GLOBAL CHANGE BIOLOGY
BIOENERGY

GCB Bioenergy (2014) **6**, 14–25, doi: 10.1111/gcbb.12047

Global simulation of bioenergy crop productivity: analytical framework and case study for switchgrass

SHUJIANG KANG*, SUJITHKUMAR SURENDRAN NAIR*, KEITH L. KLINE*†, JEFFREY A. NICHOLS‡, DALI WANG*, WILFRED M. POST*, CRAIG C. BRANDT§, STAN D. WULLSCHLEGER*, NAGENDRA SINGH‡ and YAXING WEI*

**Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA, †Global Sustainable Bioenergy Project crop modeling task and ORNL Center for Sustainable Bioenergy, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA, ‡Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA, §Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA*

Abstract

A global energy crop productivity model that provides geospatially explicit quantitative details on biomass potential and factors affecting sustainability would be useful, but does not exist now. This study describes a modeling platform capable of meeting many challenges associated with global-scale agro-ecosystem modeling. We designed an analytical framework for bioenergy crops consisting of six major components: (i) standardized natural resources datasets, (ii) global field-trial data and crop management practices, (iii) simulation units and management scenarios, (iv) model calibration and validation, (v) high-performance computing (HPC) simulation, and (vi) simulation output processing and analysis. The HPC-Environmental Policy Integrated Climate (HPC-EPIC) model simulated a perennial bioenergy crop, switchgrass (*Panicum virgatum* L.), estimating feedstock production potentials and effects across the globe. This modeling platform can assess soil C sequestration, net greenhouse gas (GHG) emissions, nonpoint source pollution (e.g., nutrient and pesticide loss), and energy exchange with the atmosphere. It can be expanded to include additional bioenergy crops (e.g., miscanthus, energy cane, and agave) and food crops under different management scenarios. The platform and switchgrass field-trial dataset are available to support global analysis of biomass feedstock production potential and corresponding metrics of sustainability.

Keywords: biofuel, biomass, EPIC, model, soil organic carbon, sustainability

Received 18 October 2012 and accepted 9 November 2012

Introduction

Models that can incorporate management practices and quantify environmental effects are necessary to assess indicators of sustainability associated with deployment of biomass crops at global scale. Models that provide transparent evaluations of potential opportunities and trade-offs among energy options across highly variable agro-ecosystems would be useful for guiding policy decisions. Geographically variable climate (e.g., precipitation, temperature), soil properties, and crop management practices (e.g., fertilizer, irrigation, tillage, seed, and harvest) affect productivity and should be integrated into a modeling architecture to reflect the implications of alternative land-use scenarios. Climate change, geospatial allocation of land to more productive uses, and management intensification will affect bioenergy, food, and other important ecosystem services (Millennium Ecosystem Assessment (MEA), 2005). The use of agro-ecosystem models capable of incorporating these factors is critical for quantifying future biological production, as well as economic and environmental effects (Lobell *et al.*, 2011).

The Global Sustainable Bioenergy (GSB) project is an international initiative that seeks to expand our understanding of the potential to meet a significant share of future energy demand with bioenergy without compromising production of food or environmental services (Kline *et al.*, 2011a; Lynd *et al.*, 2011). The GSB project and other recent analyses (Fischer & Schrattenholzer, 2001; Hoogwijk *et al.*, 2003; Smeets *et al.*, 2007; Offermann *et al.*, 2011) have highlighted a dichotomy in current global estimates of bioenergy potential by noting that some studies show very large potential, whereas others show minimal potential. The recent Intergovernmental Panel on Climate Change (IPCC) Special Report on Renewable Energy illustrated bioenergy supply potentials ranging from 50 to 500 EJ yr^{-1} (Intergovernmental Panel on Climate Change (IPCC), 2011). The

Correspondence: Shujiang Kang, tel. + 865-574-5948, fax + 865-574-9501, e-mail: kangs@ornl.gov

GSB project is currently assessing factors that influence the potential for bioenergy to sustainably supply 150 EJ annually, (target level in the IEA Blue Map Scenario (International Energy Agency (IEA), 2010) corresponding to 23% of future primary energy supply) without compromising food production and other services. Better data and analytical tools are necessary to develop more consistent and reliable simulations of the potential productivity and effects of dedicated bioenergy crops. Effects of bioenergy production on food supply and security reflect legitimate and important concerns that can be addressed more effectively when production systems for both can be analyzed using an integrated framework.

A range of process-based crop models have been successfully adapted or developed to simulate biomass yield for a number of emerging bioenergy crops (Nair *et al*., 2012). The difficulties of conducting such simulations at large scales are underscored by the lack of high-resolution global modeling platforms and barriers associated with availability of spatially explicit input data, visualization, and postprocessing and analysis of model outputs. Zhang *et al.* (2010) proposed a framework to simulate crops in a nine-county area of southwest Michigan using data with 56 meter resolution. A national simulation of switchgrass and miscanthus was conducted and highlighted the large and dynamic variability in biomass production at a spatial resolution of 32 km across the conterminous United States (Miguez *et al*., 2012). However, applying high-resolution spatial-temporal process-based models and managing corresponding outputs present significant computational challenges (Washington *et al*., 2009; Nichols *et al*., 2011), particularly as scales increase beyond field or county size. High-performance computing (HPC) technology using clusters, grids, or cloud computing can be used to scale-up from individual field-level simulations to high-resolution global simulations. Wang *et al.* (2005) designed generic software architecture for implementing spatially explicit ecosystem models on computing grids. Recently, Nichols *et al.* (2011) demonstrated the potential of parallel processing in modeling agricultural systems, achieving a 40-fold increase in speed by running 140 000 simulations concurrently on a Linux-based computing cluster.

Although we are aware of other efforts with similar aims, our review of the literature has not identified a documented platform that provides a process-based model capable of simulating dedicated bioenergy crop production and environmental effects at a global scale. Other efforts have focused primarily on (a) assembling global datasets (e.g., the Global Agro-ecological Zones datasets (FAO/IIASA, 2007), the Global Earth Observation – Benefit Estimation Project of European Union (International Institute for Applied System Analysis (IIASA), 2010), and databases specific to agriculture and yields such as Monfreda *et al*., 2008 and Ramankutty *et al*., 2008); or (b) generating productivity estimates based on statistical reports and historic yield data (e.g., Mueller *et al*., 2012; and the Global Trade Analysis Project (GTAP) as described by Dimaranan, 2006). Most of these analyses did not include energy crops, focused on a few management variables (e.g., fertilization), or were not geospatially explicit and therefore could not downscale and compare simulation results to historic production data from specific locations for validation and calibration. Other efforts (e.g., USEPA, 2010) have attempted to combine multiple model simulations with datasets that estimate carbon stocks and deforestation to calculate the effects of crop expansion, but these efforts lack causal analysis to support the hypothetical relationships, suffer from inconsistency among data sources and model assumptions, and therefore, are also limited in their ability to perform systematic calibration and validation against historic data (Kline *et al*., 2011b). This study does not attempt to repeat those efforts or to address many other well-documented data-related issues (e.g., International Institute for Applied System Analysis (IIASA), 2010; Monfreda *et al*., 2008; CBES (Center for BioEnergy Sustainability), 2009). Instead, we focus on describing a High-Performance Computing-Environmental Policy Integrated Climate (HPC-EPIC) model platform and testing the hypothesis that it can perform rapid, global-scale, biophysical process simulations.

