Bioenergy crop models: descriptions, data requirements, and future challenges

SUKJITHKUMAR SURENDRAN NAIR*, SHUJIANG KANG†, XUESONG ZHANG‡, FERNANDO E. MIGUEZ§, R. CESAR IZAURRALDE‡, WILFRED M. POST†, MICHAEL C. DIETZE¶, LEE R. LYND|| and STAN D. WULLSCHLEGER†

* Agricultural and Resource Economics, University of Tennessee, 2621 Morgan Circle Drive, Knoxville, TN 37996, USA, †Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831 USA, ‡Joint Global Change Research Institute, Pacific Northwest National Laboratory/University of Maryland, 5825 University Research Court, Suite 3500 College Park, MD, 20740, USA, §Department of Agronomy, Iowa State University, 1206 Agronomy Hall, Ames, IA 50011, USA, ¶Department of Plant Biology, University of Illinois, Urbana, IL, 61801 USA, ||Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755-8000, USA

Abstract
Field studies that address the production of lignocellulosic biomass as a source of renewable energy provide critical data for the development of bioenergy crop models. A literature survey revealed that 14 models have been used for simulating bioenergy crops including herbaceous and woody bioenergy crops, and for crassulacean acid metabolism (CAM) crops. These models simulate field-scale production of biomass for switchgrass (ALMANAC, EPIC, and Agro-BGC), miscanthus (MISCANFOR, MISCANMOD, and WIMOVAC), sugarcane (APSIM, AUSCANE, and CANEGRO), and poplar and willow (SECRETS and 3PG). Two models are adaptations of dynamic global vegetation models and simulate biomass yields of miscanthus and sugarcane at regional scales (Agro-IBIS and LPJmL). Although it lacks the complexity of other bioenergy crop models, the environmental productivity index (EPI) is the only model used to estimate biomass production of CAM (Agave and Opuntia) plants. Except for the EPI model, all models include representations of leaf area dynamics, phenology, radiation interception and utilization, biomass production, and partitioning of biomass to roots and shoots. A few models simulate soil water, nutrient, and carbon cycle dynamics, making them especially useful for assessing the environmental consequences (e.g., erosion and nutrient losses) associated with the large-scale deployment of bioenergy crops. The rapid increase in use of models for energy crop simulation is encouraging; however, detailed information on the influence of climate, soils, and crop management practices on biomass production is scarce. Thus considerable work remains regarding the parameterization and validation of process-based models for bioenergy crops; generation and distribution of high-quality field data for model development and validation; and implementation of an integrated framework for efficient, high-resolution simulations of biomass production for use in planning sustainable bioenergy systems.

Keywords: biomass, climate change, crop models, data management, land use, productivity, sustainability

Received 28 December 2011 and accepted 27 January 2012

Introduction
Substitution of liquid transportation fuels derived from petroleum with a renewable source of bioethanol has prompted a worldwide interest in determining how much lignocellulosic biomass can be grown for the production of biofuels (McLaughlin et al., 2006; Ragauskas et al., 2006; Heaton et al., 2010). Current estimates of local and regional supplies of biomass, however, are limited by the availability of data that quantify harvestable yield for herbaceous and woody energy crops across a variety of site conditions. Fortunately, field trials promise to provide such information (Heaton et al., 2004; Aylott et al., 2008; Christian et al., 2008; Wang et al., 2010; Wullschleger et al., 2010) and therein improve our agronomic understanding of how soils, climate, genetics, and crop management practices like fertilization influence potential biomass production.

