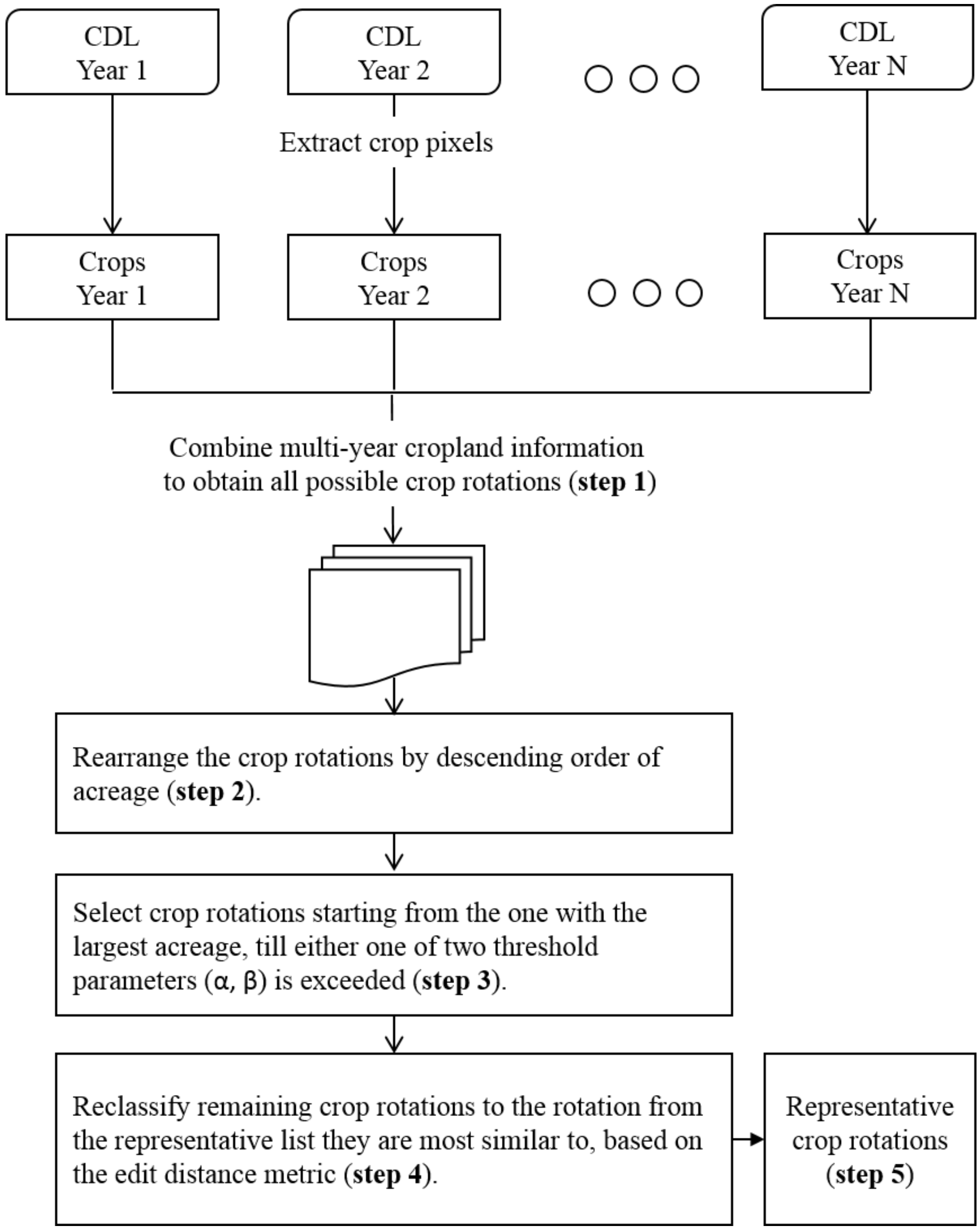

Fig. 2-1. Flowchart of the RECRUIT algorithm. Individual steps of the algorithm are indicated in bold font.

Table 2-1. Number of unique crop rotations by state and region.

	Number of rotations	Largest rotation	Largest rotation (ha)	Smallest rotation (ha)	% of cultivated area occupied by largest rotation
ND	9,619	soybean-spring wheat-soybean	345,046	0.09	4.1%
SD	6,091	corn-soybean-corn	942,059	0.09	15.4%
NE	3,906	corn-soybean-corn	1,410,024	0.09	19.5%
MN	5,554	corn-soybean-corn	1,774,178	0.09	22.9%
IA	1,108	corn-soybean-corn	3,077,256	0.09	33.4%
WCB	13,588	corn-soybean-corn	7,495,100	0.09	19.3%

We computed the accuracy of the crop rotation product as the percentage of pixels for which our crop rotation product differs in its land-cover class compared to the CDL, averaged across multiple years of rotation. For example, consider the following three year rotations: corn-soybean-corn and soybean-soybean-corn occupying 100 and 20 ha respectively. If we merge the second rotation into the first, we will get a representative rotation of corn-soybean-corn. Our accuracy for the new rotation would be 94.4%, because out of the total acreage of 360 ha over the 3 year duration, only the soybean fields spread over 20 ha in the first year would have been incorrectly reclassified as corn, while the remaining 340 ha are classified correctly. If we did not merge the second rotation into the first and used the first rotation to map the entire area, the soybean-soybean-corn rotation spread over 60 ha across the three-year duration would be unaccounted for, leading to an accuracy of 83.3%. Therefore, the marginal increase in accuracy by merging the second rotation into the first is equal to the difference in accuracies for the two scenarios or 11.1%.

3. *Results and Discussion*

3.1 Crop rotation acreage variation across the WCB

The mechanics of the RECRUIT algorithm can be better understood by examining the variation in crop rotation acreages across the WCB. We plotted the marginal contribution of each crop rotation to the total cultivated acreage in Fig. 2-2. A gradual increase, as is the case in North Dakota, indicates a wider diversity of crop rotations as compared to Iowa where the largest crop rotation by acreage accounts for one-third of the cultivated area (Table 2-1), while the smallest crop rotation in each state reflects the pixel size (0.09 ha). The RECRUIT algorithm is based on selecting the crop rotations with the greatest marginal contribution as representative of the entire set.

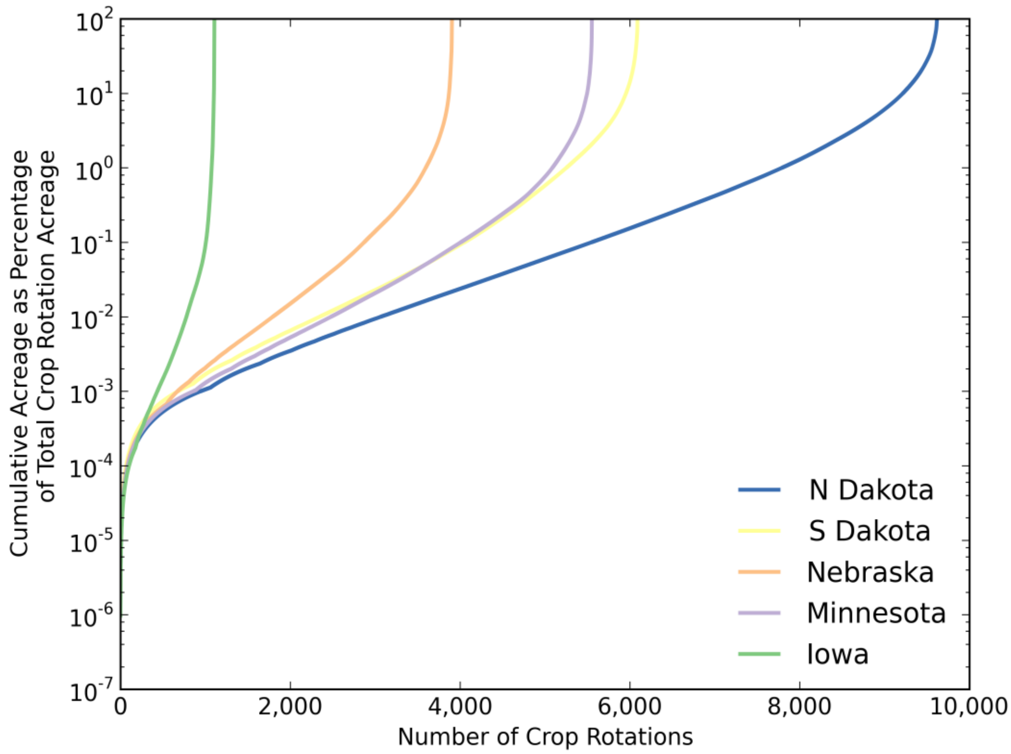

Fig. 2-2. Marginal contribution of each crop rotation to the total cultivated acreage of the state.

The crop rotations representative of the WCB from 2010 to 2012 were mapped (Fig. 2-3). They were selected by running the RECRUIT algorithm, with parameter values α (75%) and β (5%). Each pixel in the representative crop rotation raster had three attribute values: one land-cover for each year from 2010 to 2012. Overall, 75.2% of the pixels in the representative crop rotation product had the same land-cover as the CDL data from which it was derived. The acreage estimate for each representative crop rotation, listed by state, is presented (Table 2-2). Nearly 24.2 million ha or 62.5% of the total cultivated acreage is occupied by corn-soybean-corn and soybean-corn-soybean rotations. From an agronomic point of view these rotations are the same. However, were the latter rotation merged into the former, the effective area devoted to corn production would increase, thus reducing the total acreage devoted to soybean production by nearly 2 million ha/yr. Several crop rotations, e.g. continuous soybean were shown to have zero acreage in several states (Table 2-2). This does not imply that no such crop rotation exists in that state. Instead, its acreage was low enough for RECRUIT to combine it with a higher frequency crop rotation based on their similarity.

Table 2-2. Crop rotations from fig. 2-2 with acreage figures for each state and arranged in descending order of its total acreage in the WCB. Crop names are given as abbreviations: Swht - Spring Wheat, Soyb - Soybean, Cano - Canola. NonAg includes all includes all non-agricultural land-cover classes that are in rotation with crops for this three year period.

2010	2011	2012	ND	SD	NE	MN	IA
Corn	Soyb	Corn	925,987	2,544,827	3,138,292	3,641,786	3,769,169
Soyb	Corn	Soyb	766,928	1,511,005	1,458,197	2,369,365	4,146,642
Corn	Corn	Corn	0	389,559	1,662,749	707,676	1,301,801
Soyb	Swht	Soyb	2,457,665	0	0	536,270	0
Swht	Swht	Swht	1,698,744	0	0	0	0
Soyb	Corn	Corn	0	331,269	547,324	485,527	0
Swht	NonAg	Swht	1,150,006	0	0	0	0
NonAg	NonAg	Corn	0	717,099	0	0	0
Corn	Corn	Soyb	0	180,132	424,756	0	0
Swht	Soyb	Swht	511,866	0	0	0	0
Soyb	Soyb	Corn	214,859	186,577	0	0	0
Swht	Canola	Swht	320,383	0	0	0	0
Soyb	Soyb	Soyb	314,193	0	0	0	0
NonAg	Corn	Soyb	0	241,432	0	0	0
Swht	Soyb	Soyb	156,184	0	0	0	0

There are significant land-cover changes taking place in the WCB as a result of grassland conversion to cultivated areas (Wright and Wimberly, 2013). There are three representative crop rotations with NonAg land-cover class for at least one year (Fig. 2-2, Table 2-2). This could either indicate a fallow year which was misclassified as grassland or it could be Conservation Reserve Program (CRP) land coming out of enrollment. Much less frequently, the NonAg class could denote a land-cover that is highly unlikely to be part of a rotation e.g. urban. Using the edit-distance metric to combine low frequency rotations with the representative ones automatically adjusts for the rotations that are probably misclassified.

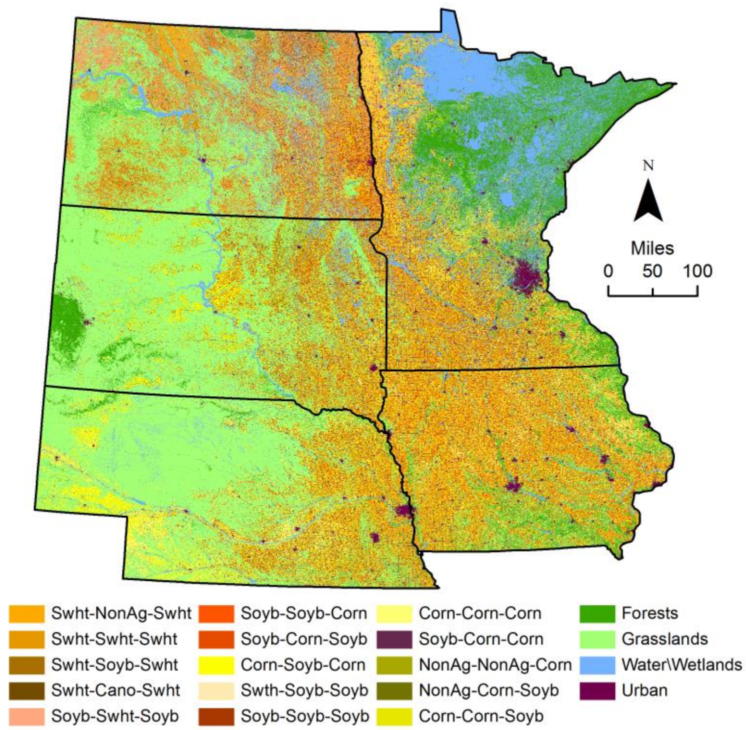

Fig. 2-3. Crop rotation patterns across the WCB from 2010 to 2012. Crop names are given as abbreviations: Swht - Spring Wheat, Soyb - Soybean, Cano - Canola. NonAg includes all non-agricultural land-cover classes that are in rotation with crops for this three year period.

3.2 Accuracy of representative crop rotations

Regional variability in the choice and acreage of crops planted by farmers manifests itself in the observed crop rotation patterns. In Iowa, selecting two representative crop rotations (corn-soybean-corn and soybean-corn-soybean) achieve ~85% accuracy (Fig. 2-4). In contrast, more than 100 representative crop rotations are needed to map North Dakota at a similar accuracy (Fig. 2-4). In all, 400 data points were plotted, representing RECRUIT results for 80 combinations of α and β for each of the five states. The number of visible data points in fig. 2-4 is less than 400 because several α and β combinations result in the same accuracy. The logarithmic x-axis implies that any additional accuracy gains involve a tradeoff, thereby greatly increasing the number of representative crop rotations. For example 1,108 crop rotations can map Iowa at 100% accuracy (Table 2-1), compared to just two for 85% accuracy.

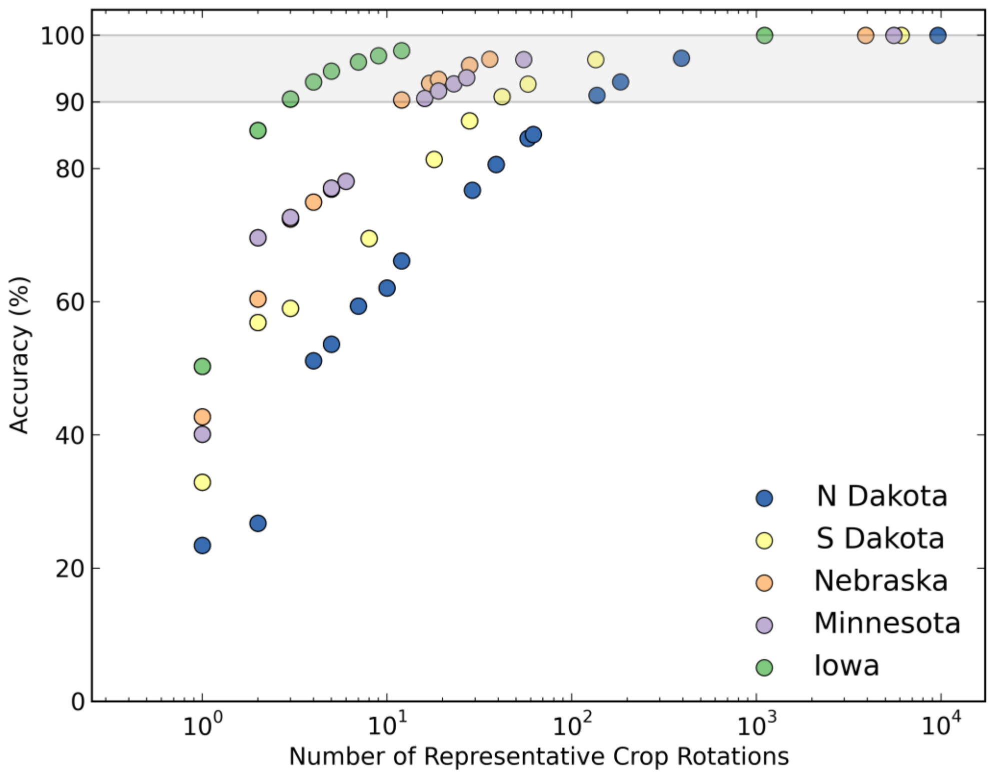

Fig. 2-4. Tradeoff between product accuracy and the number of representative crop rotations selected for each state in the WCB. The gray band indicates the number of representative crop rotations for different combinations of α and β, with accuracy exceeding 90%.

The accuracy of the WCB crop rotation product varies as a function of the two threshold parameters (Fig. 2-5). At one extreme, on the lower left corner of the heatmap, we do not select any crop rotation resulting in an accuracy of 0%. On the lower right corner, we select all crop rotations as representative resulting in an accuracy of 100%. Between these two extremes, the number of representative crop rotations varies by three orders of magnitude. The isolines for accuracy levels of

60%, 70%, 80% and 90% indicate the α and β values for attaining specific accuracy targets. Notably, part of the 60% isoline is perfectly horizontal. This is because no crop rotation in the WCB has a marginal contribution exceeding 60% of the acreage of the representative crop rotations selected before it. Since RECRUIT selects representative crop rotations till either the cumulative acreage or the marginal increase in acreage threshold is exceeded, at high values of α (> 60%), we need to select a β value of 5% or lower to obtain an accuracy exceeding 80% (Fig. 2-5). For lower values of α, the value of β is not as much of a constraining factor for product accuracy.

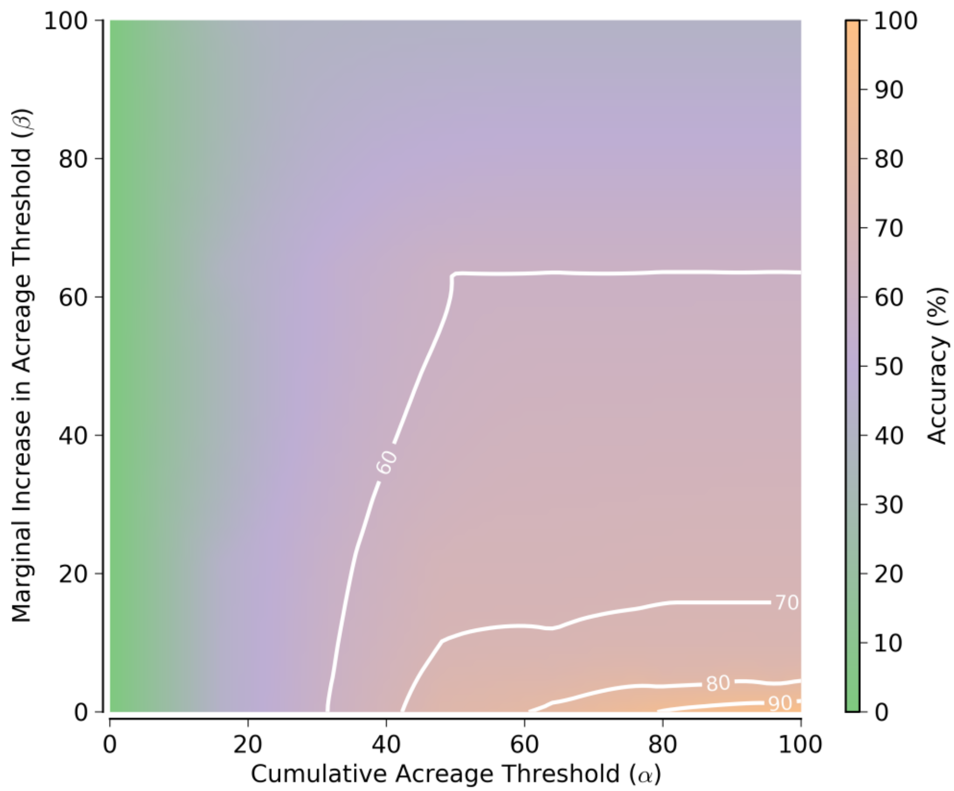

Fig. 2-5. Variation in accuracy of the WCB crop rotation product as a function of α and β parameters used in RECRUIT. The contours indicate the isolines for accuracy levels of 60%, 70%, 80% and 90%.

Merging the low frequency crop rotation with the representative set increases the accuracy of the crop rotations product (Fig. 2-6). Each of the eight panels (A to H) plots the percentage marginal increase in accuracy for a distinct value of the marginal increase in acreage threshold (0%, 0.5%, 1%, 5%, 10%, 20%, 60% and 100%) while the cumulative acreage threshold varies. In each panel, the percentage marginal increase in accuracy rises from 0% when α is 0% to around 25% or slightly higher when α is 15%. The higher percentage marginal increase for Nebraska rather than Iowa can be understood on the basis of the much greater diversity and number of crop rotations in Nebraska (Fig. 2-2). There are simply more crop rotations to merge with the representative set in the case of Nebraska. In panel A, the percentage marginal increase in accuracy drops to 0% when α is 100% because all crop rotations have been selected as representative at that point, leaving no additional crop rotations to merge.

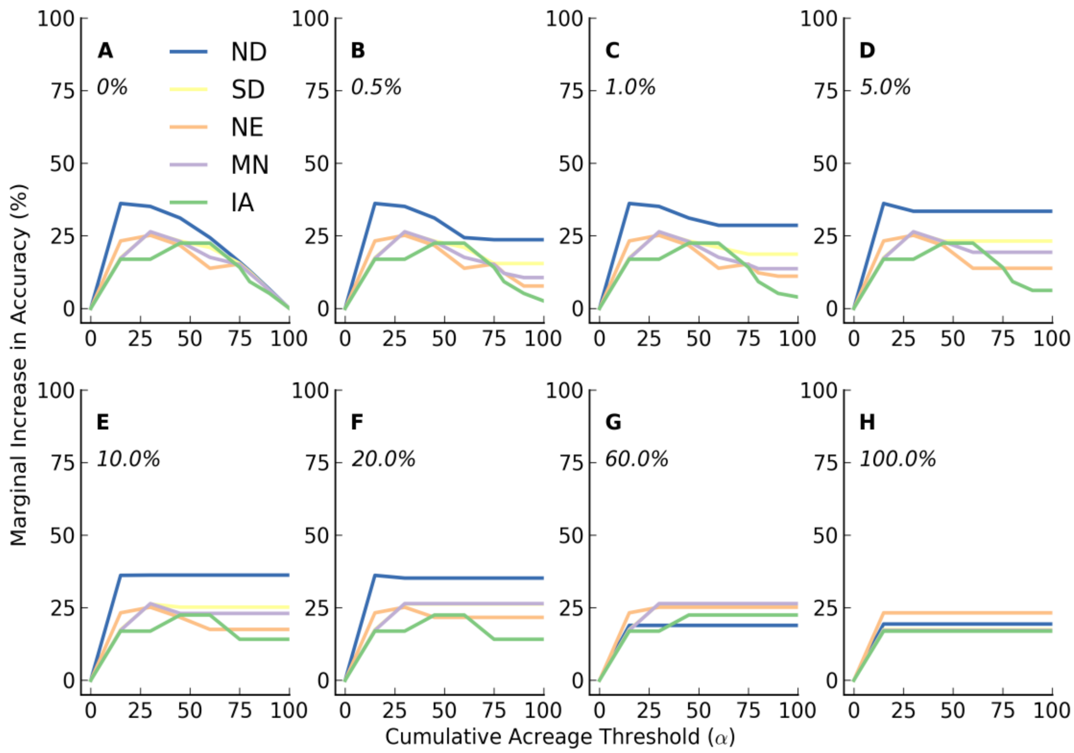

Fig. 2-6. Marginal increase in accuracy of crop rotation product for each state, obtained by combining all crop rotations with the representative crop rotations. Each of the eight panels shows a distinct value of the marginal increase in acreage threshold (β).

While the percentage marginal increase in accuracy puts a numeric estimate on the benefit of merging all crop rotations with the representative set, it is useful to look at the percentage accuracy of the product as it will ultimately drive our selection of RECRUIT parameter values. Full accuracy is achieved when α is 100%, corresponding to a 0% marginal increase in accuracy (Fig. 2-7(A)). Consistent with fig. 2-4, accuracy scores are generally the lowest for North Dakota and highest for Iowa. Also consistent with observations in fig. 2-6, percentage accuracy starts to plateau sooner as the value of β increases.

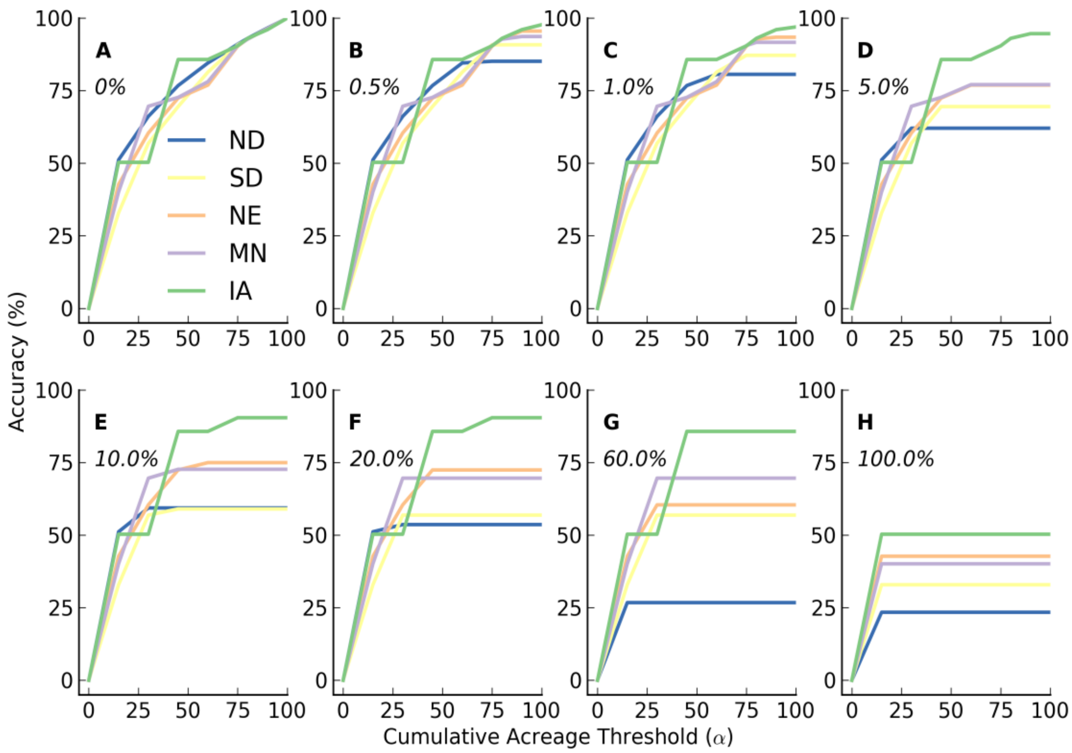

Fig. 2-7. Accuracy of crop rotation product for each state, obtained by combining all crop rotations with the representative crop rotations. Each of the eight panels shows a distinct value of the marginal increase in acreage threshold (β).