In the following sections, we describe the design of the modeling platform, our strategy for assembling large datasets, and methods used to calibrate and analyze a global simulation of a bioenergy crop, switchgrass (*Panicum virgatum* L.). Conclusions about the lessons learned, key problems encountered, and possible next steps to expand the utility of this modeling platform are discussed in the final sections. Supplemental information is posted online to provide additional details on methods, data processing efficiency, and uncertainty.

Materials and methods

Platform design

The simulation platform is based on an analytical framework designed for bioenergy crops and consisting of six major components: (i) standardized natural resources datasets, (ii) global field-trial data and crop management practices, (iii) simulation units and management scenarios, (iv) model calibration and validation, (v) high-performance computing (HPC) simulation, and (vi) simulation output processing and analysis (Fig. 1). The natural resources and management datasets provide model

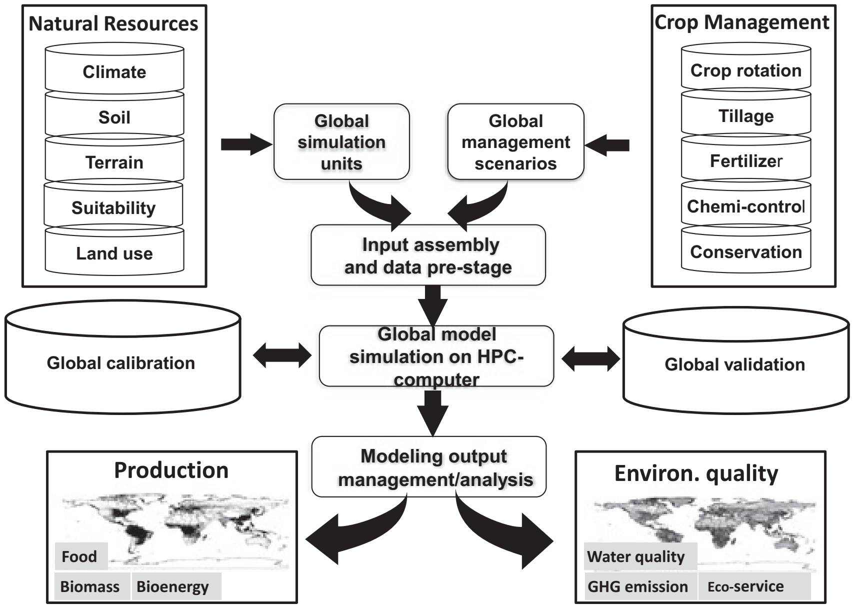

Fig. 1 Global simulation platform for sustainability assessment of bioenergy crops.

inputs of climate, soil, topography, land uses, and various management practices commonly used by biophysical crop models (e.g., fertilization, tillage, planting, and harvest timing). Half-degree spatial simulation units, determined by the climate data, are grouped by agro-ecological zones. We used HPC clusters for performing the simulations. The final component of the analytical framework involves the extraction and organization of modeling outputs to databases necessary to support analysis.

Description of EPIC and HPC-EPIC

Environmental policy integrated climate is a process-based biogeochemical model for the soil–crop–atmosphere continuum (Williams *et al.*, 1984; Izaurralde *et al.*, 2006). Major processes simulated by EPIC include plant growth, development and production, nutrient cycling and nonpoint sources pollution, hydrology, emissions of GHGs, and plant management practices. After over 30 years of improvement and over 1000 peer-reviewed publications, the model has evolved into an integrated tool to meet multiple needs of production estimation, environmental and sustainability assessment, and climate change (Easterling *et al.*, 2001; Mausbach & Dedrick, 2004; Gassman *et al.*, 2005; Liu *et al.*, 2008). Currently, EPIC is able to simulate over 100 crops including 10 bioenergy crops, such as switchgrass, miscanthus, and poplar. It has been widely calibrated and applied in over 30 countries and regions (Gassman *et al.*, 2005).

Since inception, EPIC has been applied to field- or plot-scale production and environmental management. In more recent years, it has been applied for regional-scale production and environmental analyses. For example, in the Conservation Effects Assessment Project of the U.S. Department of Agriculture Natural Resource Conservation Service (USDA NRCS), EPIC was used to estimate nutrient, soil, and carbon losses from agriculture for the contiguous states of the United States (Potter *et al.*, 2004).

Our version of HPC-EPIC is described in Nichols *et al.* (2011). HPC-EPIC enables the parallel execution of thousands of simulations. Achieving high-speed processing involves data packaging, parallelization, and dedicated processor resources. The approach, details related to data processing efficiency and the time requirements associated with the various steps required to prepare and conduct the case study simulation are described in more detail in the Supporting Online Material (SOM).

Model inputs

The specific input variables for EPIC are listed and explained in Williams *et al.* (1984). Major inputs for HPC-EPIC that are accessible from published databases include daily weather datasets, soil properties, and landscape attributes. For the design and testing of a global platform, we used the half-degree global land mask from CRU-NCEP (Viovy, 2010) to generate 62 482 simulation units and to subset daily weather input data over 30 years (1980–2010) for each of these units. Daily CRU-NECP data consist of four sets of 6 h weather data, including total shortwave solar radiation (w m^{-2}), air pressure (Pa), temperature (K), U-wind (m s^{-1}), V-wind (m s^{-1}), and precipitation (mm). We developed Python scripts (as presented in the SOM) to extract and calculate the daily radiation (MJ m^{-2} d^{-1}), maximum temperature and minimum temperature (°C), precipitation (mm), average relative humidity (%), and average wind speed (m s^{-1}) from the CRU-NCEP data. Daily maximum and minimum temperatures (°C) are the highest and lowest temperature of the daily data, respectively. Daily precipitation (mm) is the sum of the four, 6 h precipitations (mm), and daily solar radiation (MJ m^{-2} d^{-1}) is converted from the sum of total shortwave solar radiations (W m^{-2} s^{-1}). Average relative humidity (%) is calculated from temperature (K), pressure (Pa), and air-specific humidity (g g^{-1}) using the equations developed by Buck (1981). Daily average wind speed is calculated from 6 h wind speed (m s^{-1}). Monthly averaged weather data files were computed from the daily weather files for use in the HPC-EPIC modeling. A total of 62 482 weather files (one for each simulation unit) were generated.