Correspondence: Stan D. Wullschleger, tel. + 865 574 7839, fax + 1 865 574 9501, e-mail: wullschlegsd@ornl.gov

© Published 2012
This article is a U.S. Government work and is in the public domain in the USA.
Insights gained from field trials will also be critical as we develop and evaluate bioenergy crop models (Miguez et al., 2009; VanLoocke et al., 2010; Cuadra et al., 2012). Kiniry et al. (1996) were among the first to apply crop models to the analysis of biomass production for switchgrass (Panicum virgatum) using the ALMANAC (Agricultural Land Management Alternatives with Numerical Assessment Criteria) model. Later, ALMANAC-based predictions for switchgrass were validated against field data for multiple sites across the Southern US (Kiniry et al., 2005). Since the initial adaptation of the ALMANAC model for switchgrass, there has been a surge in the adaptation of other existing models or the development of a few new models specific for perennial herbaceous and woody bioenergy crops including miscanthus (Clifton-Brown et al., 2000; Hastings et al., 2009; Miguez et al., 2009), poplar and willow (Lasch et al., 2010; Amichev et al., 2011), and sugarcane (Lisson et al., 2005; Thorburn et al., 2005; Bondeau et al., 2007; Lapola et al., 2009). These models have been used to forecast biomass yields at field to regional scales, and to associate production of biomass with possible environmental consequences including soil erosion and water quality.

Our objective in this article is to review crop models that have been developed or adapted for simulating bioenergy crops. The bioenergy crops considered for this study are herbaceous energy crops (switchgrass, miscanthus, and sugarcane [Saccharum officinarum] or energy cane [Saccharum spp.]), perennial woody crops (hybrid poplar [Populus spp.] and willow [Salix spp.]), and crassulacean acid metabolism (CAM) crops adapted to arid lands (Agave and Opuntia). Current models can be classified as either empirical or mechanistic. Empirical models use data from field trials to develop relationships between yield and independent climatic, soil, and crop management variables. In contrast, mechanistic models specifically describe underlying physiological and morphological processes that determine crop growth. Although a number of empirical models exist for bioenergy crops (Heaton et al., 2004; Aylott et al., 2008; Richter et al., 2008; Grassini et al., 2009; Jager et al., 2010; Wullschleger et al., 2010), we focus on more process-based models that simulate the production of biomass for important energy crops. We provide a description of each model; discuss approaches used to simulate crop growth, phenology, and water, carbon, and nitrogen dynamics; and consider how abiotic stresses are represented in these models. Special attention is given to how models describe dry matter production and distribution of dry matter to harvested yield. Finally, we highlight a unique bioenergy yield dataset that can be used in the calibration and validation of these models and comment on the future challenges likely to be encountered given the current state of modeling bioenergy crops.

Categories of bioenergy crop models

A literature survey revealed that 14 models have been used to simulate the production of biomass for biofuels, this includes models that are exclusively developed for bioenergy crops or adapted existing models for bioenergy crop (Table 1). Eleven models are used to estimate yield of herbaceous energy crops (Table 1). EPIC, which is able to simulate both herbaceous and woody crops such as switchgrass, miscanthus, sugarcane, and poplar is a process-based model capable of simulating a wide array of ecosystem processes including plant growth, crop yield, water and nutrient balances, and soil erosion (Williams et al., 1984). ALMANAC and AUSCANE models are related to EPIC in many ways: ALMANAC uses biophysical subroutines and process descriptions from EPIC with additional details for plant growth processes and is capable of simulating several crops including switchgrass (Kiniry et al., 1996), ALMANAC is also capable of simulating multiple species competing for light, nutrients, and water such as in native prairie mixes or with intercropping. AUSCANE is an adaptation of EPIC for simulation of sugarcane yield for Australian environments (Jones et al., 1989). MISCANMOD is a spreadsheet-based model that has been widely applied in Europe to predict biomass production of miscanthus (Clifton-Brown et al., 2000). MISCANFOR is an updated FORTRAN version of MISCANMOD with additional descriptions of soil water subroutines (Hastings et al., 2009). APSIM and CANEGRO are two sugarcane models (Lisson et al., 2005; Thorburn et al., 2005). WIMOVAC is a generic plant production model (Humphries & Long, 1995) that has recently been used for simulating biomass yield of miscanthus and switchgrass (Miguez et al., 2009, 2012). Agro-BGC is a variation of Biome-BGC, a well-known terrestrial biogeochemistry model with added processes to simulate biomass production of C4 herbaceous energy crops (Di Vittorio et al., 2010). LPjGML is an adaptation of Lund–Potsdam–Jena dynamic global vegetation model (DGVM) for managed lands that is capable of simulating several crops including sugarcane grown in rainfed or irrigated environments (Bondeau et al., 2007; Lapola et al., 2009). Likewise, Agro-IBIS is an agroecological version of the Integrated Biosphere Simulator model, also a DGVM (Kucharik, 2003) that has been used to simulate production of biomass by miscanthus (VanLoocke et al., 2010) and sugarcane (Cuadra et al., 2012).