Crop acreage estimates obtained by counting pixels in the CDL are typically downward biased due to imperfect classification, cloud contamination or interference from neighboring pixels (Townshend et al., 2000). Since the crop rotation product is a simplification of a time-series of CDL rasters, we compare acreage estimates from the CDL and our crop rotation product against NASS survey data. We observe a 3% average difference in corn acreage and 4.1% average difference in soybean acreage from 2010 to 2012 between the 13,588 crop rotations in CDL (Table 2-1) and NASS (Fig. 2-8). The corresponding differences for a crop rotation product with 82 representative rotations and overall accuracy exceeding 90%, was 7.6% for corn and

9.24% for soybean. Reducing the number of representative crop rotations to 15 increased the average difference in acreage to 21.7% for corn and 18.4% for soybean (Fig. 2-7).

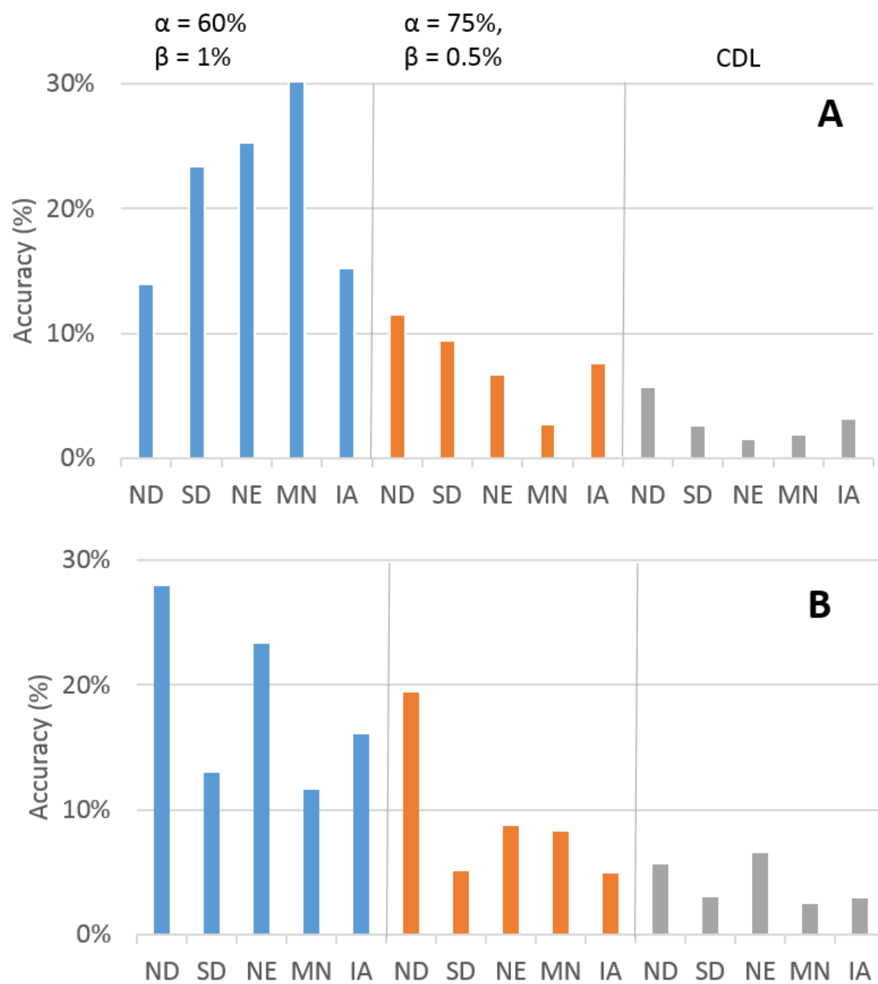

Fig. 2-8. Magnitude of average difference in crop acreages from 2010 to 2012 between the crop rotation product and NASS (blue and orange bars), and CDL layers for each year and NASS (gray bars). Panels A and B shows the average difference for corn and soybean acreages respectively. The α and β parameter values corresponding to the blue and gray bars are shown on top of panel A.

3.3 Pronounced shifts from grassland to cultivated areas in the prairie pothole region

Crop rotation patterns can provide information on land-cover change happening in ecologically sensitive areas. We used the crop rotation product identified in fig. 2-8 (82 crop rotations with overall accuracy > 90%, α(75%), β(0.5%)) to estimate grassland conversion to specific crop rotation patterns in the PPR. Recently, Wright and Wimberly (2013) reported that contemporary grassland conversion to corn and soybean cropping (GRCS) in the PPR is concentrated within a 500m area surrounding wetlands. If land-cover conversion in the PPR was induced by increased demand for corn and soybean, we should be able to observe intensive cultivation of corn and soybean in the area. To test this hypothesis, we selected representative crop rotations from 2010 to 2012 for the PPR with an overall accuracy of around 90%. Using the CDL, we found nearly 147,267 ha undergoing GRCS in close proximity to wetlands between 2010 and 2012. Of the lands undergoing GRCS, nearly 38% was devoted to continuous corn or continuous soybean cropping and another 27% was covered by alternating corn and soybean rotations (Table 2-3). These patterns are very surprising especially given the widespread consensus that continuous cropping of soybean for more than two years is not a viable choice on account of enhanced parasite activity (Secchi et al., 2011).

Table 2-3

Crop rotations in the GRCS pixels in PPR. The GRCS pixels lie within a 500m radius of wetlands. Soybean is abbreviated as Soyb.

2010	2011	2012	Area (ha)
Soyb	Soyb	Soyb	39,173
Corn	Corn	Corn	16,767
Soyb	Corn	Soyb	25,846
Corn	Soyb	Corn	13,916

4. *Discussion of limitations and future work*

The RECRUIT algorithm is temporal range agnostic. While we present results for crop rotations from 2010 to 2012, it is possible to use CDL rasters starting from 2008 for any state in the coterminous US, and going back to 1997 for states like North Dakota. Using RECRUIT over longer time periods (> 3 years), can produce patterns that include repetitions of the same crop rotation, e.g. a corn-soybean rotation repeated thrice over a six year period. Our approach can be enhanced in the future by initially identifying the repetitions within a crop rotation pattern and subsequently applying the edit-distance based approach.

The methodology, software and input data used to produce CDL has evolved during the fifteen year history of the product and it is important to understand how that can affect the accuracy of the derived crop rotation product. First, from 2006 to 2009, 56m resolution AWiFS sensor data was used to produce the CDL. An imagery comparison between the two sensors (AWiFS and Landsat-5 TM) shows that they produce equivalent results for agricultural and forestry applications despite the differences in the design of the two sensors (Johnson, 2008) and the radiometric degradation over time of the AWiFS sensor (Goward et al., 2012). Second, because of the relatively large swath width of the AWiFS sensor (740 km), angular effects arising from variable solar and viewing geometry can potentially cause significant changes in observed reflectance without any change in land-cover, phenological status or vegetation condition (Gao et al., 2013). Wheat shows stronger angular effects than corn in both red and NIR bands (Gao et al., 2013), perhaps contributing to its lower producer/user accuracies in the CDL. Third, land-cover change analysis using CDL data prior to 2006 can be subject to errors because of a change in processing software and algorithms in 2006 (Boryan et al., 2011). Fourth, while CDL has producer and user accuracies between 85%-95% for major row crops like corn and soybean, its accuracies are quite a bit lower for specialty crops, fruits and vegetables. Finally, CDL performs poorly in distinguishing between the subtle spectral signature variations amongst the various non-agricultural land-cover classes, especially in the grassland category: pasture/hay, fallow/idle cropland and native grassland. It is recommended to use NLCD for land-cover change analysis involving non-agricultural or grassland/pasture categories1.

Several opportunities exist to use the crop rotation products generated by RECRUIT. Meehan et al. (2013), used the product to examine the changes in ecosystem services due to land-use conversion from annual to perennial biofuel feedstock in the US Midwest. Zhang et al. (2014), combined the product with other spatially explicit data (climate, soils and topography) to examine the sensitivity of an agro-ecosystem model to soil data input. These applications illustrate that as an input to process-based or empirical models in a carbon monitoring system, crop rotations can be part of an attributional framework (West et al., 2013), linking ecosystem function outcomes to originating sources (Zhang et al., 2010) at multiple scales including individual fields (Daggupati et al., 2011). Further, by determining areas where non-agricultural land-covers exist in rotation with cultivated fields, we can improve the idle/fallow cropland mapping and CRP mapping process in CDL. Also, mapping of tillage extent and intensity patterns can be greatly improved by using crop rotations to produce expert-rules that determine probability of a field being tilled (Johnson, 2013). Finally, field boundaries can be generated using the crop rotations product under the assumption that each field has a unique crop rotation pattern over time.

5. *Conclusions*

Modeling approaches are increasingly using crop rotation patterns instead of single crops to estimate the environmental impacts of agricultural activities. Fine resolution, spatially explicit management information can help constrain the modeling uncertainty of spatial and temporal patterns and magnitude of terrestrial carbon fluxes. However, computational and analyst resource bottlenecks constrain modeling frameworks in the number of crop rotations they can efficiently simulate. To address this problem, we present the RECRUIT algorithm which uses multi-year CDL data to identify representative crop rotations in the WCB. We find that a small number (82) of representative crop rotations can account for over 90% of the spatio-temporal variability of the more than 13,000 rotations in the WCB (Fig. 2-8). The area estimates of the representative crop rotations are comparable to those from agricultural census data. Using the crop rotation product, we are able to detect pronounced shifts from grassland to monoculture corn and monoculture soybean cultivation within the last few years in the ecologically sensitive PPR. Given the novel capability of RECRUIT to flexibly and efficiently derive representative crop rotation patterns in a spatially and temporally explicit manner, it is expected to be a useful tool for providing input data to drive agro-ecosystem models and for detecting shifts in cropping patterns in response to environmental and socio-economic changes.

Chapter 3: Poplar Biomass Productivity Across U.S. Midwest: High-resolution Modeling Approaches and Tradeoffs Across Spatial Scales

1. *Introduction*

1.1 Background

The development of renewable energy sources is an integral step towards mitigating the carbon dioxide induced component of climate change (O'Neill and Oppenheimer, 2002). One important renewable source is plant biomass, comprising both food crops such as corn (*Zea mays*) and cellulosic biomass from short-rotation woody crops (SRWC) such as hybrid-poplar (*Populus* spp.) and Willow (*Salix* spp.). Cellulosic feedstocks, (hereafter referred to as second generation or 2G), represent an abundant and if managed properly, a carbon-neutral and environmentally beneficial resource that is expected to meet about 12 percent of global transportation fuel consumption by 2050 (Demirbas, 2008). Within the U.S., congressional legislation has set an ambitious target for biofuel production, mandating a ramping up of biofuel ethanol production from nearly 14 billion gallons in 2011 to 36 billion gallons by 2022, with at least 16 billion gallons coming from 2G feedstocks (USEPA 2010).

While there are several promising 2G candidates including corn stover, switchgrass (*Panicum virgatum*), miscanthus (*Miscanthus giganteus*) and native prairie mixes, several reasons make SRWCs an excellent choice for producing cellulosic ethanol. First, they have market acceptability as they have been under cultivation in large parts of U.S. for much longer in the past (Riemenschneider et al., 2001). Second, their overall energy balance is better (Hill et al., 2006), due to lower transportation costs and need for fewer agricultural inputs (Wang et al., 2013). When compared to conventional ethanol produced from food crops, SRWCs can reduce greenhouse-gas (GHG) emissions (Gelfand et al., 2013), improve nutrient retention, biodiversity and reduce soil erosion (Isebrands et al., 2001).

1.2 Poplar as a 2G feedstock

Amongst the various SRWC candidates, hybrid-poplar is a strong candidate for sustainable biofuel production. They grow fast, with rotation times ranging from 4 to 10 years (Jug et al., 1999), and with yields ranging from 10 – 15 Mg/ha/yr for the temperate regions of Europe and North America (Kauter et al., 2003, Aylott et al., 2008 and Tallis et al., 2012). Further, they can be grown on a wide variety of soils (Hansen et al., 1988), including marginal lands that are typically less suitable for row crop cultivation and therefore less likely to compete with food production or incur carbon debt (Fargione et al., 2008). Other traits like high light-use efficiency and photosynthetic capacity also make them a strong candidate as potential biofuel feedstock. Notably, recent scientific developments have yielded poplar clones with lignin that is more amenable to degradation and conversion to biofuels (Wilkerson et al., 2014), making it likely that the first large-scale cellulosic ethanol refinery will use poplars as a feedstock.

1.3 Consideration for modeling poplar plantations

In order to produce cellulosic ethanol from poplars in a sustainable manner, it is important to understand how their productivity will vary across a landscape. There are a growing number of field studies documenting growth, yield and environmental impacts of poplar plantations across varying edaphic and climatic conditions in the U.S. (Gamble et al., 2014, Palmer et al., 2013, Palmer et al., 2014, Xue et al., 2014 and Kaczmarek et al., 2013). While very useful in establishing ground truth, we need to extrapolate from them and predict poplar yields and environmental impacts across entire regions. One solution is to carefully calibrate a physical process based models to produce regional, national or even global impact assessments of different biofuel feedstock configurations. The standard approach is to divide the study area into a grid, with the cell size of the grid dependent on data availability, resolution as well as computational resources required to run the model for each grid cell. In the context of poplar plantations, previous modeling work has included (1) Using the 3-PG model to map poplar biomass productivity in Minnesota and Wisconsin (Headlee et al., 2013); (2) Inferring potential yields, rotation times and soil C sequestration potential of poplar for the U.S. (Wang et al., 2013); (3) Comparing water use efficiency of poplar and willow in the UK (Tallis et al., 2012). A common theme in these studies is the use of field scale data to calibrate the model but dropping down to much coarser scale data when extrapolating to a region or country. This is partly due to data availability. Field sites often have finely curated and extensive information on a variety of soil, climate and vegetation parameters, and it is infeasible both in terms of time and money to produce comparable data across broader spatial scales. However, even when high resolution data is available, for instance the Soil Survey Geographic (SSURGO) dataset in the U.S., modeling studies often either use coarser scale data, or use discrete soil property values corresponding to specific soil textures, or average the variables in the fine resolution dataset to match the resolution of a much larger climate grid cell (Wang et al., 2013). This is particularly the case with modeling studies focused on SRWC ecosystems, because the models used in such studies have been traditionally used to model forest ecosystems with less emphasis on sub-grid scale edaphic heterogeneity. Most such models typically use a small subset of available soil property information e.g. soil texture, to simulate ecosystem processes (Cramer et al., 2001). This is in contrast to agroecosystem models which need to consider finer scale details on a much wider subset of the input soil data in order to understand management impact on the sensitivity of simulated yields and carbon fluxes (Zhang et al., 2010 and Zhang et al., 2014). Therefore, it is critical to consider both the impact of averaging and discretization of soil properties in the simulation of yields in a SRWC ecosystem, and map out how this differs for different soil databases used as input in various modeling studies.

1.4 Aims of this study

Details matter! A canonical problem in spatially explicit studies is the modifiable areal unit problem (MAUP) which introduces statistical biases in the results because of the choice of grid cell boundary dimensions (Fotheringham and Wong, 1991 and Jelinski and Wu, 1996). While the issues introduced by MAUP remain unsolved for a broad spectrum of spatial modeling endeavors, our aim in this study is to demonstrate its impact or lack thereof on ecosystem modeling of poplar plantations in the U.S.. The specific objectives of the present study are to (1) parameterize the Ecosystem Demography (ED) model to predict potential non-nutrient limited yields for hybrid poplar across the U.S. Midwest; (2) Determine quantitatively how estimates of saturated conductivity (a key soil property measuring a soil's ability to transport water) vary for two soil datasets that differ in their spatial resolution and method of estimating soil properties: SSURGO and World Inventory of Soil Emission Potentials (WISE); (3) Examine how saturated conductivity estimates changes based on the resolution at which they are averaged or discretized; (4) Estimate the impact of averaging and discretization of saturated conductivity on the relative accuracy of poplar yield estimates obtained from ED.

2. *Materials and methods*

2.1 Study area

Annual crop production is the dominant land-use in our study area comprising the ten U.S. Midwest states of North Dakota (ND), South Dakota (SD), Nebraska (NE), Minnesota (MN), Iowa (IA), Wisconsin (WI), Michigan (MI), Indiana (IN), Illinois (IL) and Ohio (OH). While accounting for more than half of the corn and soybean acreage planted nationally in 2012 (http://quickstats.nass.usda.gov/), these states also overlap the Prairie Pothole Region (PPR), an ecologically sensitive wetland landscape (Johnson et al., 2010). Cellulosic biofuels are an expanding industry in the region, aided by existing farming infrastructure and presence of ethanol refineries capable of ingest cellulosic feedstocks. While not as prevalent as in south-east U.S., poplar plantations have been present in the region for more than a decade (Netzer, 2002), with more established in recent years to account for complete GHG balance of the system (Palmer et al., 2013).

2.2 Model

We used the Ecosystem Demography (ED) model (Moorcroft et al., 2001, Hurtt et al., 2002) to simulate potential biomass from poplar plantations across the U.S. Midwest. ED is an individual-based, terrestrial biosphere model (fig. 3-1). It can estimate plant growth, phenology, mortality, belowground C and N dynamics and disturbance at the level of a single tree and scale up to an entire ecosystem to estimate changes in population structure and community composition, while simultaneously modeling natural disturbances, land use, and the ecosystem dynamics lands recovering from disturbances.

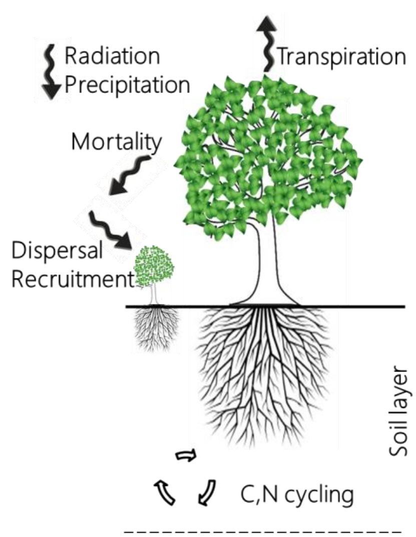

Fig. 3-1. Outline of ED model

ED uses a single soil-layer bucket model to simulate water percolation through the soil as mediated by the saturated conductivity of the soil (Klute et al., 1986). Individual plants of different functional types compete mechanistically in ED under local environmental conditions for light, water, and nutrients. These include C3 and C4 plants (as grasses), and early, mid and late successional species. The model has been validated at multiple temporal and spatial scales and used for assessment of ecosystem dynamics in South and Central America (Moorcroft et al., 2001), patterns of tree mortality and its drivers in northeastern North America (Dietze and Moorcroft, 2011), U.S. carbon sink (Pacala et al., 2001), and projections of its future, including the importance of future fire and fire suppression (Hurtt et al., 2002). The model version used in this study is based on the original version designed to evaluate the effects of land-use change on U.S. carbon balance (Hurtt et al., 2002), with modifications to parameterize a new plant functional type (poplar).

2.3 Model data inputs

Soils: SSURGO

A conterminous-scale soil property database was built using soil survey geographic (SSURGO) data from the U.S. Department of Agriculture Geospatial Data Gateway (http://datagateway.nrcs.usda.gov). Extracted soil properties included number of soil layers, depth of each layer, soil texture information and saturated conductivity value.

Soils: WISE

The World Inventory of Soil Emission (WISE) dataset is global in nature, with a spatial resolution of approximately 10km at the equator. Introduced in the 1990's (Batjes, 1996), it has grown in the number of soil properties available as well as filling in the missing information gaps globally (Batjes, 2005 and Batjes, 2009). It has been widely used to parameterize crop simulation models (Gijsman et al., 2007). We obtained soil depth information from WISE and use soil texture information from WISE to index a lookup table relating soil texture to saturated conductivity values (Cosby et al., 1984).

Climate:

To determine the climatic inputs required by ED, we used the observation constrained CRUNCEP dataset (http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm). This dataset has been used in model inter-comparison projects and while relatively coarse resolution at half-degree, it is global in scope (Huntzinger et al., 2013).

2.4 Parameterization of poplar plant functional type

Following Albani et al., 2006, we parameterized the poplar PFT based on height-diameter relationships as specified in Pacala et al., 1996 (table 3-1).

Table 3-1. Parameterization of poplar spp. in ED.

PFT	Allometric equation parameters		
	Leaf biomass equation	*Structural biomass equation*	*Height-DBH equation*
Populus spp.	a_1=0.0047; b_1=2.249	a_1=0.0265; b_1=2.259	b_{1h}=22.68; b_{2h}=-0.0653

2.5 Optimization of model parameters

We used the R package, hydroPSO (Zambrano-Bigiarini and Rojas, 2013), to calibrate ED. Model optimization often suffers from equi-finality, wherein multiple parameter value sets achieve the same model output (Kennedy and Eberhart, 1995), making it difficult to select one. hydroPSO offers several user-friendly diagnostics to help select the best parameter set. The site data used to calibrate ED came from Netzer, 2002, and the individual sites used for calibration and validation are listed below.

Table 3-2. Site data used for calibrating and validating model.

Site	State	Lat	Lon
Ashland	WI	46.6	-90.9
Fargo	ND	46.5	-96.5
Granite Falls	MN	44.4	-95.7
Milaca	MN	45.8	-93.7
Mondovi	WI	44.5	-91.5
Sioux Falls	SD	43.5	-96.5

Based on literature analysis, the parameters listed in table 3-3 were used in the calibration scheme. Their optimized values are listed in the table, alongside the parameter range supplied to hydroPSO. These range from relatively well understood parameters like Vm_0, which controls the rate of carbon assimilation and respiration, to the poorly constrained mortality parameters.

Table 3-3. Parameters used in model calibration.

Parameter		Units	Range	Optimized
GR_{resp}	Growth respiration	–	0.33–0.5	0.34
SD_m	Seedling mortality	–	0.8–0.99	0.82
K_w	Water uptake stress	–	20–200	182.85
N_1	Plant nitrogen uptake parameter	–	0.1–1	0.48
N_2	Microbial nitrogen uptake parameter	–	10–80	40.54
m_1	Mortality coefficient	–	0.01–1	0.79
m_2	Mortality coefficient	–	1–20	6.97
m_3	Mortality coefficient	–	1–30	13.24
B_f	Biomass fraction (above/below ground)	–	0.7–0.8	0.77
Vm_0	Maximum rate of carboxylation	μ molm^{-2}s^{-1}	35–55	53.2
L_t	Leaf-off temperature	°C	8–12	8.21
R_{fract}	Fraction of excess C to seed reproduction	–	0.1–0.5	0.14

2.6 Discretization and averaging soil properties

To keep modeling calculations tractable and be able to run multiple scenarios within a reasonable amount of time, models perform several simplifications. These simplifications can come in several forms: (1) Choice of algorithm for simulating a particular physical process. Using a bucket model to percolation of water through a soil profile is much less computationally taxing than one which considers lateral flows as well; (2) Simplifying an entire process by substituting it with an empirical relation. For instance, seedling mortality is usually simplified in dynamic vegetation growth models into a single parameter value, whereas in actuality it is a complex process by itself (Packer and Clay, 2000). Entire modeling studies are devoted to understanding how such simplifications can affect model outcomes.

While the choice of model algorithms and structure is fairly defined and less amenable to frequent updates, modeling frameworks are far more capable regarding the type, quality and quantity of input data they can use. These frameworks ingest input data and render it into a format compatible with the model, adequately parameterize the model and capture and present model outputs. The smallest spatial domain over which a model operates is typically called a grid-cell. Current research trend is towards highly spatially-explicit modeling, by trying to estimate model outcomes for the finest resolution grid-cell available (McBratney and Pringle, 1999, Zhang et al., 2010 and Gelfand et al., 2013). This entails creating input datasets for each and every unique grid-cell in the study area and simulating them using the model. The benefit of this approach is that it allows us to address the variation in sub-grid scale topographical and edaphic properties. As originally designed, ED does include biotic heterogeneity within each grid cell but does not simulate any abiotic heterogeneity within the grid-cells. While this was a necessity when the model was originally released in 2001 due to the lack of available high resolution soil datasets, it is simply no longer the case. In this study we present a simple yet novel approach to evaluate sub-grid scale edaphic heterogeneity without greatly increasing model complexity or computation time. Our approach is to discretize the soil parameters used by the ED model, and run the model for the entire U.S. Midwest for a factorial combination of soil parameter values, with a model run corresponding to a single value being used for each soil parameter in ED. We will also average the soil properties for each grid-cell, varying its size from 60m to 1/2° (~ 55 km) with intermediate sizes of 1/12°, 1/6°, 1/4°, 1/3° and 5/12° (table 3-4).

Table 3-4. Discretization and averaging of soil properties for running ED.

Dataset	Discretization	Averaging
SSURGO	1, 2, 5, 10, 20, 25, 50	60m, 1/12°, 1/6°, 1/4°, 1/3°, 5/12°, 1/2°
WISE	1	60m, 1/12°, 1/6°, 1/4°, 1/3°, 5/12°, 1/2°

2.7 ED simulations

As noted in the previous section, ED will be run multiple times for the discretization case. ED estimates of above-ground biomass will be collected from these runs and assembled outside in a GIS framework. The resulting raster will be a high-resolution product (upto 60m x 60m in resolution), composed from several hundred ED runs made at a half-degree resolution. The ED runs are made at half-degree resolution to match the spatial resolution of the CRUNCEP climate dataset used in this study. To compare model results from different soil dataset discretization and averaging experiments, we will use a taylor diagram which has been used successfully in the climate change literature to compare model outcomes from multiple climate change models (Taylor, 2001). The taylor diagram plots the pearson product-moment correlation coefficient (R, Pearson et al. 1895) between all pairs of model outcomes, and the standard deviation of each model on a single plot, providing an easy way to visualize how closely a pattern of group thereof matches observations. In our case, since we are directly comparing model outputs, with the only difference being the soil datasets used in the models, we assume that the 60m SSURGO based model output is the standard against which all other model outputs are matched.