Soil data for each simulation unit were produced from the Harmonized World Soil Database (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). The HWSD data were first resampled to the half-degree simulation cell mask by selecting the dominant soil type within each simulation cell for two soil layers (0–30 cm and 30–100 cm). Fractions of sand, silt and clay, gravel content, reference bulk density, organic carbon, cation exchange capacity, texture, and pH for the dominant soil type were then extracted. Preliminary slope data were calculated from 30 arc second global elevation data, GEOTOPO30 (United States Geological Survey (USGS), 1996) using the Spatial Analyst of ArcGIS10 (ESRI, 2011). Slopes for each half-degree simulation unit were the majority value resampled from the calculated slopes at 30 arc second scale. The goal of this case study is to test performance of the simulation platform. Given the coarse resolution of the simulation units, we made no prejudgment about slope constraints on potential productivity. The model platform allows for the inclusion of such constraints in future simulations along with a range of management practices, from mechanized production to manual systems, depending on the slope of the terrain.

We chose switchgrass (*Panicum virgatum* L.) as a representative perennial bioenergy crop for this study. Switchgrass is a C4 warm-season grass, naturally distributed in Africa, North America, and South America (Parrish & Fike, 2005). Each ecotype of switchgrass includes multiple cultivars. Intensive experimental studies of switchgrass have been conducted in North America. Studies in Asia and Europe have also been reported (e.g., Alexopoulou *et al*., 2008; Ma *et al*., 2011). Two identified ecotypes of switchgrass (lowland and upland) are closely associated with climate and topographic conditions (Sanderson *et al*., 1996; Casler, 2005; Parrish & Fike, 2005; Casler *et al*., 2007). An upland ecotype generally adapts better to areas that are in high latitudes (>40°N or 40°S) or arid and sloping terrain whereas a lowland ecotype switchgrass is more suited to areas that are relatively warm (lower latitudes), humid, and flat terrain (Casler *et al*., 2007).

Agronomic characteristics and management for switchgrass cultivars in different regions were obtained from the US field-trial database developed by Wullschleger *et al.* (2010). This database was expanded by adding yield and management data from published and unpublished research resources from other countries. Over 1400 observations from five continents describing site, cultivar, management, harvesting practices, and yield data were assembled. The dataset was further classified into upland and lowland ecotypes and then screened to exclude outlier observations that recorded extremely low or high yield without explanation or clear description of management practices. The classification and screening process generated a total of 84 data points across eight ecological zones that were used for model calibration.

We used ecological zones, latitude, and slope to assign the most suitable cultivars of switchgrass to each simulation unit. The 20 ecological zones classified by the Food and Agricultural Organization (FAO) reflect potential plant water availability and vegetation distribution from the evergreen tropical rainforest zone to the boreal tundra zones (Forest Resources Assessment Program (FRA) of UN-FAO, 2001). Upland ecotype switchgrass is simulated in the arid and cold ecological zones. Specifically, if the latitude is greater than 40°N or 40°S, or if the majority value slope of the simulations unit (as defined above) is greater than 5%, or if the area is located in an arid or semiarid ecological zone, an upland switchgrass cultivar is assumed to be planted in the area; otherwise, management and productivity files for a lowland cultivar are applied. We selected 33 representative switchgrass cultivars from the calibration dataset and assigned these to the simulation units (Table 1).

Eighty management files were constructed for the 33 switchgrass cultivars. Each management file specified major types of tillage, planting and harvesting dates, potential heat units, seeding rate, fertilizer and pesticides application time and amount, and irrigation. In this study, we designed management practices based on the databases developed by Kiniry *et al.* (1996), Wullschleger *et al.* (2010), and Nichols *et al.* (2011). For the initial design and testing of the global platform, we simplified some management practices to facilitate analysis of model functions. For example, we uniformly specified 60 kg ha^{-1} of nitrogen fertilizer annually, and 40 kg ha^{-1} phosphorus every 6 years, for all management files. We also assumed no irrigation use and that switchgrass is replanted every 12 years in all management files. However, the planting and harvesting dates were derived from the field-trials database for eight zones. Twelve ecological zones lacked experimental data (see Table 1) and for these, we either used data for the region with the most similar climate or we adapted the planting and harvest dates for other major crops grown in the zone. For example, wet and dry seasons were considered to establish the planting and harvest dates appropriate for tropical shrubland and tropical rainforest zones.

Model calibration

Adequate data and methods to conduct calibration and validation are critical in developing reliable estimates of productivity and environmental effects using process-based models. The quality and representativeness of data inputs and parameterization methods affect the reliability of the simulation outputs. In HPC-EPIC, 57 crop parameters can be adjusted for different Switchgrass cultivars during calibration. The most commonly used and sensitive parameters for model calibration include radiation use efficiency, leaf area indices, harvest index, base and optimal temperature, nutrient parameters, root ratio, and a few other physiological thresholds such as maximum stomatal conductance, and aeration factor. These parameters are measureable for switchgrass and should ideally be calibrated although many are not reported in the field-trial literature.

Table 1 Ecotypes and cultivars of switchgrass are determined for regional calibration and global simulation with three criteria (global ecological zone, slope, and latitude), where N/A indicates that the criterion is not applied. Latitude unit is degree. The total number of simulation units (Cells) in the region identified for a specified ecotype and cultivar, number of management practice scenarios (Management), and number of calibration sites used in the case study (Calibration) are listed for each ecological zone. The ecological zones are based on Forest Resources Assessment Program (FRA) of UN-FAO (2001)

Global ecological zone	Slope	Latitude	Cells	Ecotype	Cultivar-ID	Management	Calibration
Tropical rainforest	≤5%	N/A	5781	Lowland	1	3	0
	>5%	N/A	204	Upland	2	2	0
Tropical moist deciduous forest	≤5%	N/A	3751	Lowland	3	3	0
	>5%	N/A	238	Upland	4	3	0
Tropical dry forest	≤5%	N/A	2720	Lowland	5	3	0
	>5%	N/A	144	Upland	6	3	0
Tropical shrubland	≤5%	N/A	62	Lowland	7	3	0
	>5%	N/A	2685	Upland	8	3	0
Tropical desert	N/A	N/A	4291	Upland	9	2	0
Tropical mountain system	N/A	N/A	2283	Upland	10	3	0
Subtropical humid forest	≤5%	N/A	1924	Lowland	11	3	3
	>5%	N/A	327	Upland	12	3	2
Subtropical dry forest	≤5%	N/A	791	Lowland	13	3	3
	>5%	N/A	156	Upland	14	3	3
Subtropical steppe	≤5%	N/A	1737	Lowland	15	3	3
	>5%	N/A	201	Upland	16	3	3
Subtropical desert	N/A	N/A	2490	Upland	17	2	0
Subtropical mountain system	N/A	N/A	2018	Upland	18	2	2
Temperate oceanic forest	≤5%	<40	1161	Lowland	19	2	1
	≤5%	≥40	790	Upland	20	2	3
	>5%	<40	69	Upland	21	2	3
Temperate continental forest	≤5%	<40	437	Lowland	22	2	4
	≤5%	≥40	26	Upland	23	2	3
	>5%	<40	2907	Upland	24	2	2
Temperate steppe	≤5%	<40	317	Lowland	25	2	4
	≤5%	≥40	2477	Upland	26	2	4
	>5%	<40	45	Upland	27	2	4
Temperate desert	N/A	N/A	2315	Upland	28	2	0
Temperate mountain system	N/A	N/A	3871	Upland	29	2	5
Boreal coniferous forest	N/A	N/A	5445	Upland	30	2	0
Boreal tundra woodland	N/A	N/A	2782	Upland	31	2	0
Boreal mountain system	N/A	N/A	4544	Upland	32	2	0
Polar	N/A	N/A	3493	Upland	33	2	0

Our calibration strategy starts with parameter values obtained from the literature and previous switchgrass simulation (e.g., Kiniry *et al.*, 1996, 2008, 2011; Parrish & Fike, 2005). Separate calibrations are conducted for lowland and upland cultivars in zones containing sufficient field-trial data. We manually adjusted the radiation use efficiency, leaf area index, plant heat unit, and plant density to best match major physiological features and biomass production to the degree data permit. Table 1 shows the zones and cultivars for which calibration datasets were generated. A description of parameters for the cultivars in the zones without calibration is presented in SOM. It is noteworthy that the zones lacking field-trial data are those least likely to be used for switchgrass production as they encompass four high-latitude (polar) zones, three desert zones, and five tropical zones.