SECRETS and 3PG are two models used for simulating woody perennial bioenergy crops (Table 1). SECRETS is a modular, process-based model developed
to originally simulate growth and development of mixed-species forests (Sampson & Ceulemans, 2000). SECRETS has recently been used to simulate the biomass production from aspen and poplar (Deckmyn et al., 2004; Lasch et al., 2010). The 3PG model is a process-based model that has been successfully applied to predicting forest productivity in plantations of fast-growing trees including poplar and willow (Landsberg & Waring, 1997; Amichev et al., 2011).

Although it lacks the complexity of other bioenergy crop models, and is better characterized as an empirical model, the environmental productivity index (EPI) is the only model used to estimate biomass production of CAM plants (Table 1). The EPI is calculated based on the philosophy that crop growth is constrained by several factors with multiplicative impacts (García de Cortázar & Nobel, 1990). In this approach, constraints on crop growth are quantified using simple environmental stress indices (e.g., temperature and water index), where each index indicates the fraction of maximal net CO₂ uptake expected based on the prevailing value of that environmental factor.

### Bioenergy crop models: general descriptions

Bioenergy crop models vary in complexity and in their approach to simulating crop growth and other processes across space and time (Table 1). Most of the models operate at a daily time step; however, 3PG operates at monthly time step (Landsberg & Waring, 1997) and some processes (e.g., carbon assimilation) in WIMOVAC (Humphries & Long, 1995) and SECRETS (Sampson & Ceulemans, 2000) operate at hourly or sub-hourly time steps. In Agro-BGC, most of the processes operate at a daily time-step, although some pools update at annual time steps (Thornton, 1998; Golinkoff, 2010). In general models with shorter time step includes a detailed description of a process or part of a process. However, time step is also linked to the approach followed to explain a single process or several processes in the model. For example, radiation use efficiency (RUE) based biomass growth simulation follow daily time step. But, biomass growth based on a more detailed process oriented approach for photosynthesis follow hourly/sub-hourly time step, which is data and computational intensive. Many of the models simulate biomass production for site or field-scale application. Agro-IBIS (Kucharik, 2003), LPJmL (Bondeau et al., 2007), and Agro-BGC (Di Vittorio et al., 2010) were developed, however, to operate at much larger spatial scales using grid-based simulation techniques.

A general crop growth routine is used in many of the models to represent biomass growth, but approaches vary across models. APSIM uses a generic cultivar-level crop sub-model (Wang et al., 2002), but EPIC and ALMANAC (Williams et al., 1989; Kiniry et al., 1992) use a more general species-based crop growth routine. EPIC and ALMANAC are especially flexible and are currently configured to simulate growth and development for over 100 plant species including all major agricultural crops, grasses, legumes, some trees, and several emerging bioenergy crops. In Agro-BGC (Di Vittorio et al., 2010; Golinkoff, 2010), Agro-IBIS (Kucharik, 2003), and LPJmL (Bondeau et al., 2007), a broader generalization