3. *Results and discussion*

3.1 Saturated conductivity value distribution in SSURGO and WISE

The SSURGO database contains soil information gathered by trained professionals over the last several decades. This has resulted in a comprehensive dataset at a fine spatial resolution (~60m) and national scale coverage. In comparison, WISE, while being global is much coarser in resolution (~10km). The data collection procedures employed in collecting data which feeds into WISE are also not as harmonized as the SSURGO ones. Saturated conductivity is an important parameter determining rate of flow of water through a soil layer. The difference in soil saturated conductivity parameter values for these two datasets is shown in fig. 3-2. The largest saturated conductivity estimate in WISE is around 15.7 μm/sec, compared to >400 μm/sec in SSURGO. In fact, nearly 30% of the total area in U.S. Midwest is occupied by soils with saturated conductivity values exceeding the largest estimate in the WISE dataset. WISE tends to overestimate the acreage occupied by the lower saturated conductivity soils.

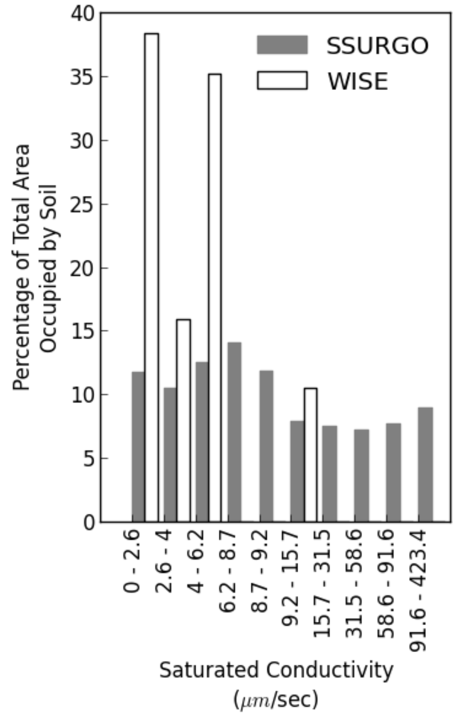

Fig. 3-2. Distribution of soil saturated conductivity values for SSURGO and WISE.

Soil parameter value is also a function of the scale at which soil properties are averaged and fed into a model. In the case of SSURGO, when no averaging is performed, the median saturated conductivity value is around 10μm/sec (fig. 3-3). When upscaled and averaged to be the same resolution as WISE, the median increases to ~ 25μm/sec, while the overall spread decreases. The highest saturated conductivity value in the dataset decreases from around 420μm/sec to 200μm/sec. Any further reduction in resolution upto 1/6°, 1/4°, 1/3°, 5/12°, 1/2° respectively does not affect the median value while progressively reducing the spread of saturated conductivity values. Reduction in the spread of values is a necessary consequence of averaging, however its impact on model output does not necessarily have to follow the same pattern. This is because a single soil parameter like saturated conductivity is one of many affecting model outputs. Further, the sensitivity of the model to the saturated conductivity value will determine variations in model outcome for soil datasets at different resolutions.

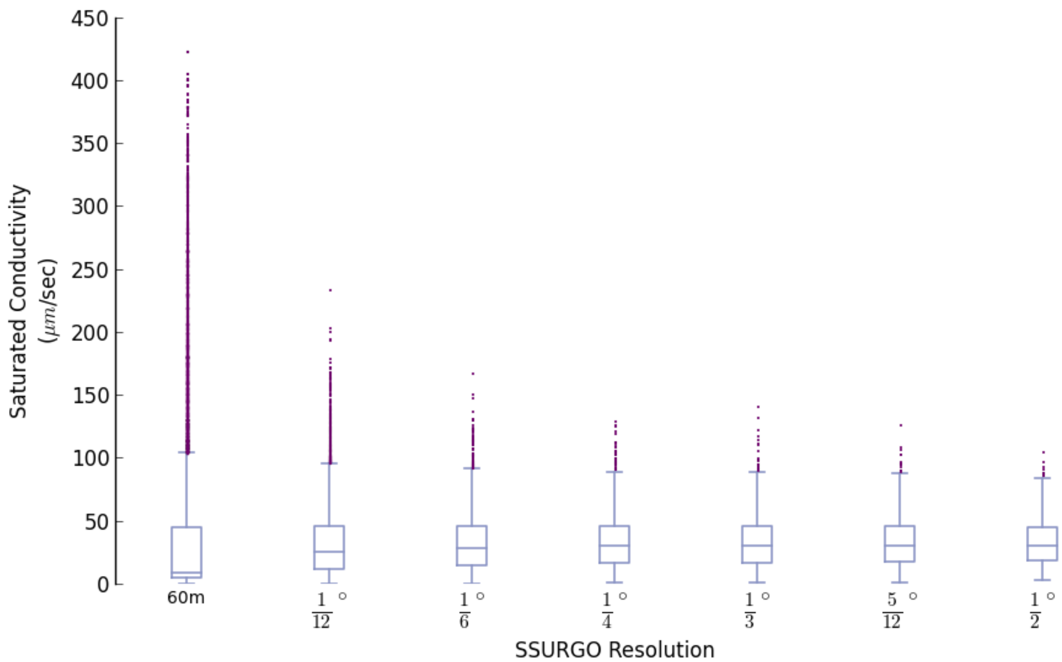

Fig. 3-3. Impact of averaging on saturated conductivity values in SSURGO.

In contrast to SSURGO, the median saturated conductivity value for WISE stays almost constant at 3.2 μm/sec, irrespective of the resolution of the dataset (fig. 3-4). New saturated conductivity values are introduced on averaging, as the initial 1/12° resolution dataset contains the 12 saturated conductivity values obtained from the Cosby et al., 1984 lookup table. The range of values in the highest resolution WISE dataset is an order of magnitude less than the corresponding range in the 60m resolution SSURGO dataset. This implies an even lesser sensitivity of model outcomes to averaging when using the WISE soil dataset as input, when compared to

SSURGO. Notably, when averaged to a half-degree resolution, there is no perceptible difference in saturated conductivity values between WISE and SSURGO (fig. 3-6).

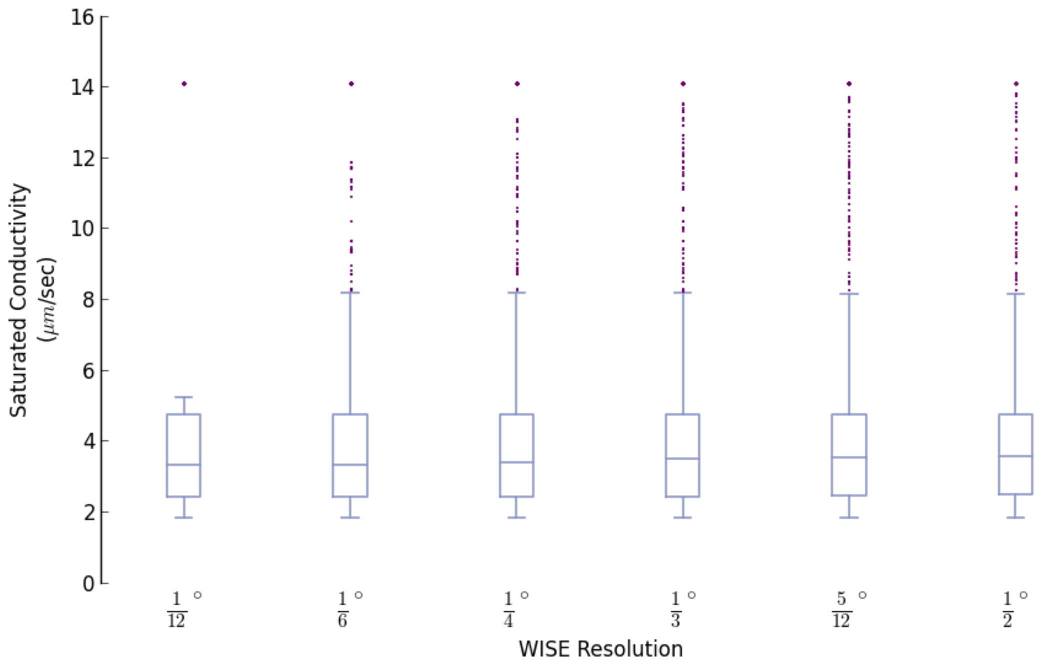

Fig. 3-4. Impact of averaging on saturated conductivity values in WISE.

There are around 1 million unique SSURGO soil identifiers in the U.S. Midwest. Modeling each of them separately would greatly increase the time required to run ED. To reduce the computational burden in this scenario, we developed a simple discretization approach to sample the SSURGO soil parameter value at specific intervals, and rather than modeling all million unique values, just model the sampled few. We vary the sampling intensity from an extreme of just two values representing the entire dataset to a maximum of 100 values, chosen at the corresponding percentile.

What sampling intensity in this discretization scheme is enough? To determine this, we plot a taylor diagram representing the sampling at 100 values (or 1P) as the standard, and comparing the rest against it. The standard deviation increases from ~ 50µm/sec for 1P to greater than 100µm/sec for a sampling intensity of just two values (50P). The correlation coefficient value drop from >0.95 when correlating 2P against 1P, to around 0.4 for the correlation between 50P and 1P. Of course, this reduction in correlation comes with a fifty-fold reduction in the number of saturated conductivity values that need to be used as input data in ED.

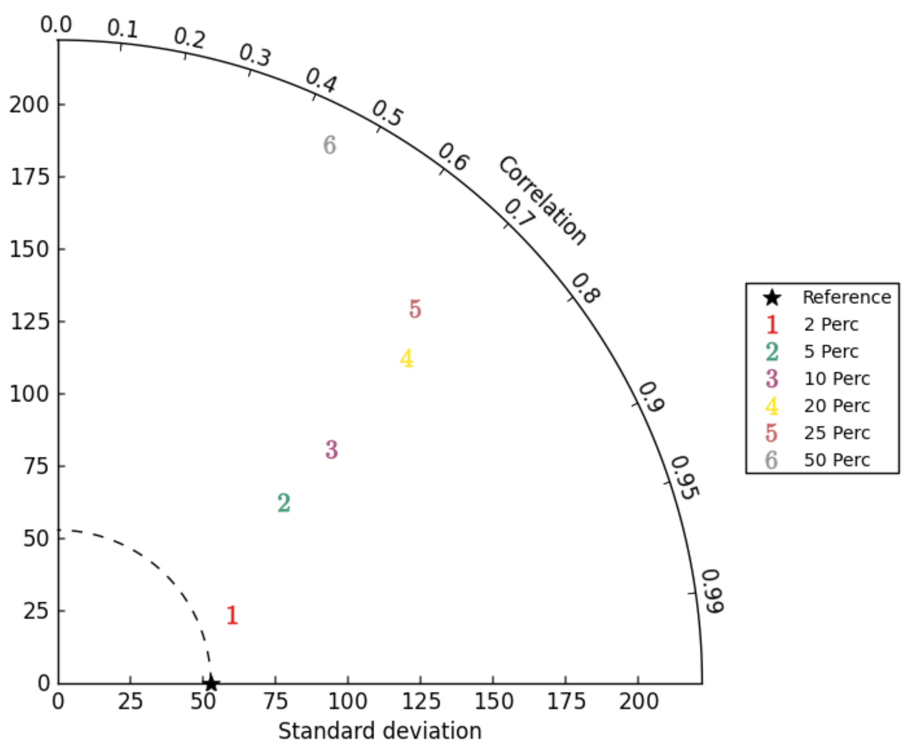

Fig. 3-5. Taylor diagram for discretized saturated conductivity values for SSURGO

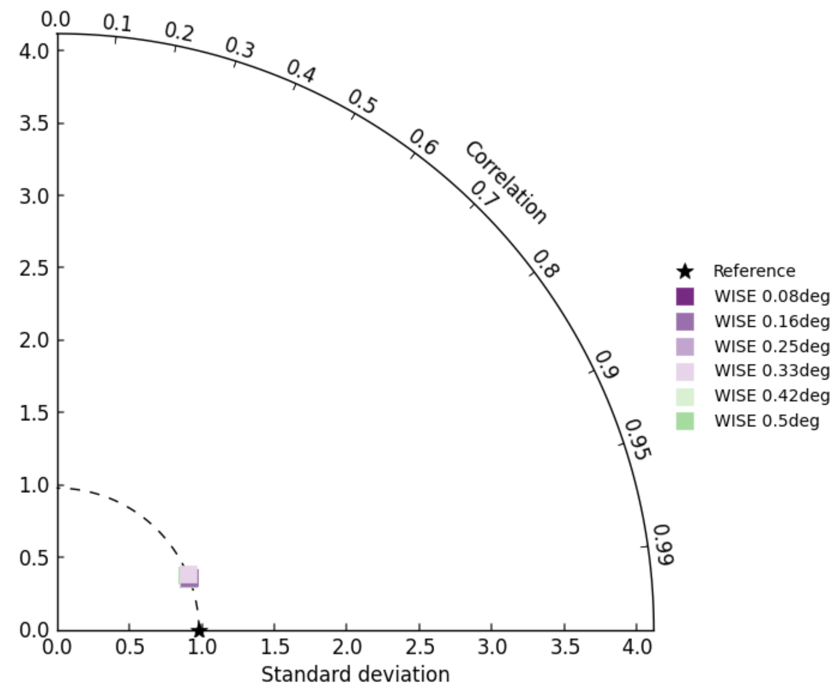

Fig. 3-6. Taylor diagram comparing WISE saturated conductivity values, averaged at various resolutions to SSURGO

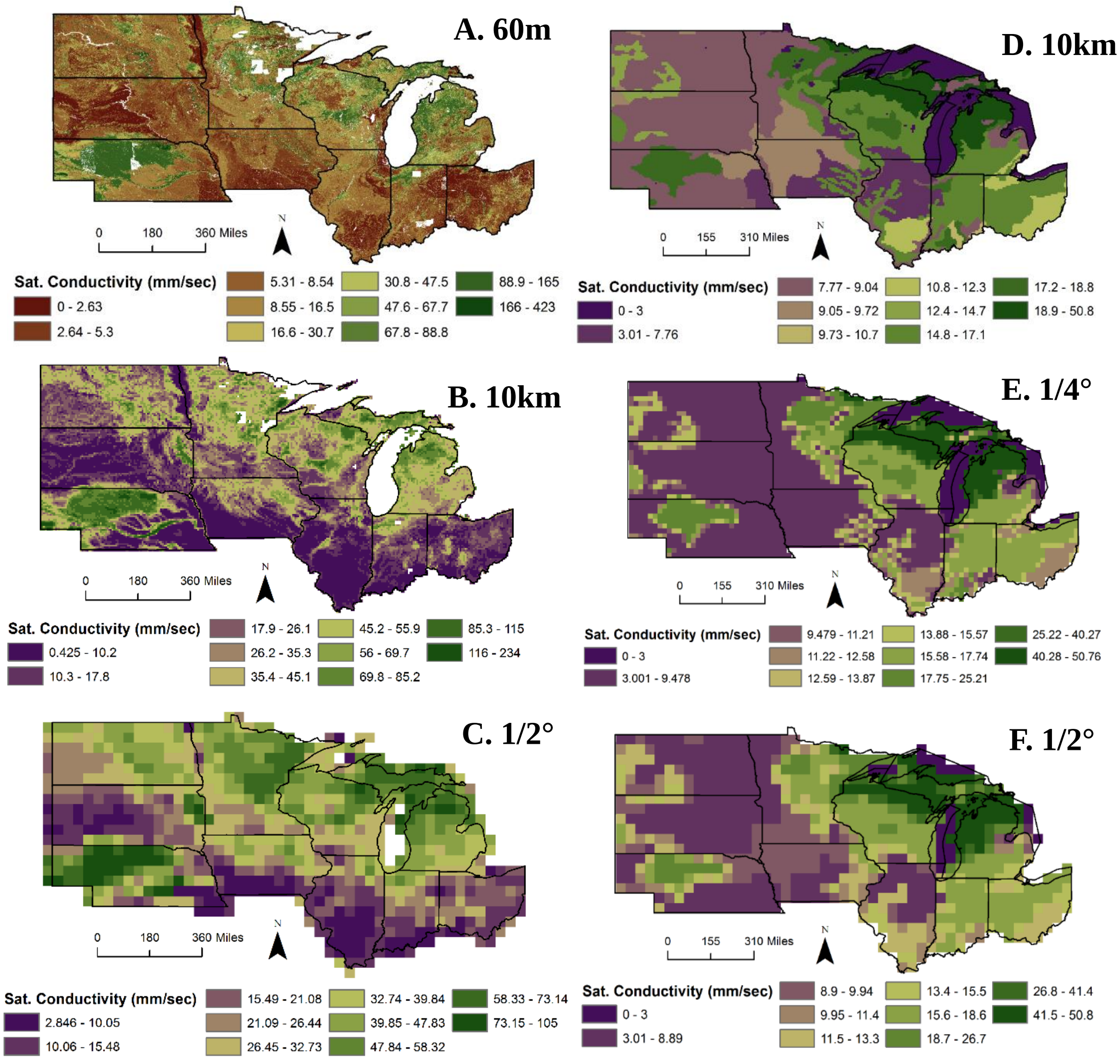

Fig. 3-7. Distribution of saturated conductivity values for SSURGO at a resolution of (A) 60m, (B) 10km and (C) half-degree respectively; WISE at a resolution of (A) 60m, (B) quarter-degree and (C) half-degree respectively.

3.2 Above-ground biomass comparisons for different discretization and averaging scenarios

As mentioned in the previous section, differences in parameter value between SSURGO and WISE as well as across the different discretization and averaging schemes, need not necessarily translate into numerically large differences in the biomass estimates produced from an ecosystem model like ED. The answer to this question is not only site-specific, but also depends on the model complexity and algorithms used to simulate water flow through the soil layers.

For SSURGO, when comparing different discretization schemes, averaged at a resolution of half-degree, the taylor diagram (fig. 3-8, saturated conductivity values mapped in fig. 3-7) shows that there is no perceptible difference in the results obtained from the different discretization schemes. I.e, using just 2 values of saturated conductivity to represent the entire U.S. Midwest at a half degree resolution works just as well for computing above-ground biomass as using 100 distinct values (or 1 million!).

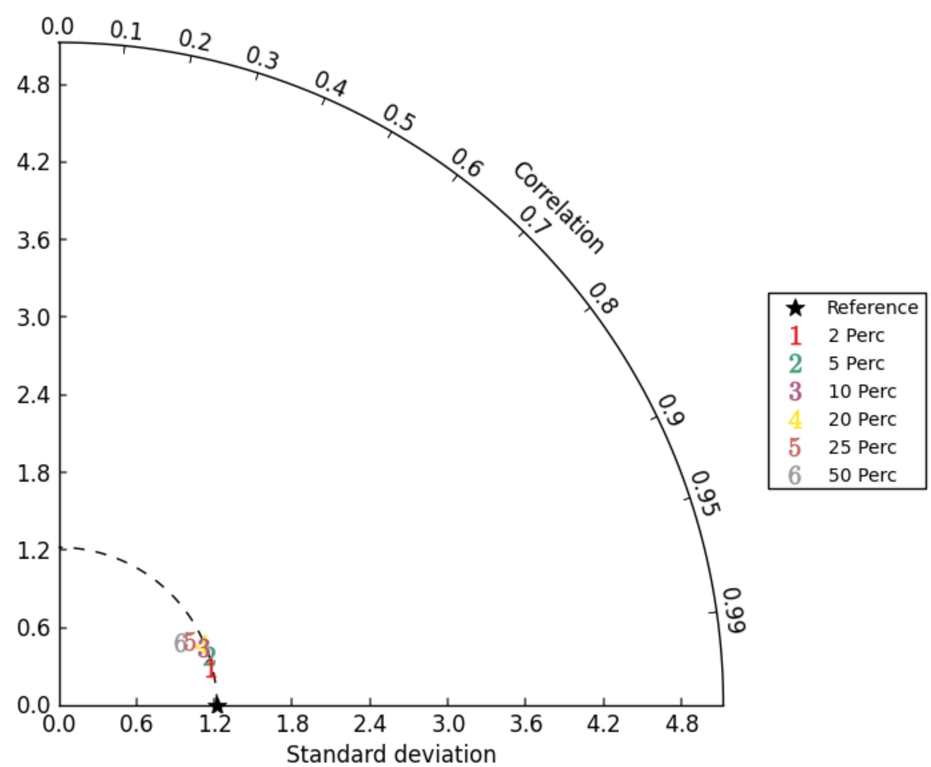

Fig. 3-8. Taylor diagram for above-ground biomass estimates from discretized saturated conductivity values for SSURGO.

In order to gain a comprehensive view of the performance of ED when using data from SSURGO and WISE, we also consider the minimum and maximum normalized error (error divided by mean) for several scenarios. While the correlation coefficient gives us an idea of how well two quantities vary against each other, we are also interested in the magnitude of difference between two measures of above-ground biomass obtained from two different soil datasets (fig. 3-9).

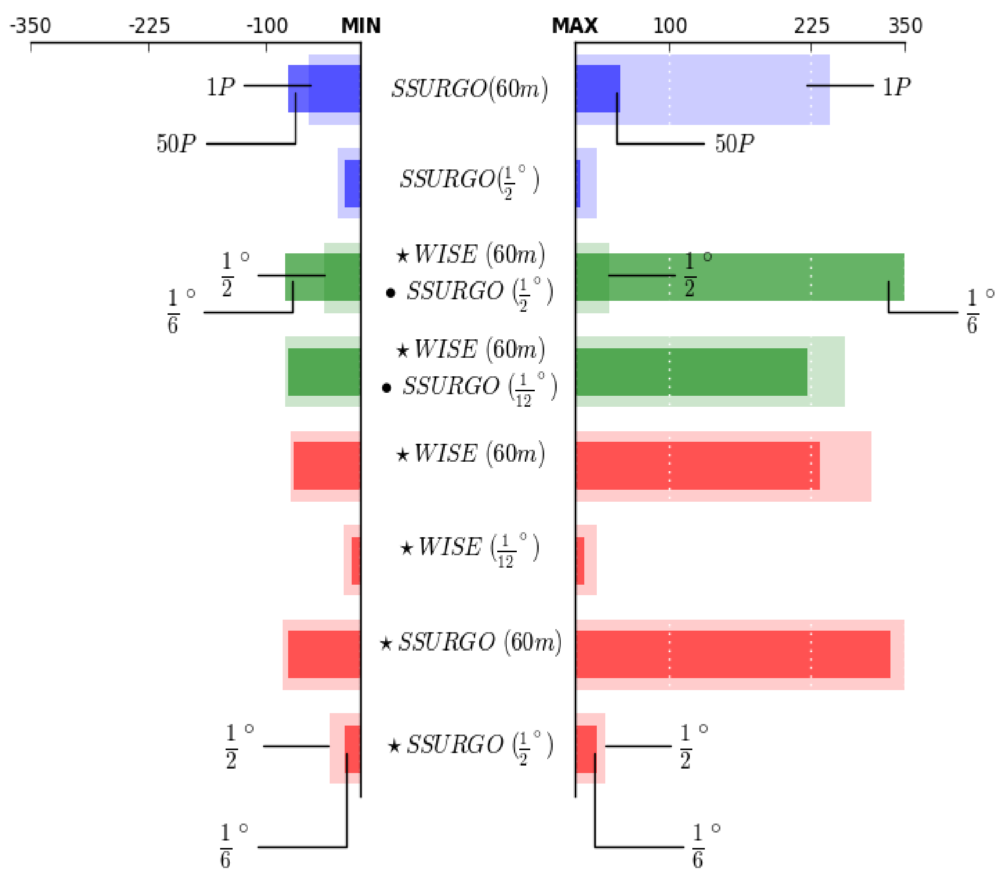

Fig. 3-9. Above-ground biomass comparisons between SSURGO and WISE for varying averaging and discretization schemes.

The left-side panel of fig. 3-9 shows the most negative difference between above-ground biomass values obtained when computing the normalized error, and the right-side panel shows the most positive differences. The red colored bars show the corresponding differences for SSURGO and WISE datasets, for their native resolutions as well as half-degree for SSURGO (maps in fig. 3-10) and 60m for WISE. Clearly, comparing estimates at high-resolution (60m) tends to produce the highest magnitude errors. The green bars represent differences between WISE and SSURGO, with errors increasing when the resolution is lowered from 1/2° to 1/6°. Finally, the blue bars correspond to the discretization scheme. Again, a higher resolution comparison (60m) provides much greater normalized errors compared to one done at half-degree.

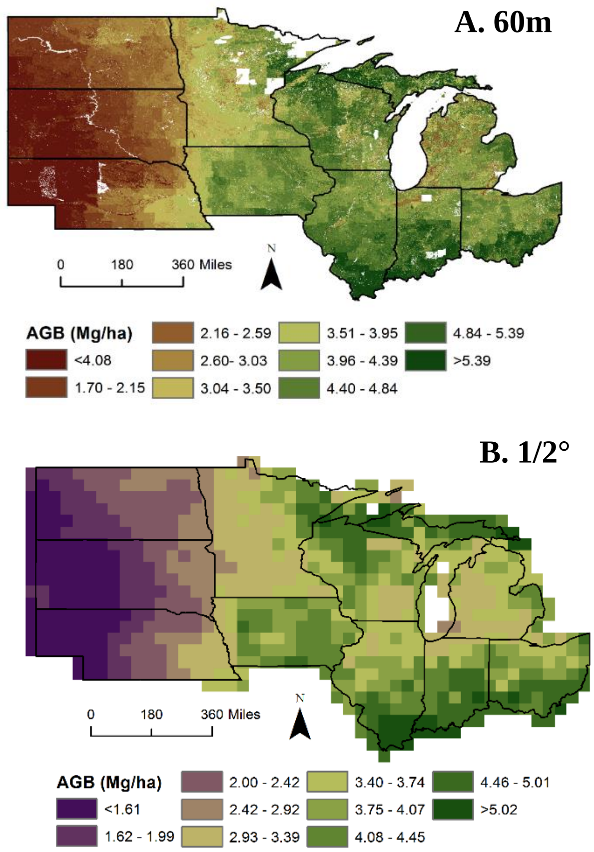

Fig. 3-10. Distribution of above-ground biomass values for SSURGO at a resolution of (A) 60m, (B) half-degree respectively.