The platform design includes a validation process that could not be conducted due to data limitations. We found that many of the observations reported from field trials could not be used for calibration or validation because management practices or trial conditions were not adequately specified. For example, some field trials tested special management practices such as multiple seasonal cuts or harvests. Other trials reported only 2 or 3 years of data which limited their utility for validation.

Simulation and postdata processing

After preparing model inputs and calibration, we assembled 50 packages for HPC-EPIC execution on a 56 processor Linux cluster at Oak Ridge National Laboratory. Custom Python scripts were developed to parse, check, and import the simulation output data into an open-source PostgreSQL v8.3 database to facilitate future validation and analysis. Using PostgreSQL queries, we extracted data for production and environmental analysis and generated maps in ArcGIS 10 (ESRI, 2011). The initial simulation results provide estimates of switchgrass productivity and environmental effects across all zones although switchgrass is not expected to be grown in zones such as boreal mountains, tropical forests, and tundra. Switchgrass is also unlikely to be planted extensively on current productive cropland.

To better illustrate potential for a bioenergy crop on more representative lands, we restricted the initial simulation results of the projected potential switchgrass production to pasturelands. We used the delineations in Ramankutty *et al.* (2008) which identified approximately 28 million square kilometers of pastureland. These pasturelands are located in 15 of the 20 ecological zones. However, the majority (about 65%) of pasturelands occur in the eight ecological zones where field-testing data supported model calibration. For this case study, we did not consider specialized management systems adapted to arid and highly marginal lands. Therefore, when identifying potential biomass production, any simulation units where the average annual productivity was estimated to be less than 2 Mg ha^{-1} yr^{-1} were not considered. This screening was done by converting these low-productivity simulation cells to 'no data' in ArcGIS and then recalculating the total available pastureland area and pastureland switchgrass biomass productivity. The data presented in Table 2 were calculated using national map boundaries and zonal statistics in ArcGIS. The average annual potential switchgrass biomass production is calculated based on 30 years of simulated production. The production values were determined by first multiplying the average annual productivity (Mg ha^{-1}) by the area of the simulation unit (ha), and then by the percentage of the simulation unit that was in pasture (Ramankutty *et al.*, 2008) to get average production (Mg) per simulation unit. Then, all units are summed for a given country (Table 2).

Results

Model calibration analysis

High-quality data are critical for successful HPC-EPIC calibrations. Our calibration results indicated that HPC-EPIC was able to simulate global switchgrass productivity reasonably well, with r^2 = 0.78 for lowland ecotype cultivars and r^2 = 0.55 for upland ecotype cultivars (Fig. 2) for the ecological zones and subzones where data were available (Table 1). Root mean square errors (RMSE) for the two switchgrass ecotype cultivars are less than 3.17 Mg ha^{-1}. The calibration for lowland ecotype switchgrass cultivars was slightly better than that of upland ecotype switchgrass cultivars. In general, biomass productivity estimates for the lowland ecotype were within 1 standard deviation of the observed values. However, HPC-EPIC overestimated productivity of upland switchgrass ecotypes in some zones, especially in lower yield ranges. Although the calibrations of the different cultivars of two switchgrass ecotypes for the eight major ecological zones appeared to be acceptable, more data are needed to improve the calibration of the upland ecotype, extend calibration across other ecological zones not adequately represented in the

Table 2 The total area of pastureland, the area where simulated production exceeded the 2 Mg threshold (pastureland considered), dry biomass production on considered pastureland, and the average annual production over the 30 year simulation on considered pastureland are provided for selected countries

Country	Simulated production (10^6 Mg)	Average annual yield (Mg ha^{-1} yr^{-1})	Pastureland considered (10^3 km^2)	Total pastureland (10^3 km^2)
Brazil	2500	14	1700	1700
China*	1300	9	1400	2700
United States*	1100	7	1500	2200
Colombia	500	17	300	300
Australia	500	4	1200	2700
Sudan	500	7	700	1000
Russia	400	6	700	800
Mozambique	400	10	400	400
Mexico	300	6	500	600
Argentina	300	10	300	900
Madagascar	300	12	200	200
Mongolia	200	4	400	800
France*	90	9	90	100
Germany*	40	8	50	50
United Kingdom*	20	5	40	80
Italy*	20	8	30	30

*Calibration conducted.

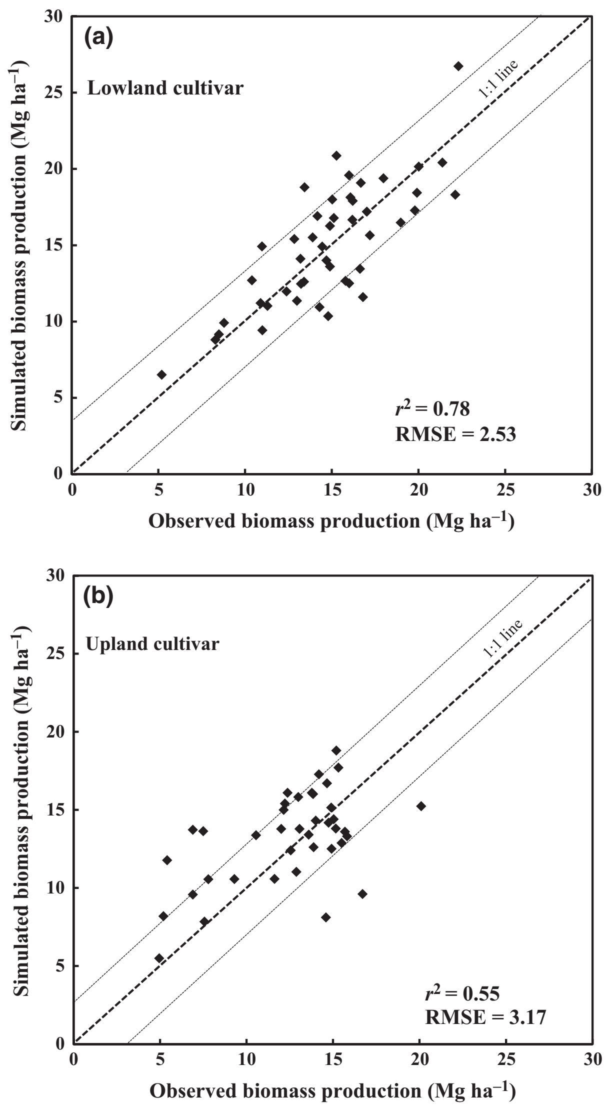

Fig. 2 Global calibration of switchgrass for the HPC-EPIC model: (a) calibration for lowland ecotype switchgrass and (b) calibration for upland ecotype switchgrass. The dashed line is 1 : 1 line, and the dotted lines show 1 standard deviation from the 1 : 1 line.

field-trial database, and to conduct formal validation procedures.