### Table 1  General characteristics of selected models

<table>
<thead>
<tr>
<th>Models</th>
<th>Scale</th>
<th>Sub-models</th>
<th>Bioenergy crops covered</th>
<th>Crop model</th>
<th>First reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herbaceous perennial grass</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPIC</td>
<td>Field</td>
<td>P, W, N, C</td>
<td>Switchgrass, Miscanthus</td>
<td>Generic, dynamic</td>
<td>Williams et al. (1984)</td>
</tr>
<tr>
<td>ALMANAC</td>
<td>Field</td>
<td>P, W, N</td>
<td>Switchgrass, Miscanthus</td>
<td>Generic, dynamic</td>
<td>Kiniry et al. (1992)</td>
</tr>
<tr>
<td>MISCANMOD</td>
<td>Field</td>
<td>P</td>
<td>Miscanthus</td>
<td>Crop specific</td>
<td>Clifton-Brown et al. (2000)</td>
</tr>
<tr>
<td>MISCANFOR</td>
<td>Field</td>
<td>P, W</td>
<td>Miscanthus</td>
<td>Crop genotype specific</td>
<td>Hastings et al. (2009)</td>
</tr>
<tr>
<td>Agro-BGC</td>
<td>Ecosystem</td>
<td>P, W, N</td>
<td>Switchgrass</td>
<td>Generic PFT, dynamic</td>
<td>Di Vittorio et al. (2010)</td>
</tr>
<tr>
<td>APSIM</td>
<td>Field</td>
<td>P, W, N</td>
<td>Sugarcane</td>
<td>Generic, dynamic</td>
<td>Keating et al. (1999)</td>
</tr>
<tr>
<td>AUSCANE</td>
<td>Field</td>
<td>P, W</td>
<td>Sugarcane</td>
<td>Specific, dynamic</td>
<td>Jones et al. (1989)</td>
</tr>
<tr>
<td>LPJmL</td>
<td>Ecosystem</td>
<td>P, W</td>
<td>Sugarcane</td>
<td>Generic CFT</td>
<td>Bondeau et al. (2007)</td>
</tr>
<tr>
<td>CANEGRO</td>
<td>Field</td>
<td>P, W</td>
<td>Sugarcane</td>
<td>Specific, dynamic</td>
<td>Inman-Bamber (1991)</td>
</tr>
<tr>
<td>Woody perennials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P, Plant growth; W, water; N, nitrogen; C, soil carbon dynamics (soil organic matter is included).

*Water and nutrient index.

© Published 2012
This article is a U.S. Government work and is in the public domain in the USA., GCB Bioenergy, doi: 10.1111/j.1757-1707.2012.01166.x
of vegetation is incorporated based on crop/plant functional types (CFT/PFT). WIMOVAC is a general vegetation model in which photosynthesis can be switched between C3 and C4, and individual crops need to be parameterized by adjusting parameters specific to photosynthesis, leaf area index, canopy architecture, and carbon allocation (Humphries & Long, 1995). Some models simulate other processes such as water, nitrogen, and carbon cycling in an agroecosystem. MISCANFOR (Hastings et al., 2009), LPjmL (Bondeau et al., 2007), and 3PG (Landsberg & Waring, 1997) explicitly represent crop response to soil water relationships. AUSCANE (Jones et al., 1989) and CANEGRO (Lisson et al., 2005) include a soil nitrogen component along with crop and soil water relationships. EPIC (Izaurralde et al., 2006), WIMOVAC (Long et al., 1998), Agro-IBIS (Kucharik et al., 2000), Agro-BGC (Golinkoff, 2010), APSIM (Probert et al., 1998), and SECRETS (Sampson et al., 2001) have a soil organic carbon submodel in addition to soil nitrogen and soil water components.

Given the complexity of processes represented, some of these models are especially useful for generating information to understand sustainability issues related to bioenergy production. EPIC in particular can be coupled with the Soil and Water Assessment Tool (SWAT) to quantify the impact of land management practices on hydrologic processes in large watersheds. EPIC and SWAT in combination simulate crop growth, soil water and groundwater movement, and transport of sediment and nutrients (Luo et al., 2008). In addition, the landscape and watershed version of EPIC, called APEX (Gassman et al., 2010), contains algorithms that make it possible to conduct environmental analyses at different spatial scales, from small to large watersheds with variable land cover or land use. APEX simulates routing of water, sediment, nutrients, and pesticides across complex landscapes and channel systems to the watershed outlet.