4. *Conclusion: What approach to use and when?*

Traditionally, soil datasets have been averaged to match the coarser resolution climate dataset, both to save computational time and also because models have ignored sub-grid scale edaphic variations. With more modeling studies looking at producing high-resolution estimates of biomass and carbon fluxes, increasingly there is a need to use soil datasets in their entirety. We introduce a novel yet simple (discretization) scheme to address this computational burden without reducing model accuracy. We compare the discretization approach to the commonly used averaging scheme and find that,

1. When averaged to half-degree, saturated conductivity estimates from SSURGO and WISE datasets are similar over the U.S. Midwest. The differences at a 60m resolution are much greater, with essentially no correlation between the two.
2. Similar to the raw saturated conductivity values, above-ground biomass estimates for SSURGO and WISE are similar at half-degree resolution but we see large errors when compared at 60m resolution.
3. When discretized, distribution of SSURGO saturated conductivity values depends greatly on the sampling frequency.
4. Estimates of above-ground biomass in the discretization scheme are resolution dependent. At half-degree resolution, a coarsely discretized SSURGO dataset provides the same results as a finely discretized dataset. At 60m resolution, it is better to use the finely discretized dataset.

Chapter 4: Location and Management Interaction Impacts on Yield Gap Analysis of Cellulosic Ethanol production in U.S. Western Corn Belt

1. *Introduction*

1.1 Background

Demand for biofuel feedstocks in the U.S. has increased in response to the reduction in fossil fuels and foreign oil mandated by the Energy Independence and Security Act (EISA) of 2007. Currently, most of the biofuel demand is met by ethanol derived from grains, mostly corn (*Zea mays*). However, the Renewable Fuel Standard (RFS2) requires an annual increase in the production of biofuels from cellulosic biomass (hereafter referred to as second generation or 2G), reaching at least 16 billion gallons of production by 2022 (USEPA 2010). This has incentivized academic and industry research in utilizing 2G feedstocks to produce biofuels.

When managed properly, 2G feedstocks offer several environmental benefits over grain-based systems, including reduced fertilizer and energy inputs (Manatt et al., 2011), greater soil carbon sequestration (Gelfand et al., 2011), reduced greenhouse-gas (GHG) emissions (Gelfand et al., 2013), higher rates of energy return (Robertson et al., 2011) and positive impacts on biodiversity (Meehan et al., 2013) and water quality. 2G feedstocks also preclude any competition between food and fuel systems, since they are comprised either of waste or inedible plant material.

1.2 Diversification of 2G feedstocks

With the advancement of cellulosic feedstock conversion processes, further diversification of feedstock options is expected. However, despite their advantages, lack of farmer willingness to invest in growing 2G feedstocks (Jensen et al., 2007) and availability of proven, cost-effective technology to convert plant cellulose to ethanol has hampered widespread cultivation till now. While this is likely to change in the near future given recent breakthroughs in producing lignin that is more amenable to degradation and conversion to biofuels (Wilkerson et al., 2014), it is critical to consider a portfolio of 2G feedstocks as no single crop will satisfy all requirements in every agroecosystem. This is partly because biofuel feedstock productivity, like other crops, varies widely based on management inputs and climatic and edaphic conditions (Simmons et al., 2008).

The nature and availability of agroecosytems is evolving too. New agroecosystems, in the form of underutilized but arable marginal lands could be cultivated, in order to avoid competition with land currently devoted to growing food crops and incur minimal carbon debt (Fargione et al., 2008). Of the available 2G feedstocks, perennial herbaceous energy crops are expected to result in lower GHG emissions that conventional corn-grain production because they require fewer nutrient and pesticide inputs, and less intensive tillage practices. Additionally, diverse mixes of perennial crops are also expected to boost cellulosic ethanol production. Low-input, high-diversity (LIHD) grasslands can have energy yields twice that of corn grain ethanol, along with lesser GHG emissions and lower need for management inputs (Tilman et al., 2006). The higher yields in the LIHD scenarios can be explained on the basis of more efficient utilization of limited resources. By further reducing the need for excessive fertilization, enhancing pest control, reducing soil erosion and improving water quality, diverse mixes of perennial crops show immense promise as biofuels of the future. The debate in bioenergy is no longer ‘are biofuels beneficial?’, rather the key question is now to identify which feedstocks should be grown where to maximize yields without introducing additional carbon costs from the displacement of food and feed production (Davis et al., 2011).

1.3 Identifying suitable location and management options for 2G feedstocks

The search for beneficial biofuels should focus on the twin objectives of sustainable biofuel feedstocks that do not compete with food crops and do not induce either direct or indirect land-use change. To minimize GHG emissions from land-use change, we need to identify lands that are capable of producing abundant biomass with limited management inputs, while initially poor in carbon storage in soil and vegetation (Tilman et al., 2009). We also need to close the distance between actual 2G feedstock yields and potential yields: the yield gap (Cassman, 1999, Neumann et al., 2010, Tilman et al., 2011 and Ittersum et al., 2013). However, to the best of our knowledge, no information is available at a regional scale on the spatial distribution of yield gaps of 2G feedstocks and the potential for management interventions, in the form of fertilization and/or irrigation to close that gap.

The present study uses an existing dataset comprised of modeled yields of successional herbaceous vegetation on marginal lands in the U.S. Midwest (Gelfand et al., 2013), and extends it for a factorial combination of fertilization and irrigation inputs for each available marginal land site. The range of possible yields thus obtained for each site are then used to achieve the following objectives (1) Determine through stepwise multiple linear regression, the most important climatic and edaphic factors affecting yields; (2) Estimate the potential yield for each marginal land site based on the maximum yield attained by a site with similar climatic and edaphic conditions; (3) Apply adequate fertilization and/or irrigation to help attain the potential yield; (4) Examine how the management input affects other ecosystem variables like soil organic carbon, nutrient and water stress and surface runoff. We expect this analysis to give a high spatial resolution estimate of the environmental impact of cultivating 2G feedstocks in U.S. Midwest and the management interventions needed to sustainable achieve the maximum yield potential of the region.

2. *Method*

2.1 Study area

Our study area comprising the WCB states of North Dakota (ND), South Dakota (SD), Nebraska (NE), Minnesota (MN) and Iowa (IA), (Fig. 4-1). Annual crop production is the dominant land-use in the WCB, accounting for more than two-fifths of the corn and soybean acreage planted nationally in 2012 (http://quickstats.nass.usda.gov/). These states also overlap the Prairie Pothole Region (PPR), an ecologically sensitive wetland landscape (Johnson et al., 2013). While the original study that this work is based on modeled 5 additional states in the U.S.

Midwest (Ohio, Illinois, Indiana, Wisconsin and Michigan), they were not included in this analysis as the 5 WCB states account for over 95% of the around 11 million ha of marginal land in the region.

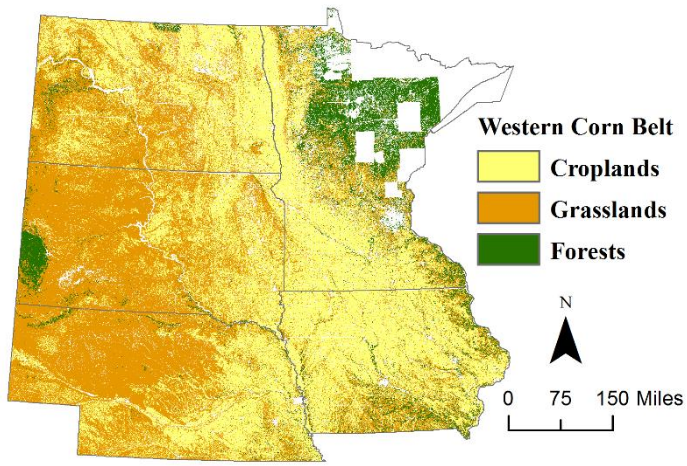

Fig. 4-1. Land-cover distribution in the Western Corn Belt

2.2 Model

We used the EPIC model (Izaurralde et al., 2006) to simulate successional herbaceous vegetation biomass yields on regional scales (Fig. 4-2a). In doing so, we used the same model parameterization as developed, calibrated and validated on site level data in Gelfand et al., 2013. EPIC is a comprehensive, field-scale, biophysical process based model that can simulate the plant yield and net primary productivity of over 100 plant species including grasses, food and specialty crops. It is based on the concept of radiation-use efficiency and growing degree days. The plant canopy intercepts a portion of photosynthetically active radiation and converts it to plant biomass, with the phenology being dependent on heat unit calculations. Potential gains in plant biomass are proportionately decreased by the most severe stress factor including vapor pressure deficits, atmospheric CO_2 concentrations and environmental stresses including nutrients, temperature, water, soil strength, aluminum toxicity and aeration (Stockle et al., 1992a, Stockle et al., 1992b). To run the model, daily weather data is needed, including solar radiation, air temperature, precipitation, wind speed and relative humidity. Other data needs include topographic parameters (slope, length), soil properties (pH, layer depths, C and N contents, bulk density, water storage capacity and texture), and management information (crop rotations, planting and harvesting dates, fertilization, irrigation). For each simulated field, EPIC provides information on the effects of management operations on water quality, soil loss and crop yields. EPIC also reports a comprehensive suite of ecosystem service indicators including C sequestration, N loss in runoff and leaching, denitrification, N fixation, evapotranspiration and runoff and lateral flows (Zhang et al., 2010).

The Spatially Explicit Integrated Modeling framework (SEIMF, Fig. 4-2b), (Zhang et al., 2010) was used to i) create a geodatabase merging all input data sets together and ii) generate input files needed to run EPIC. Applying SEIMF to the WCB resulted in approximately quarter of a million homogeneous units. All the fields which constitute a homogeneous unit, share the same topography, soil type, land use, management and climate. Since the input data sets have different resolutions ranging from 56 m for the land use to 32 km for climate, the resolution of each homogeneous unit is the greatest common denominator i.e. 56 m.

2.3 Model data inputs

The list of model data inputs and their brief description is listed below for reference. They are exactly the same as used in Gelfand et al., 2013.

Soils

A conterminous-scale soil property database was built using soil survey geographic (SSURGO) data from the U.S. Department of Agriculture Geospatial Data Gateway (http://datagateway.nrcs.usda.gov). Extracted soil properties included number of soil layers, depth of each layer, pH, bulk density, slope gradient and length, soil texture information and % organic carbon and total nitrogen. The land capability class (LCC) variable was extracted to describe marginal lands on the basis of use limitation such as erosion risk, soil depth, wetness and slope (Klingebiel and Montgomery, 1961). There are eight LCCs, ranging from class I (no limitations for agricultural use) to class VIII (severe limitations for agricultural use). Cropland agriculture is supported on classes I–IV, whereas classes V–VIII contain non-arable land.

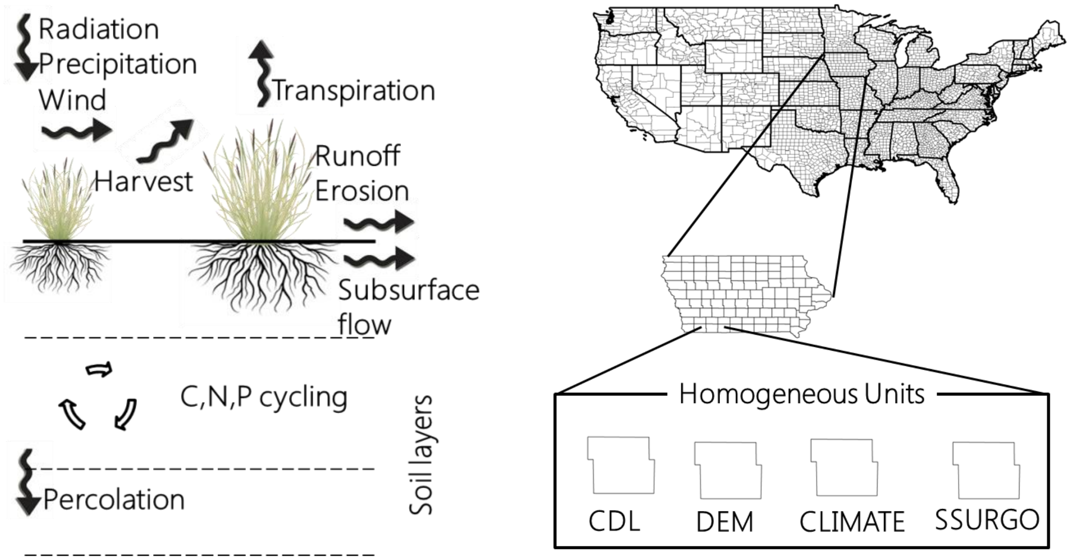

Fig. 4-2. (a) Conceptual framework of EPIC model (b) SEIMF framework

Topography

Topography was defined using data from NASA's Shuttle Radar Topography Mission, which produced a digital elevation model for the region at a resolution of 30 m (Farr et al., 2007). Slope length and gradient for each spatial modeling unit was derived through geospatial analysis of this data.

Climate Data

EPIC requires daily weather information including daily temperature (maximum and minimum), precipitation, solar radiation, wind speed and relative humidity. Daily weather files at 32-km resolution were derived from the North America Regional Reanalysis database (http://www.esrl.noaa.gov/psd/data/gridded/data.narr.html) (Messinger et al., 2006).

Successional herbaceous vegetation

Following Gelfand et al., 2013, the list of EPIC simulated successional herbaceous vegetation on each marginal land site included: Red clover (*Trifolium pratense*), Timothy (*Phleum pratense*) and Poa (*Poa annua*). Of these, Red clover is a legume and does not need fertilization. In EPIC runs where fertilization was applied, it was divided equally amongst the other two grasses. The crop parameterization, planting density and sensitivity analysis of these grasses follows Gelfand et al., 2013.

2.4 Statistical analysis of unfertilized yields

Simple and multiple linear regression analysis were used to develop equations that explain the variation in the unfertilized yields across a range of climatic and edaphic factors. The criterion for model selection was r^2 and stepwise methods were employed to eliminate some of the independent variables. The independent variables in the analysis and their associated units are listed in table 4-1. Co-linearity of the independent variables was checked with the variance inflation factor value and the significance of the models was tested with the F value. To be included in the final model, variables had to be significant at P=0.05 and the complete model at P=0.001.

2.5 Factorial EPIC simulations

Factorial combinations of fertilization and irrigation management were used to run EPIC. These included five fertilization intensities from 0 kgN $ha^{-1}yr^{-1}$ (unfertilized) upto 123 kgN $ha^{-1}yr^{-1}$, with intermediate levels of 33, 68 and 99 kgN $ha^{-1}yr^{-1}$ respectively. For each fertilization intensity, we conducted EPIC simulations with and without irrigation for a total of ten factorial combinations. The unfertilized, no-irrigation scenario represents the baseline case where no management input is applied. While the total biomass yield is expected to be the lowest in this case, it also represent the scenario with least expected negative environmental impact as excessive leaching of nutrients to ground water and erosion cannot happen in absence of fertilization or irrigation inputs. The key is to sustainably boost management inputs in a way such that yields are increased without a disproportionate increase in negative environmental impacts.

2.6 Yield gap analysis methodology

The key independent variables identified during the statistical analysis procedure, and listed in table 4-1 were split into equi-spaced bins. Since the fidelity of bin size is unknown, we used the F-ratio (Kim and Kohout, 1975), to determine the ideal bin size based on the difference in nitrogen stress days and water deficit before and after management input. The basis for this approach is that the correct bin size would give the biggest F-ratio for the before and after management intervention scenarios. Once the bin size was fixed, yield gap zones were identified as a combination of different bins, with each zone representing a unique combination of values for the significant independent variables. The maximum yield in each zone was considered to represent the maximum potential yield attainable under similar conditions. For marginal land sites where the yield fell short of the potential yield thus identified, we determined the minimum fertilization and irrigation level needed to attain the potential yield. Ecosystem variables of interest (soil organic carbon, surface runoff, water deficit etc.) were extracted to estimate change before and after management intervention.

2.7 Impact of inclusion of a legume in the native successional grass mix

Inclusion of a legume (red clover) in the native successional grass mix simulated by EPIC should reduce the amount of fertilization needed to reach a specific potential yield because of the atmospheric N fixed by that legume and made available to the other grasses sharing that land (Liu et al., 2010). To understand how the inclusion of a legume affects simulated yields, N addition, soil organic carbon and water stresses in the WCB region, we ran an additional EPIC scenario with the legume excluded from the species mix. We then compared it against the baseline case, but only extracting results for the non-leguminous timothy and poa, so that we only compared yield estimates for two grasses in the species mix in both case. The only difference being that in the baseline case, the N fixation induced by the legume should boost simulated yields, but also potentially reduce yields because of increased resource competition. The interplay of N fixation and increased resource competition would reflect in regional variations in the yield benefit conferred by including a legume in the species mix.

3. *Results and discussion*

3.1 Factorial combination of fertilization and irrigation impacts

Before performing a yield gap analysis, it is important to observe the variation in simulated yields across a gradient of management inputs. In this study, we vary fertilization across five discrete levels, both with and without irrigation (fig. 4-3). Three of the five fertilization levels (0, 68 and 123 kgN ha^{-1} yr^{-1}), were used in Gelfand et al., 2013, and the remaining two (33 and 99 kgN ha^{-1} yr^{-1}), were added for this study in order to discretize N input levels further, making the yield gap calculations more precise. As expected, increased N input reduces the level of N stress experienced by a system (fig. 4-3A), as reflected in the decrease in the number of days for which the most stress is caused due to nutrient limitation. As defined in EPIC, any stress (nutrient, water etc.) is reflective of the supply-demand imbalance of that resource in that site. Since, EPIC considers the largest stress in reducing the simulated yield, additional irrigation reduces water stress (fig. 4-3B), but concomitantly increases the number of N stress days (unfilled bars in fig. 4-3A). This does not imply that the nutrient stress has increased, rather, due to reduction in water stress, the dominant stressor is no longer moisture availability.

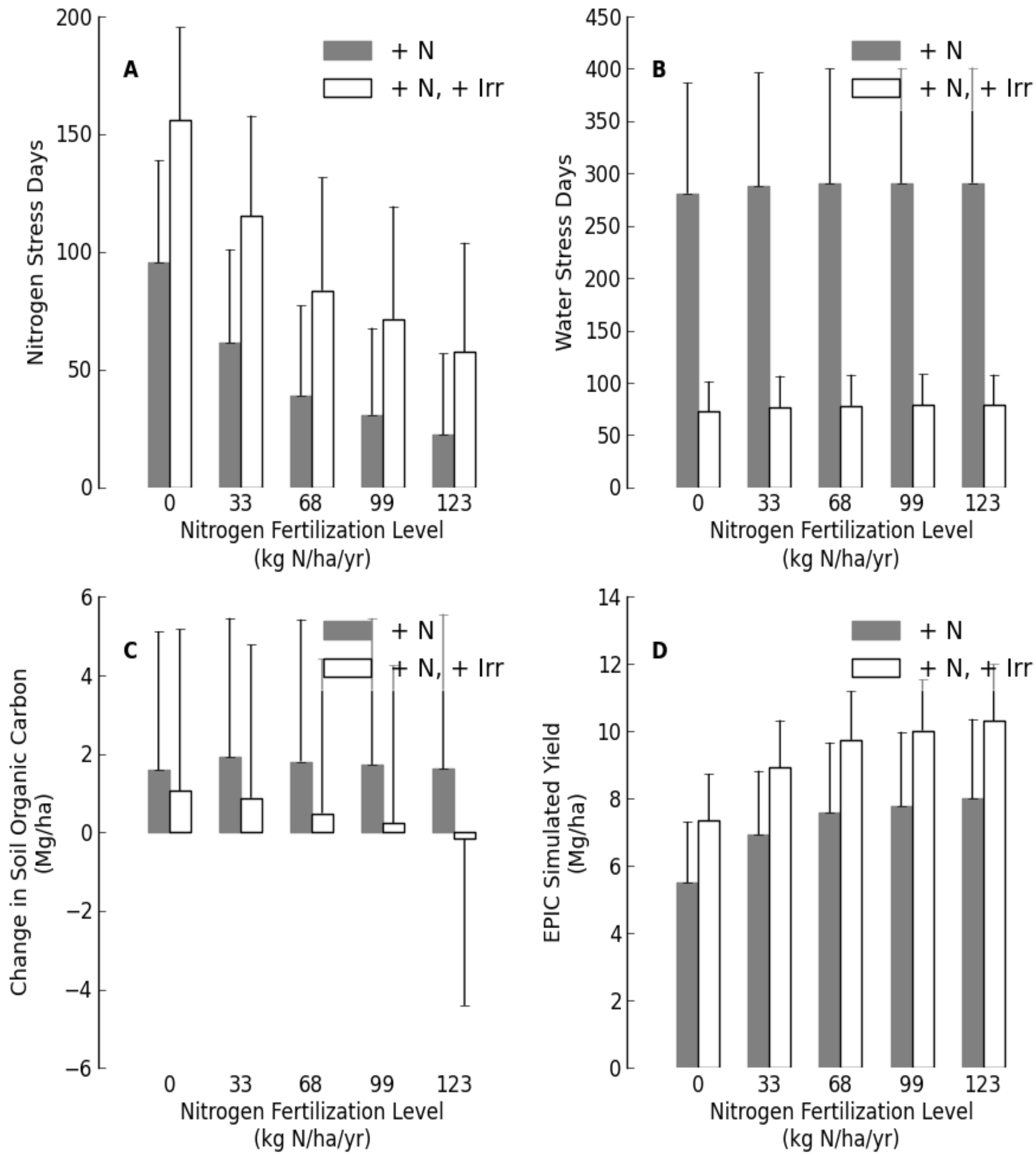

Fig. 4-3. Fertilization x Irrigation impacts on (A) nitrogen stress days; (B) water stress days (C) change in soil organic carbon; (D) EPIC simulated yield.

Notably, additional N inputs do not appreciable increase soil organic carbon levels (fig. 4-3C), with the gain in soil organic carbon stocks for an eight-year period from 2000 – 2008, hovering around 2 ± 2.5 MgC/ha. This could be attributed to the lack of clay in the marginal land soils which reduces organic C losses (Burke et al., 1989). While irrigation provides a reliable yield boost of around 2 Mg/ha, for all fertilization intensity levels; it cannot be a universal solution since it also depletes soil organic C levels due to excessive leaching (Brye et al., 2001).

3.2 Regression analysis

The results from the simple and multiple stepwise linear regression analysis conducted to regress baseline, unmanaged EPIC simulated yields against climatic and edaphic variables are listed in table 4-1. Based on their r^2 value, the following variables were determined to be significant in explaining yield variation for the baseline scenario: mean annual precipitation, pH, LCC and soil organic carbon. Notably, mean annual temperature was not found to be a significant predictor of baseline yield variation. This could be because the inter-regional variation in mean annual temperature is not high since the WCB constitutes a single eco-region (McNab et al., 2005). Additionally, the cumulative daily degree days parameterized for each of the three grasses are similar (~1300), thereby removing a factor whose interaction with mean annual temperature across the region could significantly influence yields. Precipitation does decrease moving from east to west with an arbitrary but commonly accepted transition to an arid environment west of the 100th meridian (Wright and Wimberly, 2013b). Other three factors significantly influencing yield variation are all edaphic. Of these, soil organic carbon has a well-known positive influence on estimated yields (Izaurralde et al. 2006, 2007; Causarano et al. 2007, 2008 and Apezteguía et al. 2009). As currently parameterized in EPIC, low tolerance of the three grasses to aluminium saturation induces a yield sensitivity to soil pH (Gaiser et al., 2010). Finally, the interpretive factor, LCC describes land use classes based on use limitation such as erosion, risk, soil depth, wetness and slope, with the limitation increasing as we advance from LCC class I to LCC class VIII, thus intuitively being an important factor influencing yield variation (Kang et al., 2013). Scatter plots with hexagon binning to represent density are plotted for each of the significant (fig. 4-4) and non-significant (fig. 4-5) explanatory variables.

Table 4-1. Exploring the effect of climate and soil variability on baseline native successional grass mix yields as simulated by EPIC. The 95% confidence interval (CI) and the coefficient of determination (r^2).

Variable	95% CI	r^2
Climate		
Mean annual precipitation, mm	327.83 – 1025.0	0.35***
Mean annual temperature, °C	5.50 – 12.91	0.02***
Soil		
pH	5.43 – 8.39	0.18***
Soil organic carbon, Mg/ha	0.16 – 2.47	0.12***
Bulk density, T/m^3	1.18 – 1.71	0.01***
Saturated conductivity, mm/sec	1.68 – 1009.21	0.02***
Available Water Capacity, m/m	2.29 – 15.45	0.02***

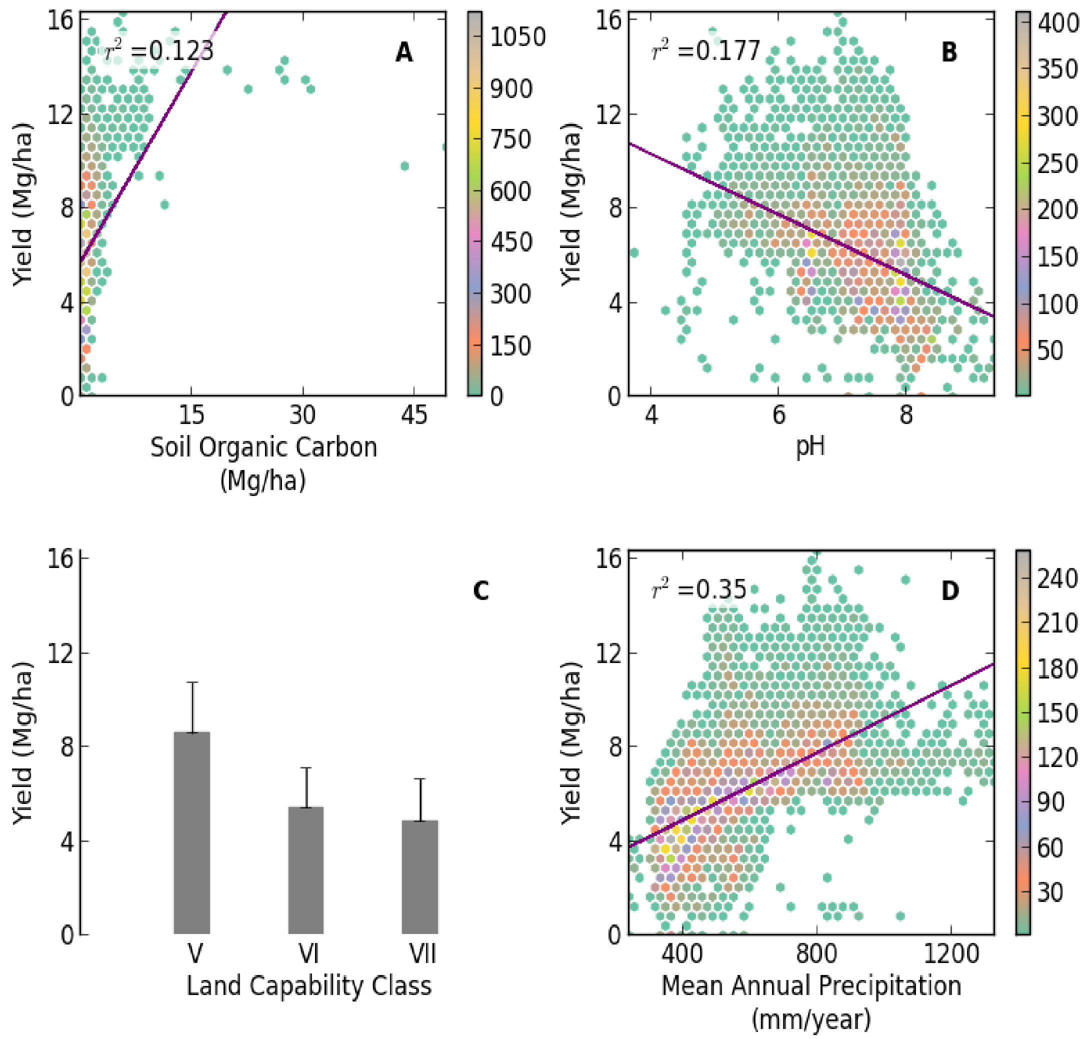

Fig. 4-4. EPIC simulated yield variation for the baseline (unmanaged) scenario for marginal lands in WCB. The four explanatory variables for which the yield variation is plotted are: (A) soil organic carbon; (B) pH; (C) land capability class and (D) mean annual precipitation.