Simulation results

We illustrate the simulated HPC-EPIC global biomass production potential of switchgrass in Fig. 3a and production on pasturelands in Fig. 3b. The potential biomass productivity reported in the figures and discussed below is the average annual value expressed in dry Mg ha^{-1} yr^{-1} based on 30 years of switchgrass simulation (1981–2010) for each simulation unit. The average switchgrass productivity ranges from near zero in boreal and desert areas to a maximum of 35 Mg ha^{-1} in a few simulation units in the lowland tropics (e.g., parts of Brazil and central Africa). With the current calibrated parameters, the areas showing the highest potential productivity are consistently in the moist tropics, but as noted above, these are also areas where reliable field-trial data were not available to support crop parameter calibration (Table 1). Therefore, we emphasize that the biomass productivity estimates in these tropical zones remain speculative until relevant field-trial data can be assembled and analyzed.

The HPC-EPIC model simulates production potential based on biophysical factors across the globe, regardless of current land cover, markets, and zoning. High simulated productivity does not necessarily identify the best opportunities for cultivating switchgrass. This case study simulates high potential productivity in areas with favorable climate and soils even though other local conditions and current land cover may limit or prohibit actual cultivation. Areas with high productive potential also include lands that have been under cultivation for food crop production (Fig. 3a). This is not surprising because agricultural land is expected to represent relatively fertile areas in the ecological zones.

Simulated productivity on pasturelands is shown in Fig. 3b. Although a screening procedure (see Methods) is designed to remove lands classified as cropland, forests, and others to focus only on pastureland, this visualization exaggerates the pastureland area because it portrays all simulation units reported as containing pasture even though many units contain only a small portion of pastureland along with other land classes.

Table 2 provides examples of estimates of total switchgrass biomass production on pasturelands by country. The total area of pastureland, the area where simulated production exceeded the 2 Mg threshold (pastureland considered), and the average annual production over the 30-year simulation are provided for selected countries (Table 2). While this simulation estimates the highest production potential in Brazil (over 2 billion tons yr^{-1}), this is due largely to the high average productivity which requires further data to support model calibration. Australia has larger pastureland area than Brazil and the United States, and Australia has lower annual production because of low average pastureland productivity as simulated by EPIC.

Environmental effects

In addition to total aboveground biomass production, the HPC-EPIC platform has the capacity to calculate environmental effects associated with crop management and production, such as changes in soil organic carbon

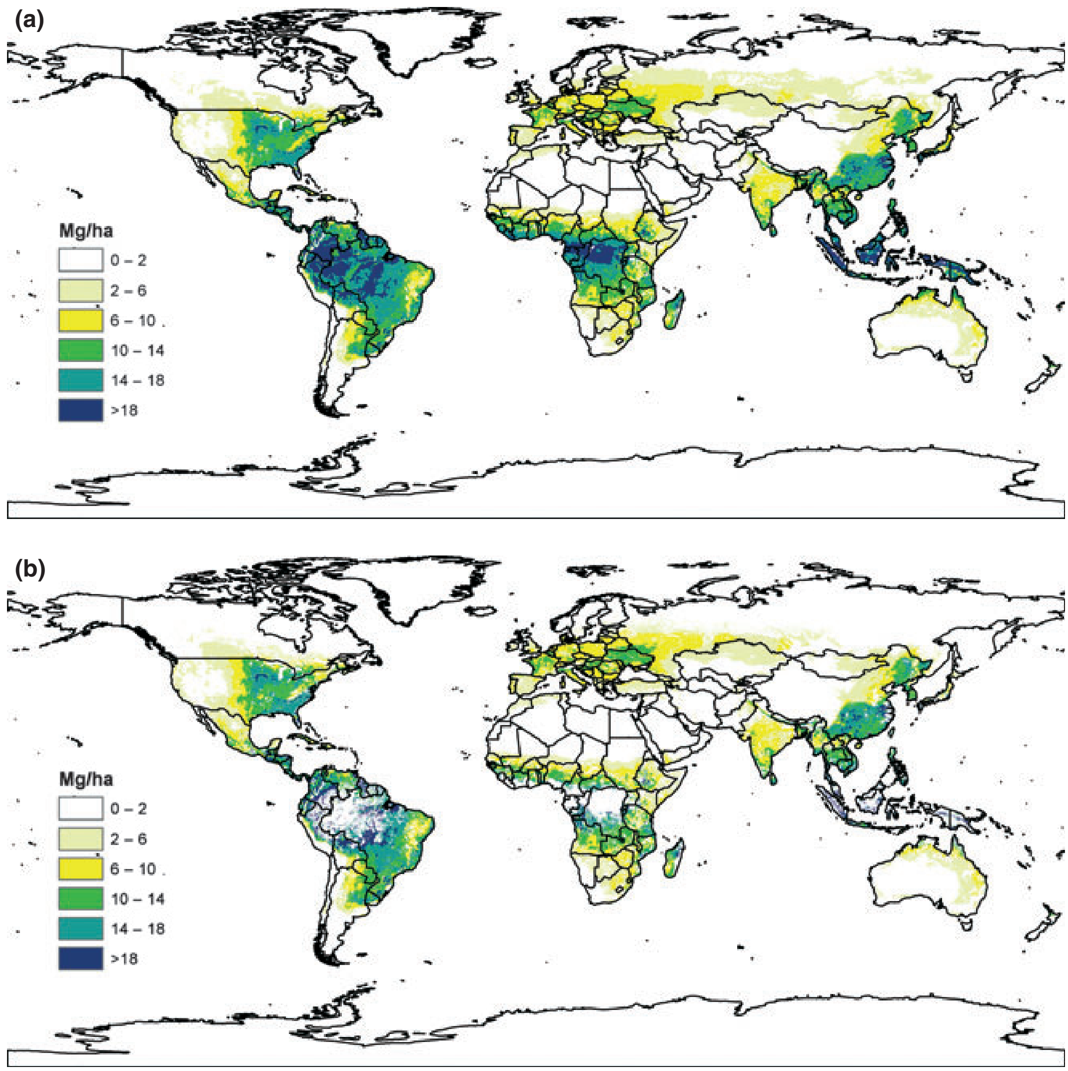

Fig. 3 Maps of simulated switchgrass productivity (Mg dry biomass ha^{-1} yr^{-1}) on all land globally (a) and on pastureland (b) at half-degree resolution. Note limitations on estimates for tropical areas discussed in text.