Specific crop growth subroutines

Phenology, leaf area dynamics, radiation interception and utilization, and crop growth and carbon partitioning are major processes that determine harvestable biomass. Thus, a comparison of these key physiological and morphological processes is warranted.

Phenological stages of development

In all the bioenergy crop models summarized here, phenological stages are linked in some manner to cumulative growing degree days (GDD). This measure of thermal period is computed as an average of daily maximum and minimum temperature above the base temperature of a specific crop (Williams et al., 1989). Some models explicitly simulate different phenological stages of bioenergy crops (e.g., flowering), while other models dispense with simulations of specific developmental stages and instead use developmental curves to capture important temporal dynamics (e.g., canopy leaf area as a function of GDD). APSIM, MISCANMOD, MISCANFOR, WIMOVAC, Agro-IBIS, Agro-BGC, 3PG, and SECRETS belong to the first category, and EPIC, ALMANAC, AUSCANE, and LPjml belong to the second category (Table 2). In some models, growing season length spans the last frost in the spring to the first frost in the fall (e.g., WIMOVAC), while in other models the sum of heat units from planting or tiller emergence to maturity is used to determine growing season duration (e.g., ALMANAC).

Leaf area dynamics

In general, leaf area development in crop growth models is simulated using two approaches – a functional approach and a mechanistic approach (Adam et al., 2011). The majority of bioenergy crop models follow a functional approach in simulating leaf area dynamics (Table 2). The form of that function, however, does vary across models. A linear function is used in MISCANMOD and MISCANFOR (Clifton-Brown et al., 2000; Hastings et al., 2009) and a sigmoid function is used in ALMANAC, EPIC, and LPjml to represent pre-senesence growth of leaf area index (LAI). A decline in LAI during the post-senescence period is represented by a power function in EPIC, ALMANAC, and LPjml (Kimiri et al., 1996; Bondeau et al., 2007; Williams et al., 2008).

In contrast to the functional approach, leaf area dynamics is calculated internally in a mechanistic approach by using total accumulated biomass, a crop-specific leaf partitioning coefficient, and developmental stage of the crop (Adam et al., 2011). A mechanistic approach is applied in crop models in two ways – specific leaf area method (SLA) and individual leaf area method (ILA) (Keating et al., 1999; Adam et al., 2011). In the SLA approach, leaf area is estimated by using dry mass of leaf and unit weight of leaf, whereas in the ILA approach leaf area is calculated as a function of leaves per stalk and unit leaf area. Some models treat SLA as a constant value (e.g., Agro-IBIS, Agro-BGC, WIMOVAC, SECRETS), while others treat SLA as a variable over developmental stages (e.g., 3PG) (Amichev et al., 2011). APSIM and CANEGRO simulate LAI using ILA (Inman-Bamber, 1994a, b; Keating et al., 1999) (Table 2). LAI in APSIM is calculated by simulating the green leaf area per stalk, number of leaves per stalk (maximum
of 13 leaves), and area of a single leaf, which is determined by the genotype coefficients of the crop (Keating et al., 1999). CANEGRO allows variation in number of leaves per stalk with GDD and cultivar (Inman-Bamber, 1994a, b).

### Biomass production and partitioning

In general, biomass production and partitioning in bioenergy crop models is represented as a three step process. In the first step, the processes involved in light interception and estimation of the amount of solar radiation captured by crop canopy are addressed. The second step converts intercepted light energy into biomass, and then in the third step, biomass is partitioned into different plant parts. In the bioenergy crop models described here, the representation of these three processes varies considerably.