Not all yield gap analysis studies conduct a regression analysis beforehand, instead using variables based on an intuitive understanding of the region (Licker et al., 2010). As a result, the commonly used variables include mean annual temperature, mean annual precipitation and a metric of water deficit based on the ratio of actual to potential evapotranspiration (Prentice et al., 1992). Yield gap results are highly sensitive to our choice of environmental variables and a rigorous analysis can help inform results that are likely to remain constant when transferred from a computer model to the real world (Lobell et al., 2009).

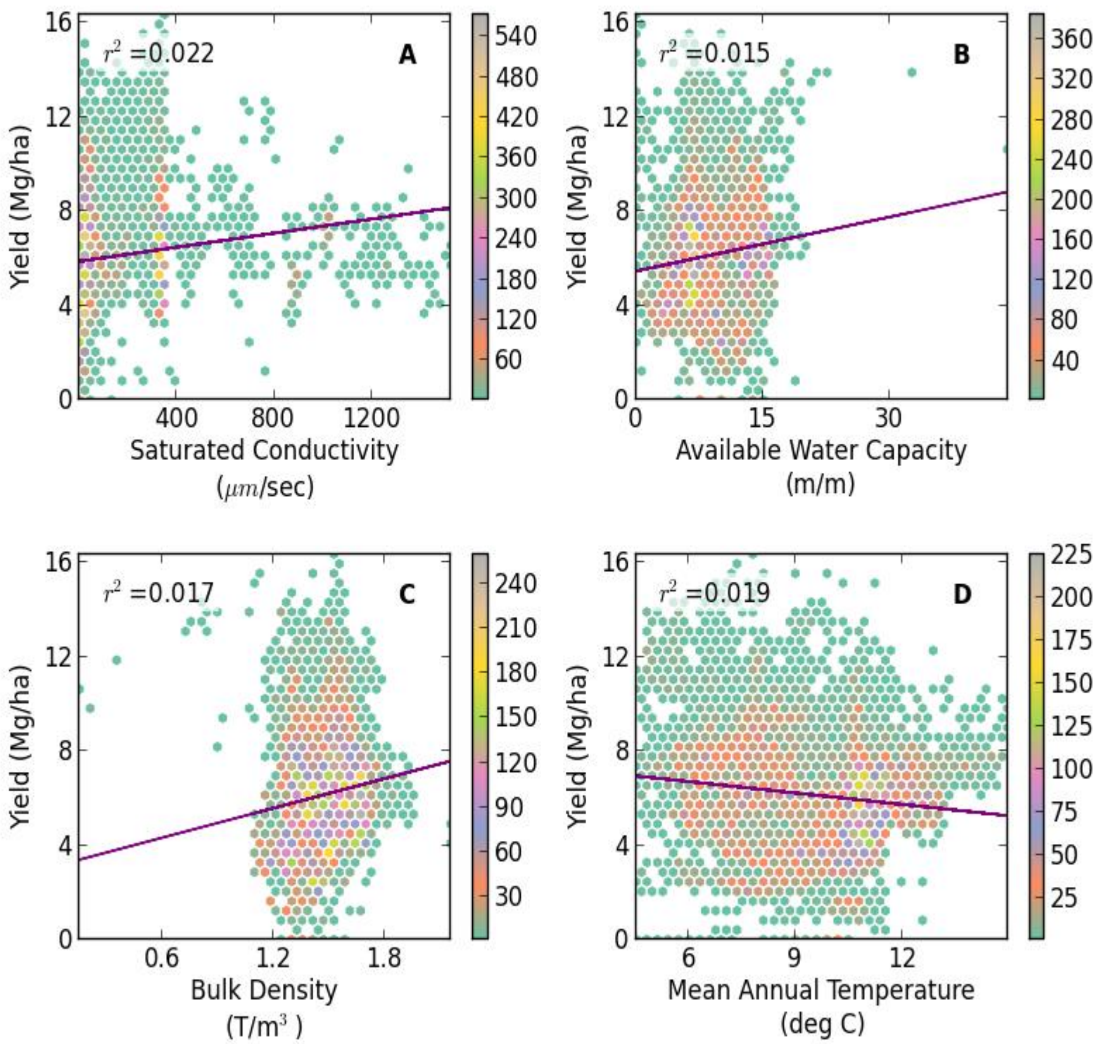

Fig. 4-5. EPIC simulated yield variation for the baseline (unmanaged) scenario for marginal lands in WCB. The four explanatory variables for which the yield variation

is plotted are: (A) saturated conductivity; (B) available water capacity; (C) bulk density and (D) mean annual temperature.

3.3 Determining bin size for yield gap analysis

Bin size or the numerical range of each unique bin can dramatically influence our estimates of yield potential. At one extreme, if the bin size is small enough such that each marginal land site occupies a unique environmental niche, we will find that all sites are already attaining their potential yield. This would mean that no management inputs are needed. On the other hand, if the bin size is big enough to include the entire study region in one bin, almost all sites will be below potential and require immense amounts of nutrient and water inputs to attain yields similar to a few highly productive sites. In this case, the potential yield might be so high so as to be considered an outlier. In this analysis, we adopt a simple yet novel approach to identify an appropriate bin size. Our reasoning is that an appropriate bin size for the system would automatically partition the sites into two groups which are quantitatively different in key ecosystem parameters, when compared before and after management inputs have been applied. For the present study, we use the F ratio (Hartley et al., 1950), which describes the ratio of variance between groups to the variance within groups. The first group in this case is the set of marginal land sites that were not provided with any new management inputs and the second group consists of the sites that were. We then compute F ratio for the number of nitrogen stress days and water deficit for these groups. Subsequently, we find the bin size which maximizes a Cobb-Douglas production function (Cobb and Douglas, 1928) with these two ecosystem variables. In our present case, the production function is maximized when each of the significant explanatory variables affecting unmanaged yields (soil organic carbon, mean annual precipitation and pH) is divided into 14 equi-spaced bins (fig. 4-6). Land capability class being ordinal in nature, is always divided into just three bins (LCC classes V, VI and VII). Overall, we get 8,232 (14^3*3) unique combinations of the four explanatory variables for which we compute the potential yields.

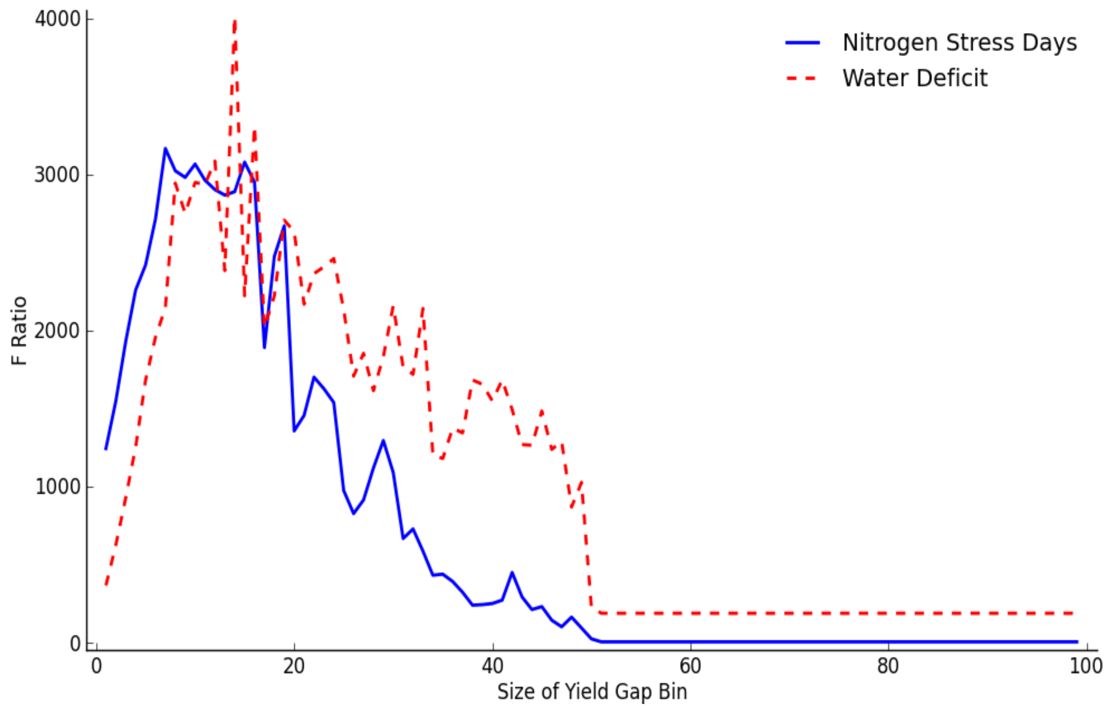

Fig. 4-6. Variation of F ratio for nitrogen stress days and water deficit with changing size of yield gap bin.

For the analysis, we wanted to choose ecosystem metrics that would be affected directly with changes in management. Fertilization affects nitrogen stress days (fig. 4-3A), whereas irrigation affects water stress days (fig. 4-3B). However, since calculation of nitrogen and water stress days is linked in EPIC with a reduction in one automatically causing the other one to increase, we decided to use another metric of moisture availability in the system: water-deficit (Licker et al., 2010).

We conducted a sensitivity analysis on changing the bin size. A larger bin size means that more sites have to be provided with N or water inputs (fig. 4-7A, D), therefore more sites are below their potential yields (fig. 4-7B). Increased management inputs also bring down water and nitrogen stress days (fig. 4-7C).

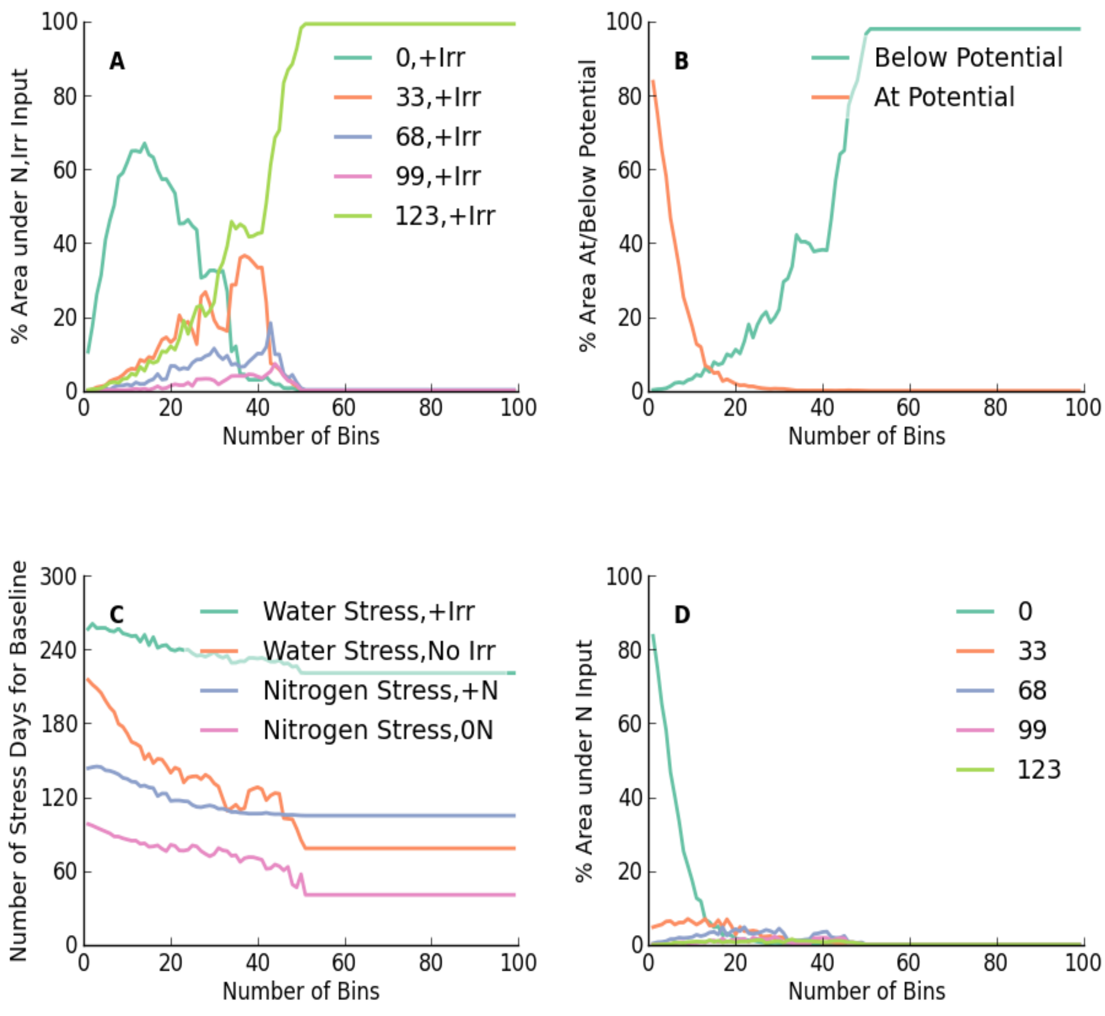

Fig. 4-7. Variation with bin size of (A) Area supplied with both N and water; (B) area above or below potential yield; (C) number of stress days; (D) area supplied with N input.

3.4 Yield gap analysis

Another way of observing where the management inputs are applied is to compare ecosystem metrics for site where no new management was applied versus sites where new management inputs were applied (fig. 4-8). Simulated yield was slightly lower for sites where our algorithm suggests that we apply fertilizer to raise it (fig. 4-8A), this is also reflected in the higher number of N stress days (fig. 4-8C) for these sites. There is no appreciable difference in the nitrogen-use efficiency and nitrogen mineralization estimates for the two sets of sites (fig. 4-8 B, D).

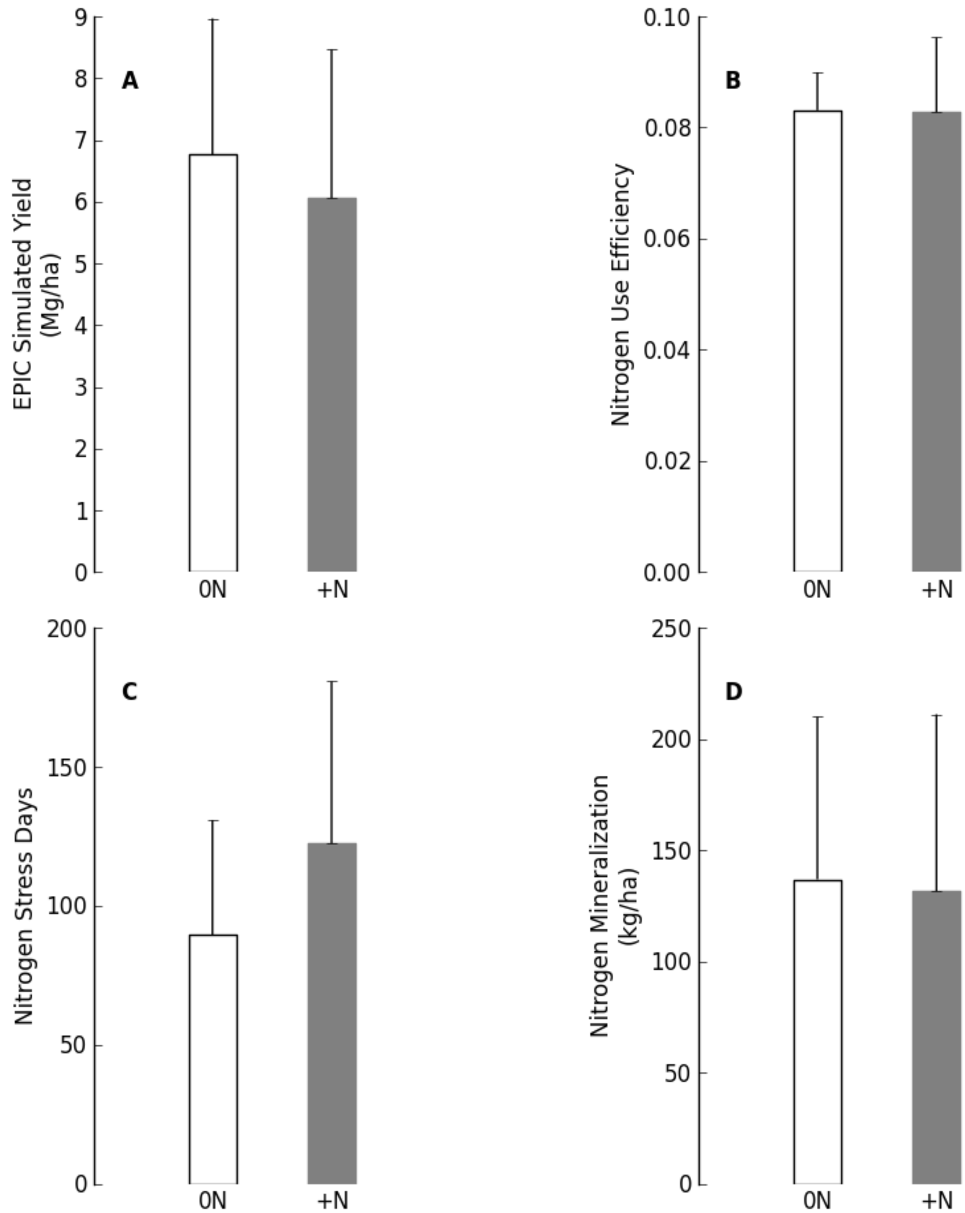

Fig. 4-8. Two groups of sites, one where no fertilization is done (0N) and the other has fertilizer applied (+N). We note their ecosystem metrics before any fertilization.

In the case of irrigation, a similar comparison shows that sites chosen for irrigation tend to have higher water deficit (fig. 4-9A), higher water stress (fig. 4-9B), lower precipitation input (fig. 4-9D) and slightly lower water use efficiency as compared to sites which have not been chosen for any irrigation.

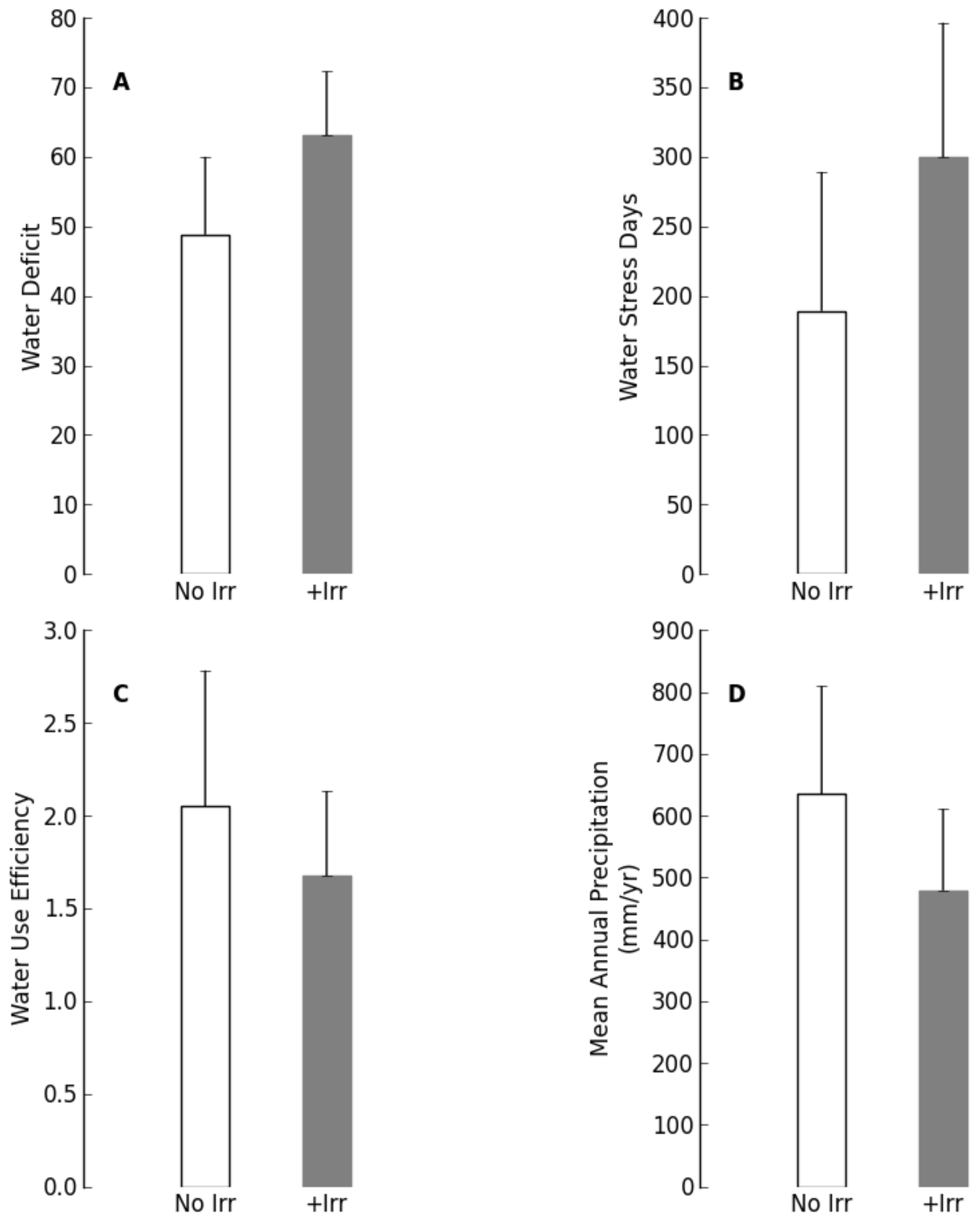

Fig. 4-9. Two groups of sites, one where no irrigation is done (No Irr) and the other has irrigation performed (+Irr). We note their ecosystem metrics before any irrigation.

Sites selected for irrigation also tend to have lower denitrification, N volatilization and surface runoff values (fig. 4-10 A, C and D) than the sites not selected for irrigation. Since water mediates denitrification, this means that the selected sites have a lower potential for loosing N in saturated and moist conditions.

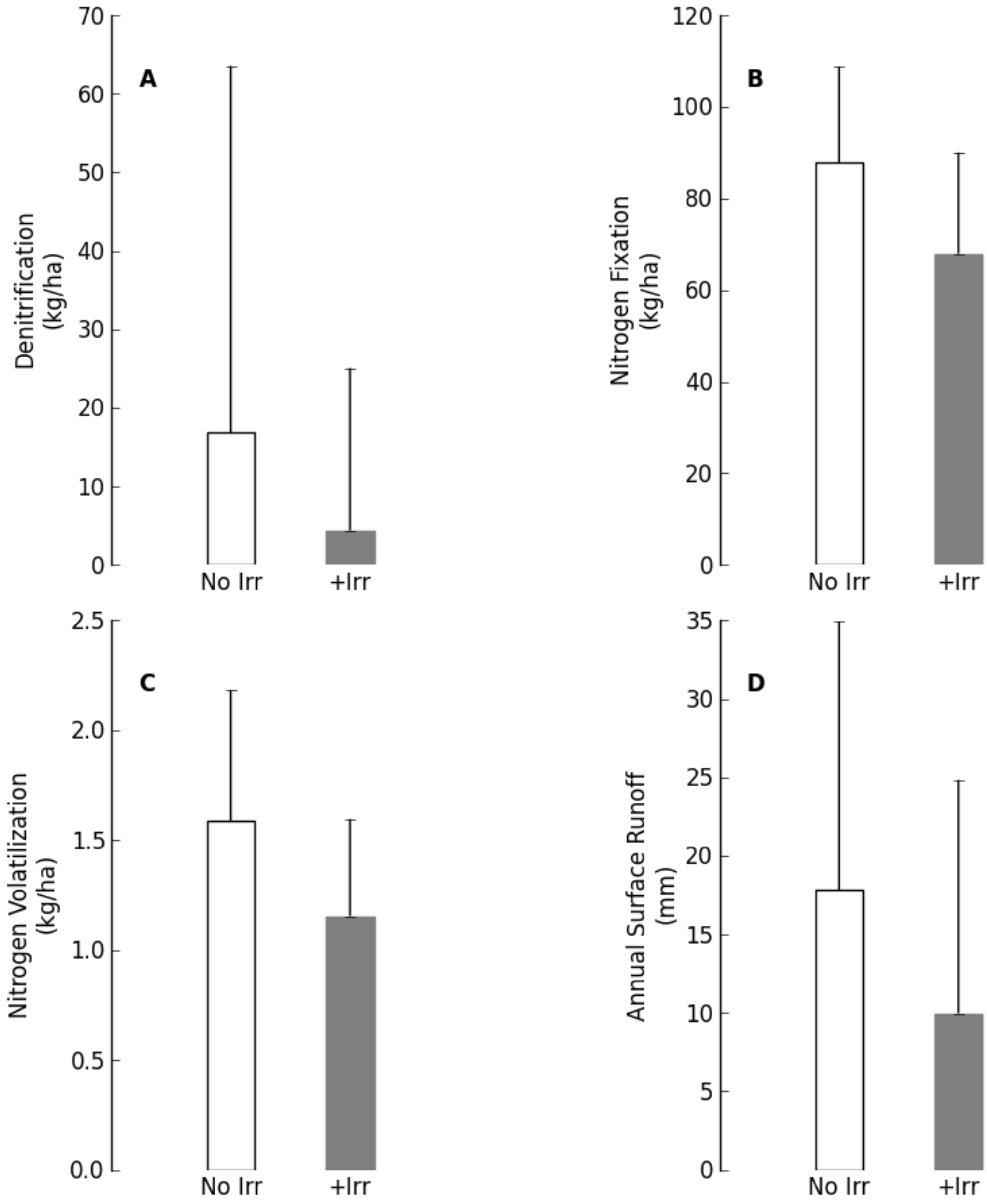

Fig. 4-10. Two groups of sites, one where no irrigation is done (No Irr) and the other has irrigation performed (+Irr). We note their ecosystem metrics before any irrigation.