(SOC), belowground biomass, erosion, and GHG emission. We used our switchgrass case study to estimate changes in SOC as a demonstration of how this platform could be applied to assess environmental effects of land management choices at large scales (Fig. 4). As illustrated, switchgrass cultivation could cause losses of SOC in areas with low productivity because of soil disturbance and biomass removal leading to greater losses than switchgrass growth replaces. As very few data points were available for the calibration of environmental effects of switchgrass production, calibration of SOC changes could not be performed. Thus, the SOC simulations from this initial testing of the HPC-EPIC platform should be regarded as illustrative of model function and capabilities, but highly uncertain in terms of absolute values.

Efficiency of HPC-EPIC simulation

The primary aim of this study was to test the hypothesis that an HPC-EPIC platform could be designed to perform rapid, global-scale, biophysical process simulations. The global switchgrass simulation processed 30 years of climate data, 80 management files, and over 60 000 simulation units. The simulation of 30 years of switchgrass cultivation at a global scale required less than 3 h using the HPC-EPIC platform on ORNL computers. Approximately 12 months of research staff time was invested to prepare data and management files and to develop, calibrate, and test the modeling platform (see SOM Figures S1 and S2). Several considerations affecting the efficiency of the simulation and modeling process, and the utility of this investment to

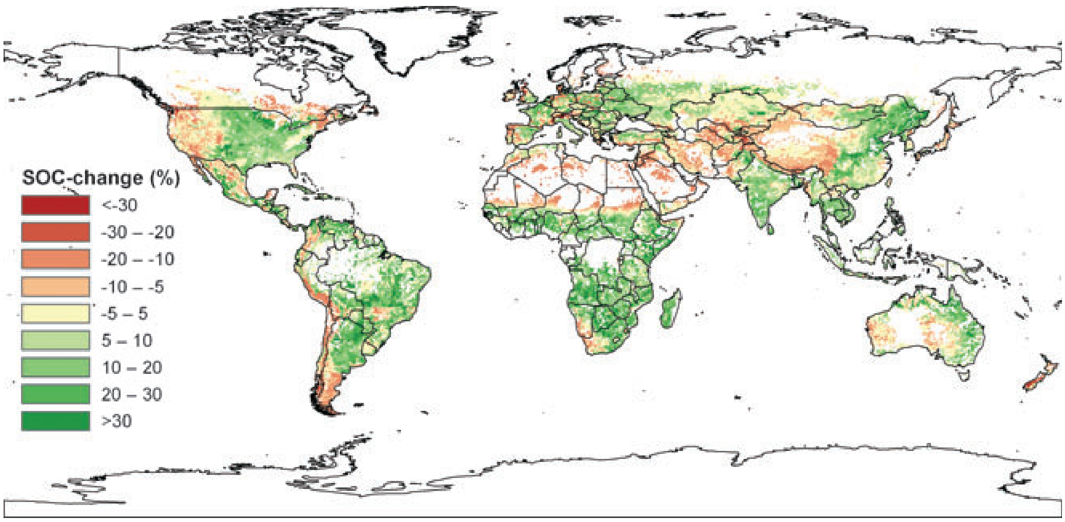

Fig. 4 Map illustrating soil organic carbon change (%) on pasturelands at half-degree resolution with 30 year switchgrass cultivation simulated by the HPC-EPIC model.

facilitate future simulations, are further described in the SOM.

Discussion

Uncertainty in global biomass productivity simulation of switchgrass

All crop model simulations require data for model input, parameterization, and validation, and the full extent of uncertainty associated with input data and model process relationships is often unknown and not easily quantified. Compared with other more traditional crops, detailed agronomic information for growth, development, management, and environmental effects of bioenergy crops is relatively scarce (Thomson *et al.*, 2005; Heaton *et al.*, 2010; Nair *et al.*, 2012). Major uncertainties in the HPC-EPIC model simulations can be attributed to three sources: model inputs, model parameters, and model calibration and validation.

We used the CRU-NCEP weather data for our simulation units. Studies have reported the underestimate or overestimates of CRU-NCEP weather variables in some regions. Smith *et al.* (2001) showed that wind speeds have been routinely underestimated in the CRU-NCEP data, and Poccard *et al.* (2000) indicated that the NCEP data underestimated precipitation during rainy season. Underestimation of temperature and precipitation by the CRU-NCEP reanalysis data in the Tibetan plateau region has also been observed (Xie *et al.*, 2007). Wind, precipitation, and temperature influence productivity. However, the use of 30 years of weather data and the resolution of the simulation mitigate the influence that these factors may have on the general productivity patterns and results presented here. Uncertainties associated with soil and management data inputs are discussed in the SOM.

The quality, extent, and quantity of the switchgrass experimental trials data are perhaps the largest source of uncertainty in this case study. Many records lacked detailed soil, management, and other information which affect productivity and therefore, the reliability of global simulation results. After a careful screening was applied to develop the calibration datasets from field-trial data, it become impossible to conduct both calibration and independent validation procedures. Several tropical zones lacked the minimum data necessary for calibration making simulations in those areas particularly uncertain. Of the 20 global ecological zones (Table 1), most bioenergy crop production is expected to fall within just 11 zones as a large majority of agriculture and pasture are located in these areas. The field-trial dataset contained information for calibration in eight of these eleven zones. To facilitate expansion and sharing of data across the research community, the switchgrass field-trial dataset used for the case study simulations was uploaded to the Bioenergy Knowledge Discovery Framework (see SOM and the web page: https://www.bioenergykdf.net/content/global-switchgrass-field-trial-production-and-management-dataset).

Although the HPC-EPIC platform is designed to provide extensive information on environmental indicators associated with modeled crop production, bioenergy crops lack sufficient data for the calibration of these variables. Thus, the projections for changes in soil organic carbon and similar environmental effects remain uncertain. In addition, to estimate environmental effects, crop models must incorporate baseline assumptions that also reflect uncertainty. As spatial scale increases, it becomes more difficult for models to reflect the heterogeneity and variability in baseline conditions, disturbances, and other factors that influence environmental outcomes.

Lessons learned and simulation improvement

Data collection, preparation, and postprocessing are the most time-intensive tasks for high-resolution global simulations (Figure S3 in SOM). For example, we generated about 250 000 input files, 625 000 output files, and an 11 GB PostgreSQL database for data processing and analysis. Higher resolution simulations will generate even larger data management demands. However, once data are assembled and screened for quality, the HPC-EPIC platform is able to run simulations quickly and to assess different hypotheses with relatively small additional effort.

Several aspects of the current simulation platform can be improved. First, we need to expand the number of field trials in the switchgrass database, particularly for tropical areas and the areas where little or no experimental data are available. Environmental effects of switchgrass production systems such as SOC change, erosion, and GHG emission should be included in the database for sustainability analysis although these variables are not often reported. Second, we need to further classify calibration and validation zones and identify the most suitable switchgrass cultivars, particularly for future high-resolution simulations such as 1/8 degree and 5 min resolutions. Third, we need to increase the resolution to provide results at a scale that best meets the needs of decision makers. The half-degree simulation unit resolution in this study limits our ability to characterize the details of biomass variability and environmental effects in topographically complex regions. We also need to improve the efficiency of large database management and data analysis for higher resolution modeling.

Future work

One of the great social and political barriers to expanding biomass production for energy is the perception of competition between biomass and food production. The issue has been difficult to resolve for many reasons, including the lack of modeling platforms capable of generating accurate results while integrating complex management and cropping systems at large scales. The HPC-EPIC platform is positioned to help address these questions and related queries about potential changes in climate and water availability. However, additional high-quality data will be needed to facilitate the future steps outlined below.

We do not foresee significant computational barriers to conducting higher resolution simulations as improved datasets become available. The current framework is flexible enough to accommodate simulations at 1/8 degree or 5 min, but the packaging of simulations should be aligned with the numbers of available computer nodes on a cluster. Simulation speed increases as work is distributed across more nodes. However, higher resolution will present new challenges for data preparation, postprocessing, and analysis. While we were able to convert half-degree weather data into the format required by EPIC in a few days, a more efficient data processing method would be needed if high-resolution simulations such as 1/8 degree or 5 min resolution are conducted. In addition, methods used for postsimulation extraction of the results from the EPIC output files into PostgreSQL databases would require updates (Nichols *et al.*, 2011).

Modeling platforms could also be further improved by analyzing and assembling complementary high-resolution datasets on omitted factors that influence productivity and environmental effects. For example, probabilities and risks associated with disturbances are likely to affect potential yield and management strategies. Incorporation of such factors in the modeling process could better reflect the range of observed and expected variability that results from cyclic drought, common pests, fire, flooding, or climate change.

This case study documents a first effort to design and test a platform capable of global simulation of a dedicated bioenergy crop. The HPC-EPIC platform estimated productivity and environmental effects of switchgrass cultivation across the globe. This platform can be adapted with relative ease to incorporate additional data as they become available and this will be necessary to further calibrate switchgrass in tropical zones. Work is also needed to parameterize and validate process-based models for other candidate bioenergy crops. Future plans include (i) integrating other emerging bioenergy crops (e.g., miscanthus, energy cane, and crassulacean acid metabolism (CAM) crops) into the simulation platform, (ii) improving the acquisition and sharing of high-quality field experimental data for model development and testing, and (iii) upgrading data management for efficient execution of large-scale simulations and processing of large input and output datasets on a supercomputer. These modifications will improve reliability, permit downscaling or upscaling, and better meet the needs of policy makers and planners who require timely assessments of the opportunities and trade-offs among alternative systems to supply food, feed, fiber, and bioenergy.

Our approach advances prior work by demonstrating a method that allows a spatially explicit, agro-ecosystem process model (EPIC) to efficiently run simulations at a global scale. The platform provides a foundation that can support more complex modeling that incorporates management variables, rotations, and quantification of agricultural and bioenergy production as well as environmental effects. Advances in global-scale process modeling will enable the scientific community to further evaluate sustainable bioenergy production systems for various levels of decision making. The research team encourages national and international collaboration to share data and assessment results. Our hope is that by sharing our datasets and modeling approach, we can encourage and facilitate the collection of additional data to permit improved simulations in the future.

Acknowledgments

Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. We gratefully acknowledge ORNL, Martin Keller, and Gary Jacobs for supporting the research and the Global Sustainable Bioenergy (GSB) Project. We would like to thank Drs. R. Cesar Izaurralde and Xuesong Zhang at Pacific Northeast National Laboratory, Jimmy Williams at Texas A&M University, and James Kiniry of USDA-ARS for their support and discussions on global HPC-EPIC model simulation and calibration. We also thank Dr. Lee Lynd at Dartmouth College, Executive Director of the GSB Project, for encouragement and guidance over the course of this study.

References

Alexopoulou E, Sharma N, Papatheohari Y, Christou M, Piscioneri I, Panoutsou C, Pignatelli V (2008) Biomass yields for upland and lowland switchgrass varieties grown in the Mediterranean region. *Biomass and Bioenergy*, **32**, 926–933.

Buck AL (1981) New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, **20**, 1527–1532.

Casler MD (2005) Ecotype variation among switchgrass populations from the northern USA. *Crop Science*, **45**, 388–398.

Casler MD, Vogel KP, Taliaferro CM *et al.* (2007) Latitudinal and longitudinal adaptation of switchgrass populations. *Crop Science*, **47**, 2249–2260.

CBES (Center for BioEnergy Sustainability) (2009) Land-use change and bioenergy. Report from the 2009 workshop, ORNL/CBES-001, U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy and Oak Ridge National Laboratory, Center for Bioenergy Sustainability. Available at: http://www.ornl.gov/sci/ees/cbes.shtml (accessed 10 October 2012).

Dimaranan BV (2006) Global Trade, Assistance, and Production: The GTAP 6 Data Base, Center for Global Trade Analysis, Purdue University. Available online at: http://www.gtap.agecon.purdue.edu/databases/v6/v6_doco.asp (accessed 20 October 2012).

Easterling WE, Mearns LO, Hays CJ, Marx D (2001) Comparison of agricultural impacts of climate change calculated from high and low resolution climate change scenarios: Part II. Accounting for adaption and CO2 direct effects. *Climate Change*, **51**, 173–197.

ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, CA.

FAO/IIASA (2007) Global agro-ecological zones. FAO/IIASA, Rome, IT.

FAO/IIASA/ISRIC/ISSCAS/JRC (2012) *Harmonized World Soil Database (version 1.2)*. FAO, Rome, IT and IIASA, Laxenburg, AT.

Fischer G, Schrattenholzer L (2001) Global bioenergy potentials through 2050. *Biomass and Bioenergy*, **20**, 151–159.

Forest Resources Assessment Program (FRA) of UN-FAO (2001) Global ecological zoning for the global forest resources assessment. FRA, Rome, IT.

Gassman PW, Williams JR, Benson VW *et al.* (2005) Historical development and applications of the EPIC and APEX models. CARD Working Paper 05-WP. Center for Agricultural and Rural Development, Iowa, US.

Heaton EA, Dohleman FG, Miguez AF *et al.* (2010) Miscanthus: a promising biomass crop. *Advances in Botanical Research*, **56**, 76–137.

Hoogwijk M, Faaij A, van den Broek R, Berndes G, Gielen D, Turkenburg W (2003) Exploration of the ranges of the global potential of biomass for energy. *Biomass and Bioenergy*, **25**, 119–133.

Intergovernmental Panel on Climate Change (IPCC) (2011) *IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation*. Prepared by Working Group III of the Intergovernmental Panel on Climate Change (eds Edenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Matschoss P, Kadner S, Zwickel T, Eickemeier P, Hansen G, Schlömer S, von Stechow C). Cambridge University Press, Cambridge, United Kingdom and New York, NY, US.

International Energy Agency (IEA) (2010) Energy technology perspectives: scenarios & strategies to 2050. OECD/IEA, Paris, France.

International Institute for Applied System Analysis (IIASA) (2010) Global earth observation – benefit estimation. Available at: http://ec.europa.eu/research/research-for-europe/environment-geo-bene_en.html (accessed 20 October 2012).