Soil organic C stock in the marginal land sites decreases due to leaching, when irrigation is performed (fig. 4-3C). What implication does it have for closing the yield gaps? This is especially important in a region where water limitation is the dominant stressor (fig. 4-11), since an increase in yields due to irrigation might be balanced by a loss in soil organic C stocks.

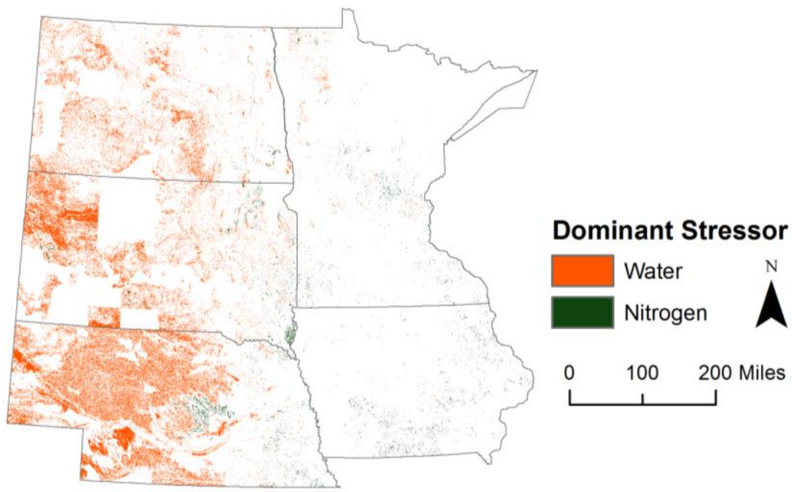

Fig. 4-11. Dominant stressor based on number of stress days estimated by EPIC.

The yield gap approach manages to increase the simulated yields (after management intervention, fig. 4-13D) by ~33%, without a corresponding decrease in soil organic C stocks (fig. 4-13A). This despite a reduction in water stress days by more than 150 and only a 25% increase in annual surface runoff (fig. 4-13C).

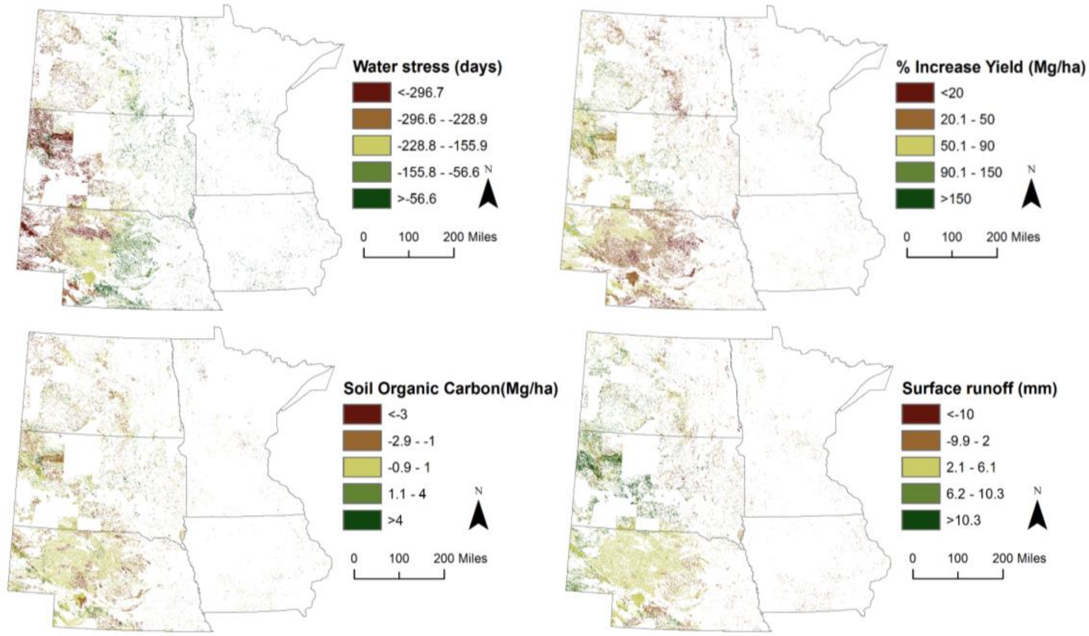

Fig. 4-12. Change in water stress days, yield, soil organic C and surface runoff, before and after management inputs have been applied (N and/or irrigation).

Corresponding to the numerical change estimates in fig. 4-13, their spatial distribution is shown in fig. 4-12. The yield increases are fairly modest (< 20%) in the Sandhills of Nebraska (Eggemeyer et al., 2006), where almost 50% of the ~11 million ha of marginal lands of the WCB are present. This can be attributed to initial low levels of soil C stocks and generally sandy soils in the region

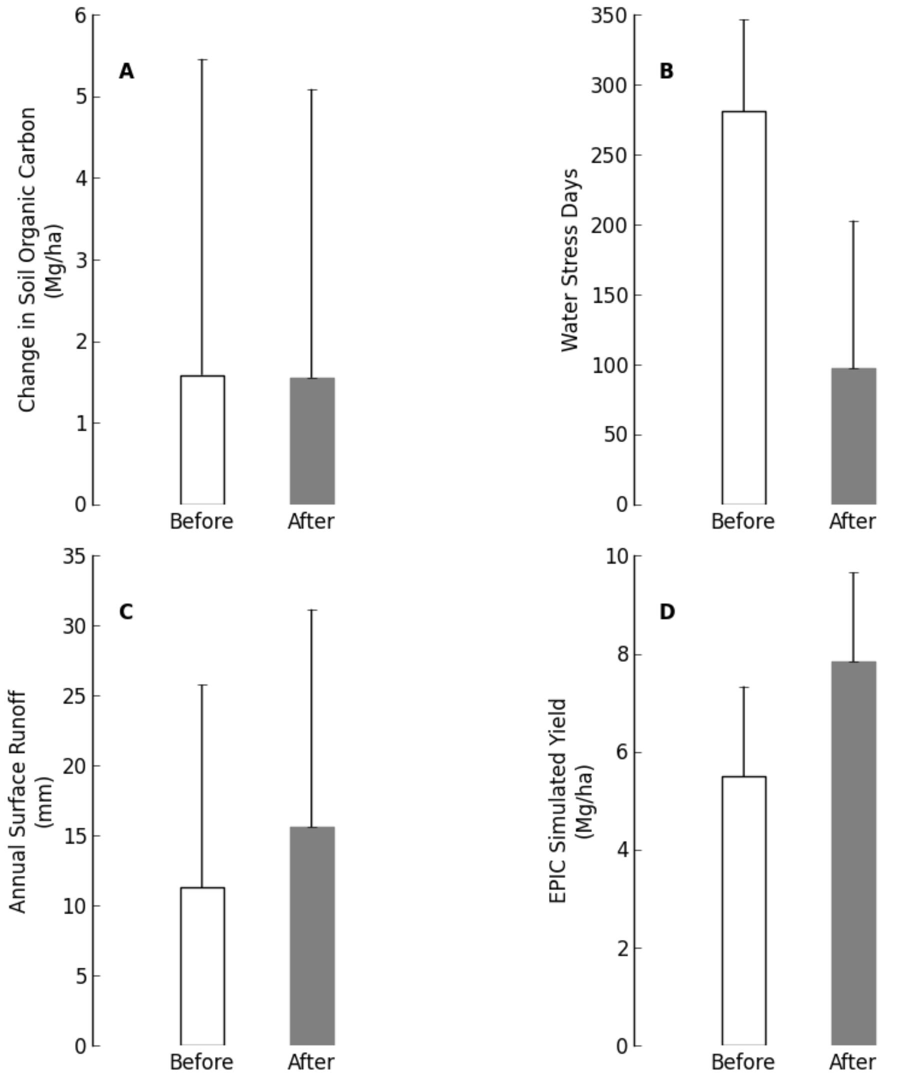

Fig. 4-13. Ecosystem metrics before and after management intervention.

In terms of magnitude, the largest increases in surface runoff can be seen in the sandy soils in western parts of SD, with a corresponding increase in the number of water stress days.

After the yield gap algorithm has been applied at a bin size of 14, we produce a map showing the distribution of management across the region (fig. 4-14). Overall, there is a 45% increase in average yield across the region from 5.49 Mg/ha to 8.0

Mg/ha. 83% of the marginal land acreage has irrigation applied to it, with 13% receiving only fertilization and 5% receiving no management inputs at all.

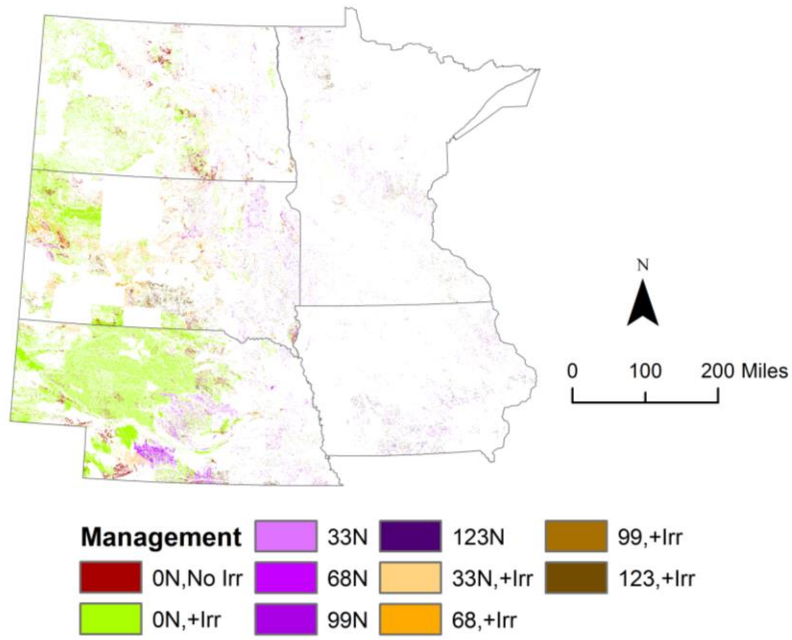

Fig. 4-14. Management distribution after yield gap analysis across the WCB

3.5 Inclusion of legume in native grass mix

As a rule of thumb, presence of around 25% of a legume in a grass mix removes the need for any N input. However, as modeled in EPIC, simulated yields are not enhanced by the presence of a legume (fig. 4-15D), probably because of the competitive pressure for water in an already arid environment. N fixation does help in the sequestration of additional soil organic C, due to the addition of extra residue into the soil (fig. 4-15C).

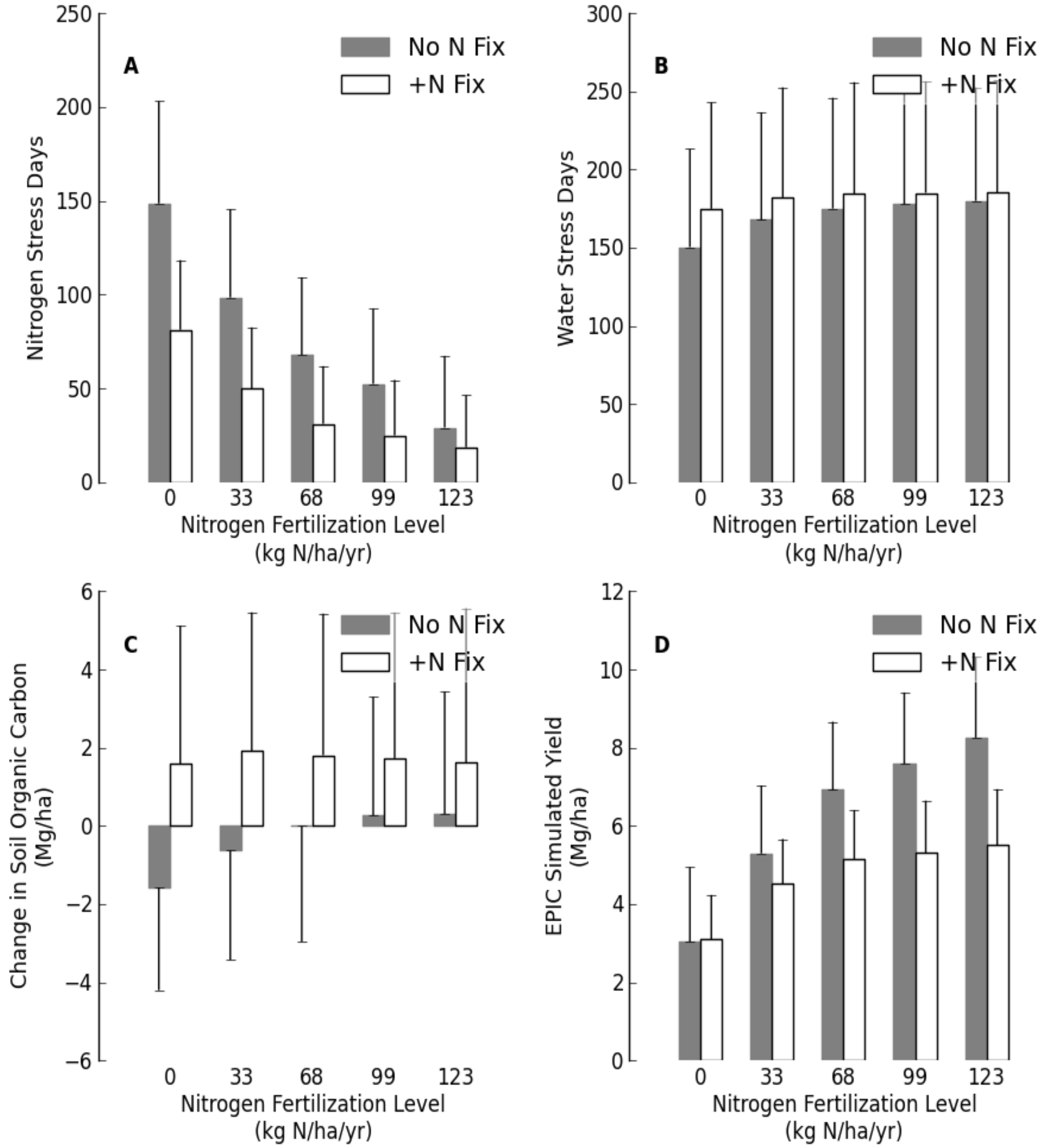

Fig. 4-15. Comparison of ecosystem metrics, with (unfilled bars) and without (filled bars) inclusion of legume.

4. *Conclusions*

In summary, statistical analysis was used to identify the significant climatic and edaphic factors influencing unmanaged yields of native successional grasses on marginal land sites in the WCB. After running EPIC for a factorial combination of management options, yield gap analysis was performed to ascertain for each site, the management intervention best suited to improve yield outcomes in a sustainable manner. More than 80% of the marginal land sites need irrigation, with a much smaller fraction (~13%) requiring N input. We expect this analysis to give a high spatial resolution estimate of the environmental impact of cultivating 2G feedstocks in U.S. Midwest and the management interventions needed to sustainable achieve the maximum yield potential of the region.

Chapter 5: Conclusions

1. *Contextual summary*

While biofuels are widely considered to be a part of the solution to high oil prices, a comprehensive assessment of the environmental sustainability of existing and future biofuel systems is needed to assess their utility in meeting U.S. energy and food needs without exacerbating environmental harm.

The following questions guide this research:

1. What is the spatial extent and composition of agricultural management systems that exist in the U.S. Midwest?

2. How does sub-grid scale edaphic variation impact our estimation of poplar biomass productivity across a gradient of spatial scales in the U.S. Midwest?

3. How do location and management interactions impact yield gap analysis of cellulosic ethanol production in U.S. Midwest?

In the first chapter, I developed an algorithm to identify representative crop rotations in the U.S. Midwest, using remotely sensed data; and used this information to detect pronounced shifts from grassland to monoculture cultivation in the U.S. Midwest. In the second chapter, a new algorithm is developed to reduce the computational burden of high resolution ecosystem modeling of poplar plantations in U.S. Midwest, with the results from the high resolution modeling being used to estimate the impact of averaging and discretization of soil properties on poplar yield estimates. In the third chapter, a novel yield gap analysis of cellulosic feedstocks on marginal lands in the U.S. Midwest is conducted to determine the management inputs needed to reach their yield potential in a sustainable manner.

The significance of this research lies in providing a spatially explicit regional scale analysis of the tradeoffs between food and fuel production and providing an understanding of which biofuel crops should be grown where to maximize production while mitigating environmental damage.

2. *Major findings and contributions*

The research in this dissertation aimed at determining the diversity of biofuel feedstocks that can be sustainable grown in the U.S. Midwest. The second chapter focused on synthesizing the agricultural diversity in the form of crop rotations prevalent in the U.S. Midwest. Crop rotations (the practice of growing crops on the same land in sequential seasons) reside at the core of agronomic management as they can influence key ecosystem services such as crop yields, carbon and nutrient cycling, soil erosion, water quality, pest and disease control. Despite the availability of the Cropland Data Layer (CDL) which provides remotely sensed data on crop type in the U.S. on an annual basis, crop rotation patterns remain poorly mapped due to the lack of tools that allow for consistent and efficient analysis of multi-year CDLs. This chapter presented the **Re**presentative **C**rop **R**otations Using Ed**i**t Dis**t**ance (RECRUIT) algorithm, implemented as a Python software package, to select representative crop rotations by combining and analyzing multi-year CDLs. Using CDLs from 2010 to 2012 for 5 states in the U.S. Midwest, I demonstrated the performance and parameter sensitivity of RECRUIT in selecting representative crop rotations that preserve crop acreage and capture land-use changes. Selecting only 82 representative crop rotations accounted for over 90% of the spatio-temporal variability of the more than 13,000 rotations obtained from combining the multi-year CDLs. Furthermore, the accuracy of the crop rotation product compared favorably with total state-wide planted crop acreage available from agricultural census data. The RECRUIT derived crop rotation product was used to detect land-use conversion from grassland to crop cultivation in a wetland dominated part of the U.S. Midwest. Monoculture corn and monoculture soybean cropping were found to comprise the dominant land-use on the newly cultivated lands.

The third chapter shifted the focus from agricultural lands to second generation biofuel feedstock production from Poplar. To achieve this, I parameterized the Ecosystem Demography (ED) model to predict potential non-nutrient limited yields for hybrid poplar across the U.S. Midwest. Subsequently, I determined quantitatively how estimates of saturated conductivity (a key soil property measuring a soil's ability to transport water) vary for two soil datasets that differ in their spatial resolution and method of estimating soil properties: SSURGO and World Inventory of Soil Emission Potentials (WISE). Subsequently, I examined how saturated conductivity estimates changes based on the resolution at which they are averaged or discretized. Finally, I estimate the impact of averaging and discretization of saturated conductivity on the relative accuracy of poplar yield estimates obtained from ED. The significance of this chapter is in providing a novel method to improve the spatial resolution over which

ED operates by three orders of magnitude (Fig. 5-1). This is crucial because site-specific poplar productivity estimation requires sub-grid scale modeling.

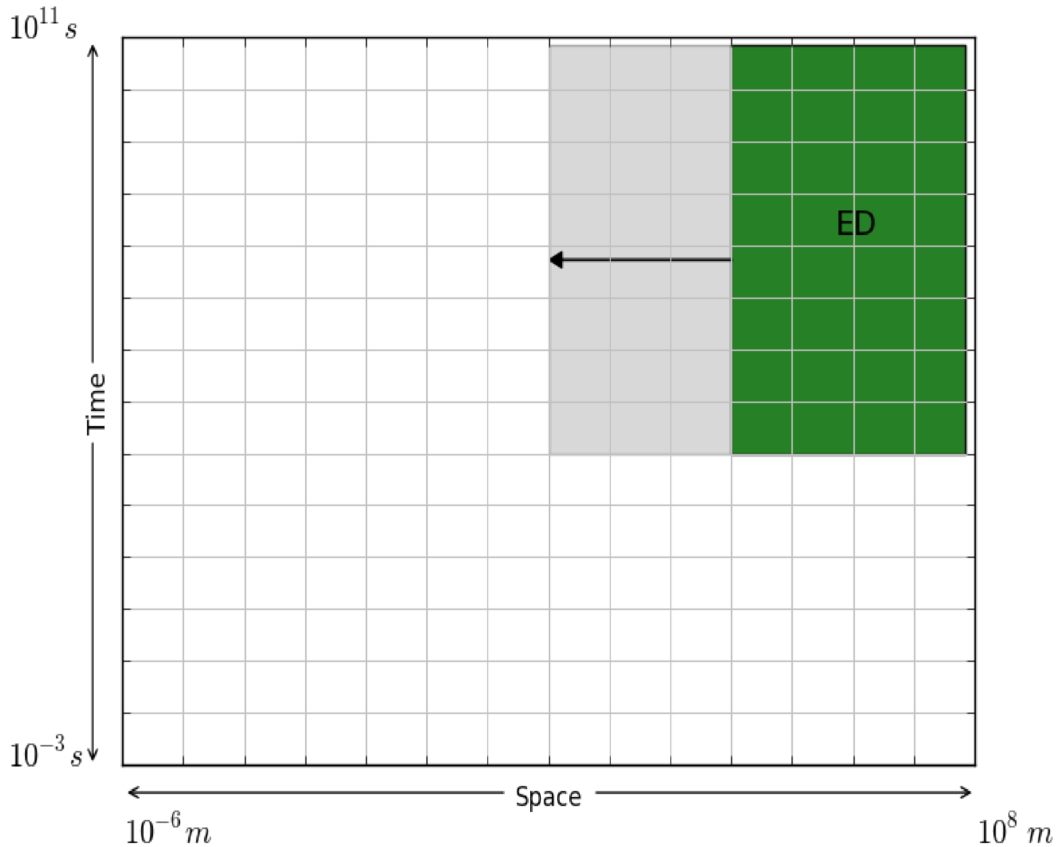

Fig. 5-1. Spatial and temporal extent of ED

The fourth chapter continued the focus on second generation biofuel feedstocks, examining the productivity of native successional grasses on marginal lands in the U.S. Midwest. It also determined the management inputs needed to close the yield gap in the region. This study used an existing dataset comprised of modeled yields of successional herbaceous vegetation on marginal lands in the U.S. Midwest (Gelfand et al., 2013), and extended it for a factorial combination of fertilization and irrigation inputs for each available marginal land site. The range of possible yields thus obtained for each site were then used to achieve the following objectives (1) Determine through stepwise multiple linear regression, the most important climatic and edaphic factors affecting yields; (2) Estimate the potential yield for each marginal land site based on the maximum yield attained by a site with similar climatic and edaphic conditions; (3) Apply adequate fertilization and/or irrigation to help attain the potential yield; (4) Examine how the management input affects other ecosystem variables like soil organic carbon, nutrient and water stress and surface runoff. This analysis resulted in a high spatial resolution estimate of the environmental impact of cultivating 2G feedstocks in U.S. Midwest and the management interventions needed to sustainable achieve the maximum yield potential of the region.

3. Broader implications of research

3.1 Pronounced shifts from grassland to cultivated areas in the prairie pothole region

Crop rotation patterns can provide information on land-cover change happening in ecologically sensitive areas. We use the crop rotation product identified in fig. 2-8 (82 crop rotations with overall accuracy > 90%, α(75%), β(0.5%)) to estimate grassland conversion to specific crop rotation patterns in the PPR. Recently, Wright and Wimberly (2013) reported that contemporary grassland conversion to corn and soybean cropping (GRCS) from 2006 to 2011 in the PPR is concentrated in the areas surrounding wetlands. Their analysis implicitly assumes that all wetlands affected by GRCS in the PPR existed in or after 2006. However, the areal extent of wetlands was based on National Wetland Inventory (NWI) maps, which were produced from aerial photography taken in the 1970s and 1980s (Johnston, 2013). Thus, we conclude that Wright and Wimberly (2013) overestimated the total area of agricultural expansion through GRCS in close proximity (within 500m) of wetlands. Instead of attributing historical wetland conversion to recent land-use changes in the PPR, here we provide separate estimates of wetland conversion from 1982 - 2007 and 2006 – 2011 using data from National Resources Inventory (NRI) and 2006 NLCD (Xian et al., 2009) respectively.

According to the NRI database, net wetland area declined by 33.3 x 10^3 ha in the U.S. Midwest between 1982 and 1987 (Brady and Flather, 1994). Another 15.0 x 10^3 ha of palustrine and estuarine wetlands in the Northern Plains comprised by the Dakotas and Nebraska were converted to other land-uses between 1992 and 1997 (Summary Report: 1997 National Resources Inventory). This conversion can be equally attributed to agricultural expansion and other causative agents including development and silviculture (Summary Report: 1997 National Resources Inventory). During 1997 - 2007, palustrine and estuarine wetland acreage further declined by 8.0 x 10^3 ha in the Dakotas (Summary Report: 2007 National Resources Inventory). In their analysis, Wright and Wimberly (2013) attribute all wetland conversion between 1982 and 2007 in the PPR to post-2006 GRCS.

Since the non-agricultural land-cover classes in CDL have been derived from NLCD, we combined both wetland classes from the 2006 NLCD, to estimate the total area of agricultural expansion through GRCS within 500 m of wetlands in the PPR. The 2006 NLCD has been previously used to derive national wetland extent and conversion (Potter et al., 2006).We found that the total grassland acreage in close proximity to wetlands that was converted to corn/soybean cropping between 2006 and 2011 was 230,097 ha or 58% of the nearly 400,000 ha estimated by Wright and Wimberly (Fig. 5-2). Notably, while Wright and Wimberly (2013) used the 2006 NLCD to depict wetland cover (Fig. 1B in their paper), they did not use it in their analysis.

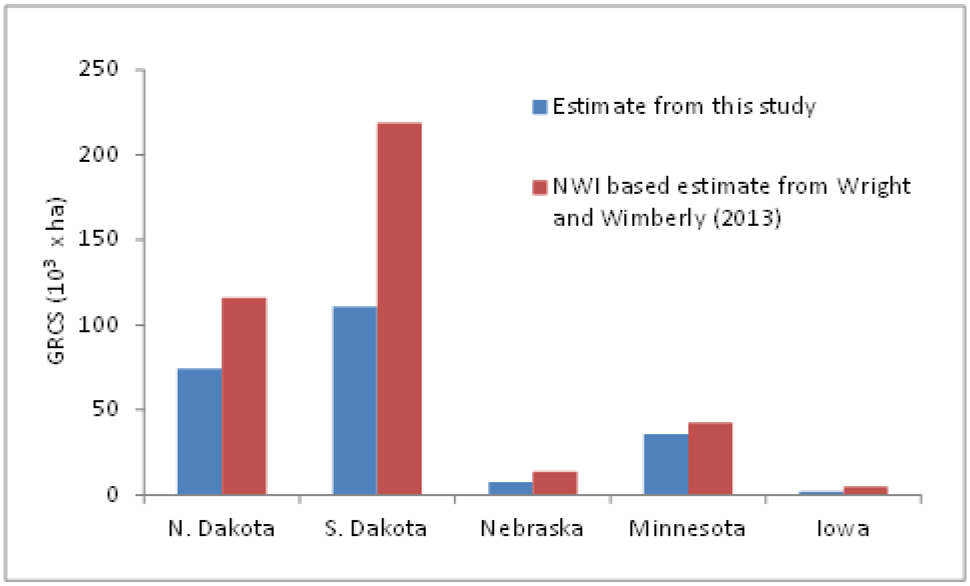

Fig. 5-2. Separate estimates for grassland conversion to crop cultivation in close proximity to wetlands in PPR.

Compared to Wright and Wimberly (2013), who consider only palustrine wetlands in their analysis, our estimate of wetland conversion is conservative since NLCD does not break down wetland classification by palustrine or estuarine type but includes two wetland classes that differ in their vegetation cover. The Energy Independence and Security Act of 2007 explicitly protects wetlands by mandating that renewable biomass may only be harvested from agricultural land that was cultivated prior to the enactment of the law. While biofuel production impacts land-use decisions, it is critical to use appropriate tools and datasets to inform our analysis of environmental impacts of these decisions, lest it divert our focus from other drivers of land-use/cover change, as well as climate change which might have a bigger impact on the fate of wetlands in the region (Johnston, 2013).

If land-cover conversion in the PPR was induced by increased demand for corn and soybean, we should be able to observe intensive cultivation of corn and soybean in the area. To test this hypothesis, we selected representative crop rotations from 2007 to 2011 for the PPR with an overall accuracy of around 90%. Of the 230,097 ha undergoing GRCS in close proximity to wetlands between 2007 and 2011, nearly 42% was devoted to continuous corn or continuous soybean cropping and another 30% was covered by alternating corn and soybean rotations (Table 5-1). These patterns are very surprising especially given the widespread consensus that continuous cropping of soybean for more than two years is not a viable choice on account of enhanced parasite activity (Secchi et al., 2011).

Table 5-1

Crop rotations in the GRCS pixels in PPR. The GRCS pixels lie within a 500m radius of wetlands. Soybean is abbreviated as Soyb.

2007	**2008**	**2009**	**2010**	**2011**	**Acreage (ha)**
Soyb	Soyb	Soyb	Soyb	Soyb	52,050
Corn	Corn	Corn	Corn	Corn	44,867
Soyb	Soyb	Soyb	Corn	Soyb	22,706

Corn	Corn	Corn	Soyb	Corn	17,895
Corn	Soyb	Corn	Soyb	Corn	14,055
Soyb	Corn	Soyb	Corn	Soyb	12,581

3.2 Errors introduced by changing soil datasets

There is a clear link between time needed to complete a modeling run and the spatial resolution of the input datasets. Inherent to this is a tradeoff between feasibility and precision of results. In our modeling of Poplar biomass using ED model, we used two soil datasets with vastly different resolutions: SSURGO at 60m and WISE at 10km. To quantify the error introduced by using a fine resolution dataset like SSURGO at coarser scales, I mapped out the normalized difference between ED biomass estimates using SSURGO data at 60m versus SSURGO averaged to half-degree but mapped out at 60m (Fig. 5-3). Some of the least values for normalized error are in the Sandhill region of Nebraska. This is a positive finding because more than half of the 12 million ha of marginal lands in the U.S. Midwest lie in the Sandhills. Therefore, biomass modeling studies using coarser resolution data in that region are less likely to be off-mark. However, it is possible for individual locations to have much larger associated errors in estimation. Consequently, it is advisable to use the highest resolution soil dataset when possible.

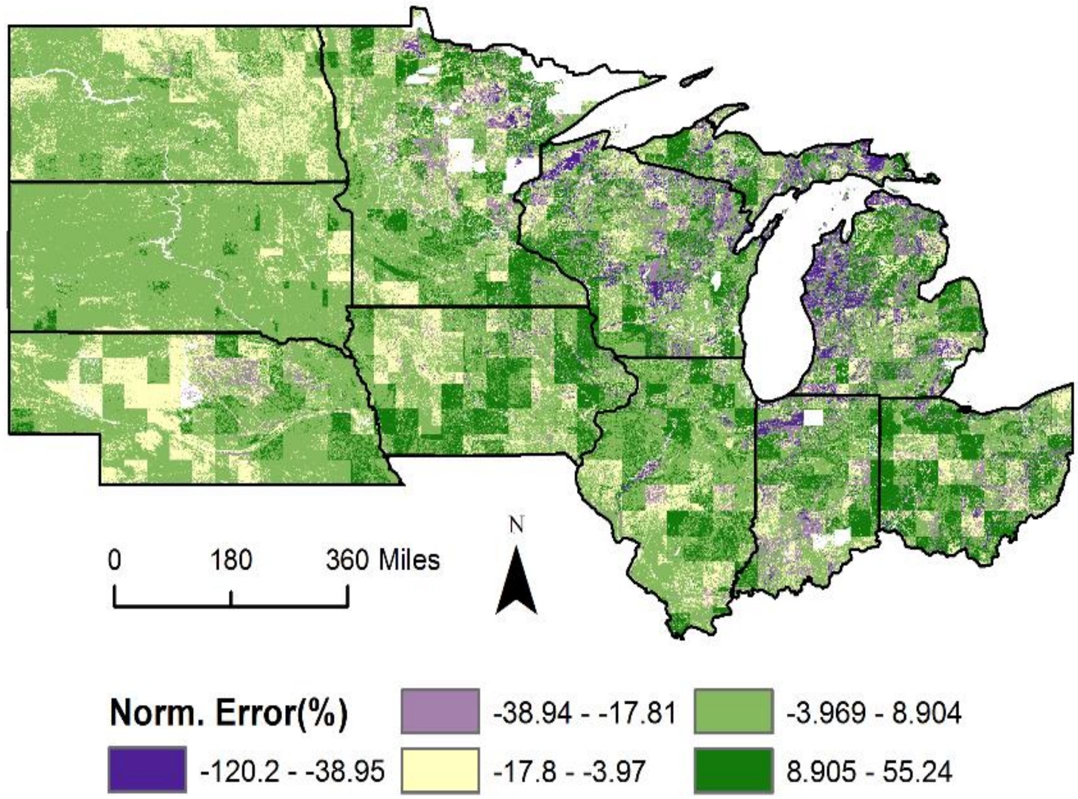

Fig. 5-3. Normalized difference between ED biomass estimates from SSURGO at 60m resolution and SSURGO averaged to half-degree resolution but mapped at 60m

3.3 Yield gap analysis of native grass modeling results

While yield gap analysis tries to identify the difference in potential yields and the yields actually achieved at a given location, it is important to place that in the context of the maximum yield that can be attained when no limitation is present (nutrient, water etc.). To understand this, I plotted four different ecosystem metrics for the native grass modeling, against three management scenarios: no input, enough inputs to close the yield gap, full irrigation and high nutrient input. The ecosystem metrics plotted included: yield, change in soil organic carbon, number of water stress days and denitrification potential. The hypothesis was that yield gap approach is similar to an optimization algorithm in that it tries to maximize yield while mitigating increase in nutrient and water stresses. Indeed, I found that the yield gap scenario closed more than half the difference in yields between the no input and the full input case (Fig. 5-4 (A)). Moreover, this was not accompanied with a reduction in soil organic carbon as in the full input case (Fig. 5-4 (B)). Since irrigation was applied as input in several locations, the number of water stress days did decrease, almost to the level of full management input (Fig. 5-4 (C)). Finally, this increase in irrigation input was accompanied by an increase in denitrification, but less than half the nearly 40 kg/ha of denitrification in the full input case.

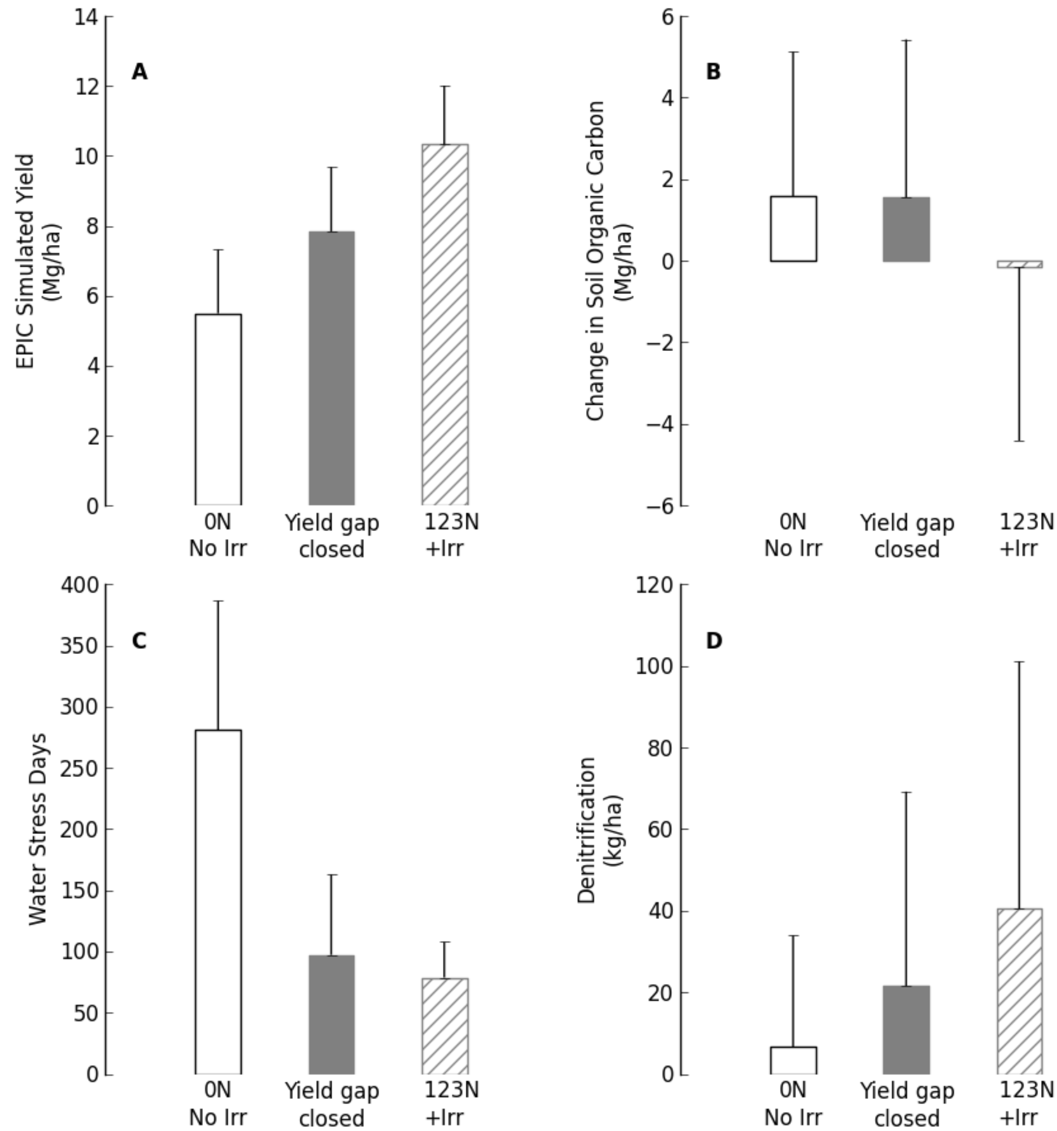

Fig. 5-4. Yield gap analysis of EPIC modeling results for native successional grasses in U.S. Midwest.

In conclusion, the search for beneficial biofuels should focus on the twin objectives of sustainable biofuel feedstocks that do not compete with food crops and do not induce either direct or indirect land-use change. The sustainable biofuel feedstocks include byproducts of human activities (crop residues, forestry wastes) and purpose-grown perennial mixes and woody bioenergy crops. Due care needs to be taken to avoid excessive harvesting of crop residues as it can intensify soil erosion by tenfold or more, increase GHG emissions and also increase eutrophication due to runoff (Pimentel et al., 2009). To minimize GHG emissions from land-use change, we need to identify lands that are initially not storing large quantities of carbon in soil and vegetation but are capable of producing abundant biomass with limited management inputs (Tilman et al., 2009).

When done right, biofuels can provide a solution to meeting the global environmental, food security and energy challenges (Robertson et al., 2008, V.H. Dale et al., 2010, Tilman et al., 2009, K. Kline et al., 2011). This dissertation addressed that by assessing biofuel feedstock production and ecosystem service tradeoffs across a gradient of agricultural management systems in the U.S. Midwest.

Bibliography

A. Ajanovic

Biofuels versus food production: Does biofuels production increase food prices?

Energy, 36(4) (2011), pp. 2070-2076

A. Demirbas

Biofuels sources, biofuel policy, biofuel economy and global biofuel projections

Energ. Convers. Manage., 49(8) (2008), pp. 2106-2116

BP.

Statistical Review of World Energy (2011)

Commission of the European Communities. An EU Strategy for Biofuels. Brussels, Belgium (2006)

D. A. Landis, M. M. Gardiner, W. van der Werf, S. M. Swinton

Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes

Proc. Natl. Acad. Sci. USA., 105(51) (2008), pp. 20552-7

D. Rajagopal, S. E. Sexton, D. Roland-Holst, and D. Zilberman

Challenge of biofuel: filling the tank without emptying the stomach?

Environ. Res. Letters, 2 (2007), pp. 44004

European Biodiesel Board. Statistics: The EU biodiesel industry (2011)

http://www.ebb-eu.org/stats.php (last accessed 23 April 2014)

F. J. J. A. Bianchi, C. J. H. Booij, T. Tscharntke

Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control

Proceedings. Biological sciences / The Royal Society, 273(1595) (2006), pp. 1715-27

G. Marland, T. A. Boden, R. J. Andres, O. R. N. L.

Global, Regional, and National Fossil Fuel CO2 Emissions Oak Ridge, Tenn., U.S.A. (2008)

I. Gelfand, T. Zenone, P. Jasrotia, J. Chen, S. K. Hamilton, G. P. Robertson

Carbon debt of Conservation Reserve Program (CRP) grasslands converted to bioenergy production

Proc. Natl. Acad. Sci., 108(33) (2011), pp. 13864-13869

J. Conti, P. Holtberg, E. I. A. (EIA)

International Energy Outlook (2011)

J. P. W. Scharlemann, W. F. Laurance

Environmental science. How green are biofuels?

Science, 319(5859) (2008), pp. 43-4

K. K. Bertine, E. D. Goldberg

Fossil Fuel Combustion and the Major Sedimentary Cycle

Science, 173 (3993) (1971), pp. 233-235

L. Cifuente, V. H. Borja-Aburto, N. Gouveia, G. Thurston, et al.

Hidden Health Benefits of Greenhouse Gas Mitigation

Science, 293(5533) (2001), pp. 1257-1259

M. M. Gardiner, D. A. Landis, C. Gratton, C. D. DiFonzo, M. O'Neal, J. M. Chacon, M. T. Wayo, N. P. Schmidt, E. E. Mueller, and G. E. Heimpel

Landscape diversity enhances biological control of an introduced crop pest in the north-central USA

Ecol. Appl., 19(1) (2009), pp. 143-154

OECD-FAO

Agricultural Outlook 2011-2020 (2011)

P. Westcott

USDA Agricultural Projections to 2019. DIANE Publishing (2010)

P. A. Kharecha, J. E. Hansen

Implications of “peak oil” for atmospheric CO2 and climate

Global Biogeochem. Cy., 22(3) (2008), pp. 1-10

P. M. Vitousek, H. A. Mooney, J. Lubchenco, J. M. Melillo

Human Domination of Earth’s Ecosystems

Science, 277(5325) (1997), pp. 494-499

Robertson, G. Philip, Dale, H. Virginia, Doering, C. Otto, Hamburg, P. Steven, Melillo, M. Jerry, Wander, M. Michele, Parton, J. William, Adler, R. Paul, Barney, N. Jacob, Cruse, M. Richard, Duke, S. Clifford, Fearnside, M. Philip, Follett, F. Ronald, Gibbs, K. Holly, Goldemberg, Jose, Mladenoff, J. David, Ojima, Dennis, Palmer, W. Michael, Sharpley, Andrew, Wallace, Linda, Weathers, C. Kathleen, Wiens, A. John, Wilhelm, W. Wallace

Sustainable Biofuels Redux.

Science. 322(5898) (2008), pp. 49-50

R. Naylor, A. Liska, and M. Burke

The ripple effect: biofuels, food security, and the environment

Environment: Science and Policy for Sustainable Development, 49(9) (2007), pp. 30-43

S. M. Swinton, K. D. Thelen, L. K. James

Profitability analysis of cellulosic energy crops compared with corn

Agron. J., 102(2) (2010), pp. 675-687

S. D. Donner, C. J. Kucharik

Corn-based ethanol production compromises goal of reducing nitrogen export by the Mississippi River

Proc. Natl. Acad. Sci., 105(11) (2008), pp. 4513

Searchinger, Timothy, Heimlich, Ralph, Houghton, R. A., Dong, Fengxia, Elobeid, Amani, Fabiosa, Jacinto, Tokgoz, Simla, Hayes, Dermot, Yu, Tun-Hsiang

Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change

Science, 319(5867) , pp. 1238-1240

T. D. Meehan, B. P. Werling, D. A. Landis, and C. Gratton

Agricultural landscape simplification and insecticide use in the Midwestern United States

Proc. Natl. Acad. Sci., 108(28) (2011), pp. 11500

T. D. Meehan, A. H. Hurlbert, and C. Gratton

Bird communities in future bioenergy landscapes of the Upper Midwest

Proc. Natl. Acad. Sci., 107(43) (2010), pp. 18533

Tilman, David, Socolow, Robert, Foley, Jonathan A., Hill, Jason, Larson, Eric, Lynd, Lee, Pacala, Stephen, Reilly, John, Searchinger, Tim, Somerville, Chris, Williams, Robert

Beneficial Biofuels—The Food, Energy, and Environment Trilemma

Science, 325(5938) (2009), pp. 270-271

V. H. Dale, K. L. Kline, J. Wiens, J. Fargione

Biofuels: implications for land use and biodiversity

Biofuels and Sustainability Reports

W. Tyner

The U.S. ethanol and biofuels boom: Its origins, current status, and future prospects

BioScience (2008)

W. Coyle

The future of biofuels

Economic Research Service, Washington, DC. (2007)

http://www.aabioethanol.com/news/uploads/20080610044428_The Future of Biofuels - USDA.pdf.

M. Arabi, J. R. Frankenberger, B. A. Engel, J. G. Arnold

Representation of agricultural conservation practices with SWAT

Hydrol. Process., 22(16) (2008), pp. 3042-3055

F. J. J. A. Bianchi, C. J. H. Booij, T. Tscharntke

Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control

Proc. R. Soc. B., 273(1595) (2006), pp. 1715-1727

C. Boryan, Z. Yang, R. Mueller, M. Craig

Monitoring U.S. agriculture: The U.S. Department of Agriculture, National agricultural statistics service, cropland data layer program

Geocarto Int., 26(5) (2011), pp. 341-358

S. J. Brady, C. H. Flather

Changes in wetlands on nonfederal rural land of the conterminous United States from 1982 to 1987

Environ. Manage., 18(5) (1994), pp. 693-705

D.G. Bullock

Crop rotation

Crit. Rev. Plant Sci., 11 (1992), pp. 309–326

M. S. Castellazzi, G. A. Wood, P. J. Burgess, J. Morris, K. F. Conrad, J. N. Perry

A systematic representation of crop rotations

Agr. Syst., 97(1) (2008), pp. 26-33

J. Change, M.C. Hansen, K. Pittman, M. Carroll, C. Dimiceli

Corn and soybean mapping in the United States using MODIS time-series data sets

Agron. J., 99 (2007), pp. 1654–1664

P. Daggupati, K. R. Douglas-Mankin, A. Y. Sheshukov, P. L. Barnes, D. L. Devlin

Field-level targeting using SWAT: Mapping output from HRUs to fields and assessing limitations of GIS input data

Trans. ASABE, 54 (2) (2011), pp. 501-514

N. K. Detlefsen, A. L. Jensen

Modelling optimal crop sequences using network flows

Agr. Syst., 94(2) (2007), pp. 566-572

C. Dimitri, A. B. Effland, N. C. Conklin

The 20th century transformation of U.S. agriculture and farm policy

U.S. Department of Agriculture, Economic Research Service, (2005), Economic information bulletin, no. 3

S. Dogliotti, W. A. H. Rossing, M. K. Van Ittersum

ROTAT, a tool for systematically generating crop rotations

Eur. J. Agron.,19(2) (2003), pp. 239-250

J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polansky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, D. P. M. Zaks

Solutions for a cultivated planet

Nature, 478(7369) (2011), pp. 337-342

F. Gao, T. He, J. G. Masek, Y. Shuai, C. B. Schaaf, Z. Wang

Influence of angular effects and adjustment on medium resolution sensors for crop monitoring

2013 Second International Conference on Agro-Geoinformatics (2013), pp. 296–301

M. M. Gardiner, D. A. Landis, C. Gratton, C. D. DiFonzo, M. O'Neal, J. M. Chacon, M. T. Wayo, N. P. Schmidt, E. E. Mueller, G. E. Heimpel

Landscape diversity enhances biological control of an introduced crop pest in the north-central USA

Ecol. Appl., 19(1) (2009), pp. 143-154

I. Gelfand, R. Sahajpal, X. Zhang, R. C. Izaurralde, K. L. Gross, G. P. Robertson

Sustainable bioenergy production from marginal lands in the U.S. Midwest

Nature, 493(7433), pp. 514–517

S. N. Goward, G. Chander, M. Pagnutti, A. Marx, R. Ryan, N. Thomas, R. Tetrault

Complementarity of ResourceSat-1 AWiFS and Landsat TM/ETM+ sensors

Remote Sens. Environ., 123 (2012), pp. 41–56

T. W. Hertel, A. A. Golub, A. D. Jones, M. O'Hare, R. J. Plevin, D. M. Kammen

Effects of U.S. maize ethanol on global land use and greenhouse gas emissions: estimating market-mediated responses

BioScience, 60(3) (2010), pp. 223-231

O. B. Hesterman, M. P. Russelle, C. C. Shaeffer, G.H. Heichel

Nitrogen utilization from fertilizer and legume residues in legume-corn rotations

Agron. J., 79(4) (1987), pp. 726–731

R. C. Izaurralde, J. R. Williams, W. M. Post, A. M. Thomson, W. B. McGill, L. B., Owens, R. Lal

Long-term modeling of soil C erosion and sequestration at the small watershed scale

Clim. Chang., 80(1-2) (2007), pp. 73-90

M. E. Jakubauskas, D. R. Legates, J. H. Kastens

Crop identification using harmonic analysis of time-series AVHRR NDVI data

Comput. Electron. Agr., 37(1) (2002), pp. 127-139

L. K. James, S. M. Swinton, K. D. Thelen

Profitability analysis of cellulosic energy crops compared with corn

Agron. J., 102(2), pp. 675-687

D. M. Johnson

A comparison of coincident Landsat-5 TM and Resourcesat-1 AWiFS imagery for classifying croplands

Photogramm. Eng. Remote Sens., 74(11) (2008), pp. 1413-1423

W. C. Johnson, B. Werner, G. R. Guntenspergen, R. A. Voldseth, B. Millett, D. E. Naugle, M. Tulbure, R. W. H. Carroll, J. Tracy, C. Olawsky

Prairie wetland complexes as landscape functional units in a changing climate

BioScience, 60(2), pp. 128-140

D. M. Johnson

A 2010 map estimate of annually tilled cropland within the conterminous United States

Agr. Syst., 114 (2013), pp. 95-105

C. A. Johnston

Wetland Losses Due to Row Crop Expansion in the Dakota Prairie Pothole Region

Wetlands, 33(1) (2013), pp. 175-182

R. Lal

Soil carbon sequestration impacts on global climate change and food security

Science, 304(5677) (2004), pp. 1623-1627

D. A. Landis, M. M. Gardiner, W. van der Werf, S. M. Swinton

Increasing corn for biofuel production reduces biocontrol services in agricultural landscapes

Proc. Natl. Acad. Sci., 105(51) (2008), pp. 20552-20557

A. E. Larsen

Agricultural landscape simplification does not consistently drive insecticide use

Proc. Natl. Acad. Sci., 110(38) (2013), pp. 15330-15335

M. Liebman, E. Dyck

Crop rotation and intercropping strategies for weed management

Ecol. Appl., 3(1) (1993), pp. 92-122

J. V. Mannering, D.R. Griffin

Value of crop rotations under various tillage systems

Purdue Univ. Coop. Ext. Serv. Pub. (Tillage) (1981) AY-230

B. A. McCarl, U. A. Schneider

The cost of greenhouse gas mitigation in U.S. agriculture and forestry

Science, 294(21) (2008), pp. 2481-2482

T. D. Meehan, C. Gratton, E. Diehl, N. D. Hunt, D. F. Mooney, S. J. Ventura, B. L. Barham, R. D. Jackson

Ecosystem-Service Tradeoffs Associated with Switching from Annual to Perennial Energy Crops in Riparian Zones of the U.S. Midwest

PLoS ONE 8(11) (2013), e80093

T. D. Meehan, A. H. Hurlbert, C. Gratton

Bird communities in future bioenergy landscapes of the Upper Midwest

Proc. Natl. Acad. Sci, 107(43) (2010), pp. 18533-18538

T. D. Meehan, B. P. Werling, D. A. Landis, C. Gratton

Agricultural landscape simplification and insecticide use in the Midwestern United States

Proc. Natl. Acad. Sci., 108(28) (2011), pp. 11500-11505

M. Mehaffey, R. Van Remortel, E. Smith, R. Bruins

Developing a dataset to assess ecosystem services in the Midwest United States

Int. J. Geogr. Inf. Sci., *25*(4) (2011), pp. 681-695

D. J. Muth Jr, K. M. Bryden, R. G. Nelson

Sustainable agricultural residue removal for bioenergy: A spatially comprehensive U.S. national assessment

Appl. Energ., 102 (2013), pp. 403-417

F.J. Pierce, C.W. Rice

Crop rotation and its impact on efficiency of water and nitrogen use

Cropping Strategies for Efficient Use of Water and Nitrogen, ASA Special Publication 51ASA/CSSA/SSSA, Madison (1988), pp. 21–42

J. D. Plourde, C. P. Bryan, K. P. Burak**Evidence for increased monoculture cropping in the Central United States**

Agriculture, Ecosystems & Environment 165(15), 50-59

C. Potter, S. Klooster, S. Hiatt, M. Fladeland, V. Genovese, P. Gross

Methane emissions from natural wetlands in the United States: Satellite-derived estimation based on ecosystem carbon cycling

Earth Interact., 10(22) (2006), pp. 1-12

N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley

Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000

Global Biogeochem. Cycles, 22(1) (2008), GB1003

E. S. Ristad, P. N. Yianilos

Learning string-edit distance

IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(5) (1998), pp. 522-532

G. P. Robertson, E. A. Paul, R. R. Harwood

Greenhouse gases in intensive agriculture: contributions of individual gases to the radiative forcing of the atmosphere

Science, 289(5486) (2000), pp. 1922-1925

S. Secchi, L. Kurkalova, P. W. Gassman, C. Hart

Land use change in a biofuels hotspot: The case of Iowa, USA

Biomass Bioenergy, 35(6) (2011), pp. 2391-2400

R. Srinivasan, X. Zhang, J. Arnold

SWAT ungauged: hydrological budget and crop yield predictions in the Upper Mississippi River Basin

Trans. ASABE, 53(5) (2010), pp. 1533-1546

A. J. Stern, P. Doraiswamy, B. Akhmedov

Crop rotation changes in Iowa due to ethanol production

Proceedings of the International Geoscience and Remote Sensing Symposium, Boston (2008), pp. V-200 - V-203

D. Tilman, K. G. Cassman, P. A. Matson, R. Naylor, S. Polansky

Agricultural sustainability and intensive production practices

Nature, 418(6898) (2002), pp. 671-677

J. R. G. Townshend, C. Huang, S. N. V. Kalluri, R. S. Defries, S. Liang, K. Yang

Beware of per-pixel characterization of land cover

Int. J. Remote Sens., 21(4) (2000), pp. 839-843

R. E. Turner, N. N. Rabalais

Linking landscape and water quality in the Mississippi River Basin for 200 years

BioScience, 53(6) (2003), pp. 563-572

U.S. Department of Agriculture (2000). *Summary Report: 1997 National Resources Inventory (revised December 2000)*. Natural Resources Conservation Service, Washington, DC, and Statistical Laboratory, Iowa State University, Ames, Iowa

U.S. Department of Agriculture (2009). Summary Report: 2007 National Resources Inventory. Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa

B. D. Wardlow, S. L. Egbert

Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains

Remote Sens. Environ., 112(3), pp. 1096-1116

W. Han, Z. Yang, L. Di, R. Mueller

CropScape: A Web service based application for exploring and disseminating U.S. conterminous geospatial cropland data products for decision support

Comput. Electron. Agr., 84 (2012): pp. 111-123

T. O. West, M. E. Brown, R. M. Duren, S. M. Ogle, R. H. Moss

Definition, capabilities and components of a terrestrial carbon monitoring system

Carbon Management, 4(4) (2013), pp. 413-422

T. O. West, W. M. Post

Soil organic carbon sequestration rates by tillage and crop rotation

Soil Sci. Soc. Am. J., 66(6) (2002), pp. 1930-1946

P. C. Westcott

Ethanol Expansion in the United States: How Will the Agricultural Sector Adjust?