Izaurralde RC, Williams JR, McGill WB, Rosenberg NJ, Quiroga Jakas MC (2006) Simulating soil C dynamics with EPIC: model description and testing against long-term data. *Ecological Modeling*, **192**, 362–384.

Kiniry JR, Sanderson MA, Williams JR *et al.* (1996) Simulating Alamo switchgrass with the ALMANAC model. *Agronomy Journal*, **88**, 602–606.

Kiniry JR, Schmer MR, Vogel KP, Mitchell R (2008) Switchgrass biomass simulation at diverse sites in the northern Great Plains of the US. *Bioenergy Research*, **1**, 259–264.

Kiniry JR, Johnson M, Mitchell R, Vogel K, Kaiser J, Bruckerhoff S, Cordsiemon R (2011) Switchgrass leaf area index and light extinction coefficients. *Agronomy Journal*, **103**, 119–122.

Kline KL, Parish E, Singh N, Wullschleger S, Preston B, Keller M, Lynd LR (2011a) Collaborators welcome: Global Sustainable Bioenergy Project (GSB). *GLP News*, **7**, 7–8.

Kline KL, Oladosu GA, Dale VH, McBride AC (2011b) Scientific analysis is essential to assess biofuel policy effects: in response to the paper by Kim and Dale on "Indirect land use change for biofuels: testing predictions and improving analytical methodologies". *Biomass and Bioenergy*, **35**, 4488–4491.

Liu J, Fritz S, van Wesenbeeck CFA, Fuchs M, You L, Obserseiner M, Yang H (2008) A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change. *Global and Planetary Change*, **64**, 222–235.

Lobell DB, Schlenker W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science*, **29**, 616–620.

Lynd LR, Aziz RA, de Brito Cruz DH *et al.* (2011) A global conversation about energy from biomass: the continental conventions of the global sustainable bioenergy project. *Interface Focus*, **1**, 271–279.

Ma Y, An Y, Shui J, Sun Z (2011) Adaptability evaluation of switchgrass (*Panicum virgatum* L.) cultivars on the loess plateau of China. *Plant Science*, **181**, 638–643.

Mausbach JM, Dedrick AR (2004) The length we go: measuring environmental benefits of conservation practices in the CEAP. *Journal of Soil and Water Conservation*, **59**, 96.

Miguez FE, Maughan M, Bollero GA, Long SP (2012) Modeling spatial and dynamic variation in growth, yield, and yield stability of the bioenergy crops *Miscanthus × giganteus and Panicum virgatum* across the conterminous United States. *GCB Bioenergy*, **4**, 509–520.

Millennium Ecosystem Assessment (MEA) (2005) *Living Beyond Our Means: Natural Assets and Human Well-Being*. Island Press, Washington, DC.

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

Mueller ND, Gerber JS, Johnston M, Deepak KR, Ramankutty N, Foley JA (2012) Closing yield gaps through nutrient and water management. *Nature*, **490**, 254–257, doi: 10.1038/nature11420.

Nair SS, Kang S, Zhang X *et al.* (2012) Bioenergy crop models: description, data requirements, and future challenges. *GCB Bioenergy*, **4**, 620–633, doi: 10.1111/j.1757-1707.2012.01166.x.

Nichols JA, Kang S, Post W *et al.* (2011) HPC-EPIC for high resolution simulations of environmental and sustainability assessment. *Computers and Electronics in Agriculture*, **79**, 112–115.

Offermann R, Seidenberger T, Thran D, Kaltschmitt M, Zinoviev S, Miertus S (2011) Assessment of global bioenergy potentials. *Mitig. Adapt. Strateg. Global Change*, **16**, 103–115.

Parrish DJ, Fike JH (2005) The biology and agronomy of switchgrass for biofuels. *Critical Reviews in Plant Sciences*, **24**, 423–459.

Poccard L, Janicot S, Camberlin P (2000) Comparison of rainfall structure between NCEP/NCAR reanalyses and observed data in tropical Africa. *Climate Dynamics*, **16**, 897–915.

Potter SR, Atwood JD, Kellogg RL, Williams JR (2004) An approach for estimating soil carbon using the national nutrient loss database. *Environmental Management*, **33**, 496–506.

Ramankutty N, Evan A, Monfreda C, Foley JA (2008) Farming the planet. Part 1: the geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, **22**, 1–19, doi: 10.1029/2007GB002952.

Sanderson MA, Reed RL, McLaughlin SB *et al*. (1996) Switchgrass as a sustainable bioenergy crop. *Bioresource Technology*, **56**, 83–93.

Smeets EM, Faaij AP, Lewandowski IM, Turkenburg WC (2007) A bottom-up assessment and review of global bio-energy potentials to 2050. *Progress in Energy and Combustion Science*, **33**, 56–106.

Smith SR, Legler DM, Kathleen VV (2001) Quantifying uncertainties in NCEP Reanalyses using high-quality research vessel observations. *Journal of Climate*, **14**, 4062–4072.

Thomson AM, Rosenberg NJ, Izaurralde RC, Brown RA (2005) Climate change s for the conterminous USA: an integrated assessment. *Climate Change*, **69**, 27–41.

United States Geological Survey (USGS) (1996) *GTOPO30, Earth Resources Observation and Science (EROS) Center*, USGS, Fioux Falls, SD, USA.

USEPA (2010) Regulation of fuels and fuel additives: changes to renewable fuel standard program; 18 Final Rule, *Federal Register*, **75**, 40 (CFR Part 80).

Viovy N (2010) CRU-NCEP reanalysis data. Available at: http://www.cru.uea.ac.uk/cru/data/ncep/ (accessed 20 October 2012).

Wang D, Carr E, Gross LJ, Berry MW (2005) Towards ecosystem modeling on computing grids. *IEEE Computing in Science and Engineering*, **7**, 44–52.

Washington WM, Lawrence M, Anthony C (2009) The computational future for climate and earth system models: on the path to petaflop and beyond. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, **367**, 833–846.

Williams JR, Jones CA, Dyke PT (1984) A modeling approach to determining the relationship between erosion and soil productivity. *Transaction of the ASAE*, **27**, 129–144.

Wullschleger SD, Davis EB, Borsuk ME, Gunderson CA, Lynd LR (2010) Biomass production in switchgrass across the United States: database description and determinants of yield. *Agronomy Journal*, **102**, 1158–1168.

Xie A, Ren J, Qin X, Kang S (2007) Reliability of NCEP/NCAR reanalysis data in the Himalayas/Tibetan Plateau. *Journal of Geographical Science*, **17**, 421–430, doi: 10.1007/s11442-007-0421-2.

Zhang X, Izaurralde RC, Manowitz DM *et al*. (2010) An integrated modeling framework to evaluate the productivity and sustainability of biofuel crop production system. *GCB Bioenergy*, **2**, 258–277, doi: 10.1111/j.1757-1707.2010.01046.x.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1 Supporting Online Materials (SOM) for: Global Simulation of Bioenergy Crop Productivity: Analytical Framework and Case Study for Switchgrass.