Economic Research Service – United States Department of Agriculture, Washington DC, USA. http://www.ers.usda.gov/Publications/FDS/2007/05May/FDS07D01/.

C. K. Wright, M. C. Wimberly

Recent land use change in the Western Corn Belt threatens grasslands and wetlands

Proc. Natl. Acad. Sci., 110(10) (2013), pp. 4134-4139

G. Xian, C. Homer, J. Fry

Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods

Remote Sens. Environ., 113(6) (2009), pp. 1133-1147

Y. Xiao, C. Mignolet, J. F. Mari, M. Benoît

Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France

Comput. Electron. Agr., 102 (2014), pp. 51-63

X. Zhang, R. C. Izaurralde, D. Manowitz, T. O. West, W. M. Post, A. M. Thomson, V. P. Bandaru, J. Nichols, J. R. Williams

An integrative modeling framework to evaluate the productivity and sustainability of biofuel crop production systems

GCB Bioenergy, 2(5) (2010), pp. 258-277

X. Zhang, R. Sahajpal, D. H. Manowitz, K. Zhao, S. Leduc, M. Xu, W. Xiong, A. Zhang, R. C. Izaurralde, A. M. Thomson, T. O. West, M. W. Post

High Resolution Agroecosystem Modeling: A Case Study on the Sensitivity of Cropland Carbon Flux Simulations to Soils Data Resolution

Sci. Total Environ., (2014), in press.

[1] http://www.nass.usda.gov/research/Cropland/sarsfaqs2.html#Section4_7.0

A. Packer, K. Clay

Soil pathogens and spatial patterns of seedling mortality in a temperate tree

Nature, 404(6775) (2000), pp. 278-281

A. Demirbas

Biofuels sources, biofuel policy, biofuel economy and global biofuel projections

Energy Convers. Manage., 49(8) (2008), pp. 2106-2116

A. B. McBratney, M. J. Pringle

Estimating average and proportional variograms of soil properties and their potential use in precision agriculture

Precision Agriculture, 1(2) (1999), pp. 125-152

A. J. Gijsman, P. K. Thornton, G. Hoogenboom

Using the WISE database to parameterize soil inputs for crop simulation models

Comput. Electron. Agr., 56(2) (2007), pp. 85-100

A. J. Peck, R. J. Luxmoore, J. L. Stolzy

Effects of spatial variability of soil hydraulic properties in water budget modeling

Water Resour. Res., 13(2) (1977), pp. 348-354

A. Jug, F. Makeschin, K. E. Rehfuess, C. Hofmann-Schielle

Short-rotation plantations of balsam poplars, aspen and willows on former arable land in the Federal Republic of Germany. III. Soil ecological effects

Forest Ecol. Manag., 121(1) (1999), pp. 85-99

A. Klute, A. W. Warrick, C. Dirksen, A. Amoozegar

Hydraulic conductivity of saturated soils: Field methods

Methods of soil analysis, 1(2) (1986), pp. 687-733

A. S. Fotheringham, D. W. Wong

The modifiable areal unit problem in multivariate statistical analysis

Environ. Plann. A, 23(7) (1991), pp. 1025-1044

B. C. O'Neill, M. Oppenheimer

Climate change: dangerous climate impacts and the Kyoto Protocol.

Science, 296(5575) (2002), pp. 1971-1972

B. J. Cosby, G. M. Hornberger, R. B. Clapp, T. Ginn

A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils

Water Resour. Res., 20(6) (1984), pp. 682-690

Cramer, Wolfgang, Alberte Bondeau, F. Ian Woodward, I. Colin Prentice, Richard A. Betts, Victor Brovkin, Peter M. Cox et al

Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models

Glob. Change Biol., 7(4) (2001), pp. 357-373

C. G. Wilkerson, S. D. Mansfield, F. Lu, S. Withers, J. Y. Park, S. D. Karlen, E. Gonzales-Vigil, D. Padmakshan, F. Unda, J. Rencoret, J. Ralph

Monolignol Ferulate Transferase Introduces Chemically Labile Linkages into the Lignin Backbone

Science, 344(6179) (2014), pp. 90-93

C. W. Cobb, P. H. Douglas

A theory of production

The American Economic Review (1928), pp. 139-165

D. A. Netzer

Growth, yield, and disease resistance of 7- to 12-year-old poplar clones in the north central United States

General Technical Report NC-229 (2002)

D. A. Netzer

Growth, yield, and disease resistance of 7- to 12-year-old poplar clones in the north central United States

General Technical Report NC-229 (2002)

D. E. Jelinski, J. Wu

The modifiable areal unit problem and implications for landscape ecology

Landscape Ecol., 11(3) (1996), pp. 129-140

D. J. Kaczmarek, D. R. Coyle, M. D. Coleman

Survival and growth of a range of Populus clones in central South Carolina USA through age ten: Do early assessments reflect longer-term survival and growth trends?

Biomass Bioenerg., 49 (2013), pp. 260-272

D. N. Huntzinger, C. Schwalm, A. M. Michalak, K. Schaefer, A. W. King, Y. Wei, A. Jacobson, S. Liu, R. B. Cook, W. M. Post, G. Berthier, D. Hayes, M. Huang, A. Ito,

H. Lei, C. Lu, J. Mao, C. H. Peng, S. Peng, B. Poulter, D. Riccuito, X. Shi, H. Tian, W. Wang, N. Zeng, F. Zhao, Q. Zhu

The North American Carbon Program Multi-scale synthesis and Terrestrial Model Intercomparison Project Part 1: Overview and experimental design

Geoscientific Model Development Discussions, 6(3), pp. 3977-4008

D. R. Legates, G. J. McCabe

Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation

Water Resour. Res., 35(1) (1999), pp. 233-241

D. Wang, D. LeBauer, M. Dietze

Predicting yields of short-rotation hybrid poplar (Populus spp.) for the United States through model-data synthesis

Ecol. Appl., 23(4) (2013), pp. 944-958

E. A. Hansen, R. A. McLaughlin, P. E. Pope

Biomass and nitrogen dynamics of hybrid poplar on two different soils: implications for fertilization strategy

Can. J. Forest Res., 18(2) (1988), pp. 223-230

G. C. Hurtt, S. W. Pacala, P. R. Moorcroft, J. Caspersen, E. Shevliakova, R. A. Houghton, B. I. I. Moore

Projecting the future of the U.S. carbon sink

Proc. Natl. Acad. Sci., 99(3) (2002), pp. 1389-1394

G. C. Hurtt, P. R. Moorcroft, S. W. Pacala, S. A. Levin

Terrestrial Models and Global Change: Challenges for the Future

Glob. Change Biol., 4(5) (1998), pp. 581-590

G. S. Campbell

A simple method for determining unsaturated conductivity from moisture retention data

Soil Sci., 117(6) (1974), pp. 311-314

H. O. Hartley

The maximum F-ratio as a short-cut test for heterogeneity of variance

Biometrika, 37 (1950), pp. 308-12

I. C. Burke, C. M. Yonker, W. J. Parton, C. V. Cole, D. S. Schimel, K. Flach

Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils

Soil Sci. Soc. Am. J., 53(3) (1989), pp. 800-805

I. Gelfand, R. Sahajpal, X. Zhang, R. C. Izaurralde, K. L. Gross, G. P. Robertson

Sustainable bioenergy production from marginal lands in the U.S. Midwest

Nature, 493(7433) (2013), pp. 514–517

J. A. Foley, I. C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, A. Haxeltine

An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics

Global Biogeochem. Cy., 10(4) (1996), pp. 603-628

J. Hill, E. Nelson, D. Tilman, S. Polasky, D. Tiffany

Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels

Proc. Natl. Acad. Sci., 103(30) (2006), pp. 11206-11210

J. Gamble, G. Johnson, C. Sheaffer, D. Current, D. Wyse

Establishment and early productivity of perennial biomass alley cropping systems in Minnesota, USA

Agroforest. Syst., 88(1) (2014), pp. 75-85

J. G. Isebrands, E. P. McDonald, E. Kruger, G. Hendrey, K. Percy, K. Pregitzer, J. Sober, D. F. Karnosky

Growth responses of Populus tremuloides clones to interacting elevated carbon dioxide and tropospheric ozone

Environ. Pollut., 115(3) (2001), pp. 359-371

J. Kennedy, R. Eberhart

Particle swarm optimization

Proceedings of IEEE international conference on neural networks, 4(2) (1995), pp. 1942-1948

Kauter, Dirk, I. Lewandowski, W. Claupein

Quantity and quality of harvestable biomass from Populus short rotation coppice for solid fuel use—a review of the physiological basis and management influences

Biomass and Bioenerg., 24(6) (2003), pp. 411-427

K. E. Taylor

Summarizing multiple aspects of model performance in a single diagram

J. Geophys. Res.-Atmos. (1984–2012), 106(D7) (2001), pp. 7183-7192

K. Pearson

Notes on regression and inheritance in the case of two parents

P. Roy. Soc. Lond., 58 (1895), pp. 240-242

M. Albani, D. Medvigy, G. C. Hurtt, P. R. Moorcroft

The contributions of land-use change, CO2 fertilization, and climate variability to the Eastern U.S. carbon sink

Glob. Change Biol., 12(12) (2006), pp. 2370-2390

M. C. Dietze, P. R. Moorcroft

Tree mortality in the eastern and central United States: patterns and drivers

Glob. Change Biol., 17(11) (2011), pp. 3312-3326

M. F. Wilson, A. Henderson-Sellers

A global archive of land cover and soils data for use in general circulation climate models

J. Climatol., 5(2) (1985), pp. 119-143

M. J. Aylott, E. Casella, I. Tubby, N. R. Street, P. Smith, G. Taylor

Yield and spatial supply of bioenergy poplar and willow short-rotation coppice in the UK

New Phytol., 178(2) (2008), pp. 358-370

M. J. Tallis, E. Casella, P. A. Henshall, M. J. Aylott, T. J. Randle, J. I. Morison, G. Taylor

Development and evaluation of Forest Growth-SRC a process-based model for short rotation coppice yield and spatial supply reveals poplar uses water more efficiently than willow

GCB Bioenergy, 5(1) (2013), pp. 53-66

M. M. Palmer, J. A. Forrester, D. E. Rothstein, D. J. Mladenoff

Establishment phase greenhouse gas emissions in short rotation woody biomass plantations in the Northern Lake States, USA

Biomass Bioenerg., 62 (2014), pp. 26-36

M. Zambrano-Bigiarini, R. Rojas

A model-independent Particle Swarm Optimisation software for model calibration

Environ. Modell. Softw., 43 (2013), pp. 5-25

N. H. Batjes

Total carbon and nitrogen in the soils of the world

Eur. J. Soil Sci., 47(2) (1996), pp. 151-163

N. H. Batjes

ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid (Version 3.0)

ISRIC-World Soil Inf. Rep, 8 (2005)

N. H. Batjes

Harmonized soil profile data for applications at global and continental scales: updates to the WISE database

Soil Use Manage., 25(2) (2009), pp. 124-127

P. R. Moorcroft, G. C. Hurtt, S. W. Pacala

A Method for Scaling Vegetation Dynamics: The Ecosystem Demography Model (ED)

Ecol. Monogr., 71(4) (2001), pp. 557–586

Palmer, M. Marin, Jodi A. Forrester, David E. Rothstein, and David J. Mladenoff

Conversion of open lands to short-rotation woody biomass crops: site variability affects nitrogen cycling and N2O fluxes in the U.S. Northern Lake States

GCB Bioenergy (2013)

Riemenschneider, Don, J. G. Isebrands, W. E. Berguson, D. I. Dickmann, R. B. Hall, C. A. Mohn, G. R. Stanosz, G. A. Tuskan

Poplar breeding and testing strategies in the north-central U.S.: demonstration of potential yield and consideration of future research needs

The Forestry Chronicle., 77(2) (2001), pp. 245-253

S. W. Pacala, C. D. Canham, J. Saponara, J. A. Silander Jr, R. K. Kobe, E. Ribbens

Forest models defined by field measurements: estimation, error analysis and dynamics

Ecol. Monogr., 66(1) (1996), pp. 1-43

S. W. Pacala, G. C. Hurtt, D. Baker, P. Peylin, R. A. Houghton, R. A. Birdsey, L. Heath, E. T. Sundquist, R. F. Stallard, P. Ciais, P. Moorcroft, J. P. Caspersen, E.

Shevliakova, B. Moore, G. Kohlmaier, E. Holland, M. Gloor, M. E. Harmon, S.-M. Fan, J. L. Sarmiento, C. L. Goodale, D. Schimel, C. B. Field

Consistent land-and atmosphere-based U.S. carbon sink estimates

Science, 292(5525) (2001), pp. 2316-2320

USEPA (United States Environmental Protection Agency)

Regulation of fuels and fuel additives: changes to renewable fuel standard program, final rule

Fed. Regist. 75(2010), pp. 14669–5320

W. L. Headlee, R. S. Zalesny Jr, D. M. Donner, R. B. Hall

Using a process-based model (3-PG) to predict and map hybrid poplar biomass productivity in Minnesota and Wisconsin, USA

BioEnergy Research, 6(1) (2013), pp. 196-210

Z. Xue, L. Qu, X. Yang

Potential production and spatial distribution of hybrid poplar as a biofuel crop in Connecticut, USA

International Journal of Agricultural and Biological Engineering, 7(2) (2014), pp. 10-18

A. A. Klingebiel, P. H. Montgomery

Land-capability classification (No. 210)

Soil Conservation Service, U.S. Department of Agriculture (1961)

B. A. Simmons, D. Loque, H. W. Blanch

Next-generation biomass feedstocks for biofuel production

Genome Biol., 9(2) (2008), 242-242

C. A. Johnston

Wetland Losses Due to Row Crop Expansion in the Dakota Prairie Pothole Region

Wetlands, 33(1) (2013), pp. 175-182

C. K. Wright, M. C. Wimberly

Reply to Kline et al.: Cropland data layer provides a valid assessment of recent grassland conversion in the Western Corn Belt

Proc. Natl. Acad. Sci., (31) (2013b), E2864

C. O. Stockle, J. R. Williams, N. J. Rosenberg, C. A. Jones

A method for estimating the direct and climatic effects of rising atmospheric carbon dioxide on growth and yield of crops: Part I—Modification of the EPIC model for climate change analysis

Agr. Syst., 38(3) (1992a)., pp. 225-238

C. O. Stockle, P. T. Dyke, J. R. Williams, C. A. Jones, N. J. Rosenberg

A method for estimating the direct and climatic effects of rising atmospheric carbon dioxide on growth and yield of crops: Part II—Sensitivity analysis at three sites in the Midwestern USA

Agr. Syst., 38(3) (1992b), pp. 239-256

D. Tilman, J. Hill, C. Lehman

Carbon-negative biofuels from low-input high-diversity grassland biomass

Science, 314(5805) (2006), 1598-1600

D. B. Lobell, K. G. Cassman, C. B. Field

Crop yield gaps: their importance, magnitudes, and causes

Annu. Rev. Env. Resour., 34(1) (2009), pp. 179

D. Tilman, C. Balzer, J. Hill, B. L. Befort

Global food demand and the sustainable intensification of agriculture

Proc. Natl. Acad. Sci., 108(50) (2011), pp. 20260-20264

D. Tilman, J. Hill, C. Lehman

Carbon-Negative Biofuels from Low-Input High-Diversity Grassland Biomass

Science, 314(5805) 2006, pp. 1598-1600

F. Mesinger, G. DiMego, E. Kalnay, K. Mitchell, P. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. Berbery, M. Ek, F. Yun, R. Grumbine, W. Higgins, L. Hong, L. Ying, G. Manikin, D. Parrish, S. Wei

North American Regional Reanalysis

Bulletin Of The American Meteorological Society, 87(3) (2006), pp. 343-360

G. P. Robertson, S. K. Hamilton, S. J. Del Grosso, W. J. Parton

The biogeochemistry of bioenergy landscapes: carbon, nitrogen, and water considerations

Ecol. Appl., 21(4) (2011), pp. 1055-1067

H. J. Causarano, J.N. Shaw, A. J. Franzluebbers, D. W. Reeves, R. L. Raper, K. S. Balkcom, M. L Norfleet, R. C. Izaurralde

Simulating field-scale soil organic carbon dynamics using EPIC

Soil Sci. Soc. Am. J., 71 (2007), pp. 1174-1185

H. J. Causarano, P. C. Doraiswamy, G. W. McCarty, J. L. Hatfield, S. Milak, A. J. Stern

EPIC Modeling of Soil Organic Carbon Sequestration in Croplands of Iowa

J. Environ. Qual., 37 (2008), pp. 1345–1353

H. P. Apezteguía, R. C. Izaurralde, R. Sereno

Simulation study of soil organic matter dynamics as affected by land use and agricultural practices in semiarid Córdoba, Argentina

Soil Tillage Res., 102 (2009), pp. 101-108

I. C. Prentice, W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, A. M. Solomon

Special paper: a global biome model based on plant physiology and dominance, soil properties and climate

J. Biogeogr. (1992), pp. 117-134

I. C. Burke, C. M. Yonker, W. J. Parton, C. V. Cole, D. S. Schimel, K. Flach

Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils

Soil Sci. Soc. Am. J., 53(3) (1989), pp. 800-805

I. Gelfand, T. Zenone, P. Jasrotia, J. Chen, S. K. Hamilton, G. P. Robertson

Carbon debt of Conservation Reserve Program (CRP) grasslands converted to bioenergy production

Proc. Natl. Acad. Sci., 108(33) (2011), pp. 13864-13869

I. Gelfand, R. Sahajpal, X. Zhang, R. C. Izaurralde, K. L. Gross, G. P. Robertson

Sustainable bioenergy production from marginal lands in the U.S. Midwest

Nature, 493(7433) (2013), pp. 514–517

J. Fargione, J. Hill, D. Tilman, S. Polasky, P. Hawthorne

Land clearing and the biofuel carbon debt

Science, 319(5867) (2008), 1235-1238

J. O. Kim, F. J. Kohout

Analysis of variance and covariance: subprograms ANOVA and ONEWAY

Statistical package for the social sciences, 2 (1975), pp. 398-433

K. D. Eggemeyer, T. Awada, D. A. Wedin, F. E. Harvey, X. Zhou

Ecophysiology of Two Native Invasive Woody Species and Two Dominant Warm-Season Grasses in the Semiarid Grasslands of the Nebraska Sandhills

Int. J. Plant Sci., 167(5) (2006), pp. 991-999

K. Jensen, C. D. Clark, P. Ellis, B. English, J. Menard, M. Walsh, D. de la Torre Ugarte

Farmer willingness to grow switchgrass for energy production

Biomass Bioenergy, 31(11) (2007), 773-781

K. G. Cassman

Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture

Proc. Natl. Acad. Sci., 96(11) (1999), pp. 5952-5959

K. R. Brye, J. M. Norman, L. G. Bundy, S. T. Gower

Nitrogen and carbon leaching in agroecosystems and their role in denitrification potential

J. Environ. Qual., 30(1) (2001), pp. 58-70

K. Neumann, P. H. Verburg, E. Stehfest, C. Muller

The yield gap of global grain production: A spatial analysis

Agr. Syst., 103(5) (2010), pp. 316-326

M. K. van Ittersum, K. G. Cassman, P. Grassini, J. Wolf, P. Tittonell, Z. Hochman

Yield gap analysis with local to global relevance—a review

Field Crop. Res., 143 (2013), pp. 4-17

R. C. Izaurralde, J. R. Williams, W. B. McGill, N. J. Rosenberg, M. C. Jakas

Simulating soil C dynamics with EPIC: Model description and testing against long-term data

Ecol. Model., 192(3) (2006), pp. 362-384

R. C. Izaurralde, J. R. Williams, W. M. Post, A. M. Thomson, W. B. McGill, L. Owens, R. Lal **Long-term modeling of soil C erosion and sequestration at the small watershed scale**

Climatic Change, 80 (2007), pp. 73-90

R. Licker, M. Johnston, J. A. Foley, C. Barford, C. J. Kucharik, C. Monfreda, N. Ramankutty

Mind the gap: how do climate and agricultural management explain the 'yield gap'of croplands around the world?

Global Ecol. Biogeogr., 19(6) (2010), pp. 769-782

R. K. Manatt, A. Hallam, L. A. Schulte, E. A. Heaton, T. Gunther, R. B. Hall, K. J. Moore

Farm-scale costs and returns for second generation bioenergy cropping systems in the U.S. Corn Belt

Environ. Res. Lett., 8(3) (2013), 035037

S. C. Davis, W. J. Parton, S. J. D. Grosso, C. Keough, E. Marx, P. R. Adler, E. H. DeLucia

Impact of second-generation biofuel agriculture on greenhouse-gas emissions in the corn-growing regions of the U.S.

Frontiers in Ecology and the Environment, 10(2) (2011), pp. 69-74

S. Kang, W. Post, D. Wang, J. Nichols, V. Bandaru, T. West

Hierarchical marginal land assessment for land use planning

Land Use Policy, 30(1) (2013), pp. 106-113

T.G. Farr, P.A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, D. Alsdorf

The Shuttle Radar Topography Mission. Rev. Geophys. 45(2) 2007, pp. RG2004

Soil Survey Staff. 2012. Soil Survey Geographic (SSURGO) Database for [U.S.]. Available online at http://soildatamart.nrcs.usda.gov, verified on March 10, 2014

T. Gaiser, I. de Barros, F. Sereke, F. M. Lange

Validation and reliability of the EPIC model to simulate maize production in small-holder farming systems in tropical sub-humid West Africa and semi-arid Brazil

Agr. Ecosyst. Environ., 135(4) (2010), pp. 318-327

T. D. Meehan, C. Gratton, E. Diehl, N. D. Hunt, D. F. Mooney, S. J. Ventura, B. L. Barham, R. D. Jackson

Ecosystem-Service Tradeoffs Associated with Switching from Annual to Perennial Energy Crops in Riparian Zones of the U.S. Midwest

PLoS ONE 8(11) (2013), e80093

USEPA (United States Environmental Protection Agency) (2010)

Regulation of fuels and fuel additives: changes to renewable fuel standard program, final rule

Fed. Regist. 75 14669–5320

W. H. McNab, D. T. Cleland, J. A. Freeouf, J. Keys, J. E, G. J. Nowacki, C. A. Carpenter Description of ecological subregions: sections of the conterminous United States. U.S Department of Agriculture, Forest Service, Washington DC (2005)

X. Zhang, R. Sahajpal, D. H. Manowitz, K. Zhao, S. D. LeDuc, M. Xu, W. Xiong, A. Zhang, R. C. Izaurralde, A. M. Thomson, T. O. West, W. M. Post
Multi-scale geospatial agroecosystem modeling: A case study on the influence of soil data resolution on carbon budget estimates
Sci. Total Environ., 479 (2014), pp. 138-150

X. Zhang, R. C. Izaurralde, D. Manowitz, T. O. West, W. M. Post, A. M. Thomson, V. P. Bandaru, J. Nichols, J. R. Williams
An integrative modeling framework to evaluate the productivity and sustainability of biofuel crop production systems
GCB Bioenergy, 2(5) (2010), pp. 258-277

Y. Liu, L. Wu, J. A. Baddeley, C. A. Watson
Models of biological nitrogen fixation of legumes. A review
Agronomy for Sustainable Development (2010)