### Light interception

Beer’s law (Monsi & Saeki, 1953), or some variation, is commonly used in bioenergy crop models to estimate the amount of light intercepted by a crop. However, representation of the light extinction coefficient (k) varies across models. As per Beer’s law for a given LAI, a higher value for k would result in a higher fraction of light intercepted by the crop. EPIC uses k value of 0.65 (Williams et al., 1989) while ALMANAC uses a species-specific value for k (Kiniry et al., 1996). The LPjmL model uses a value of 0.5 to estimate intercepted radiation (Bondeau et al., 2007). MISCANMOD and MISCANFOR use measured values of light interception and LAI to estimate k (i.e., 0.68) using Beer’s law (Clifton-Brown et al., 2000; Hastings et al., 2009). APSIM uses a fixed k value of 0.38 and CANEGRO accounts for tiller density while estimating the final light interception (Keating et al., 1999; Bezuidenhout et al., 2003). 3PG also employs Beer’s law to calculate interception using total incoming radiation and canopy leaf area index (Landsberg & Waring, 1997).

Some bioenergy crop models calculate k as a function of orientation and position of leaves, solar zenith angle, and soil albedo. WIMOVAC, Agro-IBIS, and SECRETS considered one or more of the above mentioned factors in modeling light interception by crop canopy (Sampson et al., 2001; Miguez et al., 2009; Cuadra et al., 2012). In WIMOVAC, an ellipsoid distribution of leaves is assumed, and thus a single-shape parameter is used to account for the position and orientation of leaves while estimating k (Norman, 1980; Miguez et al., 2009). Agro-IBIS calculates k using the functions of leaf orientation, transmittance, soil albedo, SLA, and LAI (Cuadra et al., 2012). In contrast, SECRETS uses a modified version of the Beer-Lambert method for calculating a light interception reduction factor by using solar zenith angle and
canopy gap fraction for randomly distributed and for clumped foliage (Sampson et al., 2001). WIMOVAC adopted a multilayer canopy formulation by dividing the canopy into 10 discrete layers and using a sunlit/shade procedure to account for interception of diffused and direct radiation separately (Miguez et al., 2009). SECRETS applied the ‘big leaf’ canopy concept along with the sunlit/shade approach to quantify interception of diffuse and direct radiation (Spitters et al., 1986; Sampson et al., 2001). In Agro-IBIS, a two-stream approximation method is used for quantifying the incidence of direct and diffuse radiation for both the visible and near-infrared wavelength bands (Foley et al., 1996; Kucharik et al., 2000).

**Biomass production.** Simulation of biomass in almost all bioenergy crop models can be grouped into three main approaches: (1) radiation use efficiency (RUE) approach, (2) photosynthesis and respiration (PR) approach, and (3) biochemical approach. RUE is a simple, robust, and straightforward approach for simulating crop biomass growth that can be implemented by directly linking measured incident radiation to total biomass produced over a crop growth period (Monteith, 1977). In the PR approach, an empirical description of both photosynthesis and respiration of the plant is included, while in the biochemical approach a mechanistic formulation of carbon uptake and assimilation is achieved by representing key biochemical processes of photosynthesis. Daily plant growth is estimated based on the RUE approach in EPIC, ALMANAC, and APSIM (Williams et al., 1989; Kiniry et al., 1992; Keating et al., 1999), the PR approach in CANEGRO and 3PG (Inman-Bamber & Thompson, 1989; Landsberg & Waring, 1997), and the biochemical approach in SECRETS, WIMOVAC, LPjmL, Agro-BGC, and Agro-IBIS (Humphries & Long, 1995; Sampson et al., 2001; Kucharik, 2003; Bondeau et al., 2007; Di Vittorio et al., 2010) (Table 2). However, there are variations among specific models within each broad category of approach.

Monteith (1977) provided a strong and convincing theoretical foundation for the RUE approach based on experimental evidence of a robust functional relationship between seasonal light interception and stress-free biomass production for several crops. RUE is defined as biomass produced per unit of intercepted photosynthetically active radiation (IPAR) (g MJ⁻¹). This definition has been considered as the most preferred concept for crop growth modeling because of its simplicity in factors needed for modeling and straightforward implementation. In MISCANMOD, RUE is considered as a fixed value of 3.3 g MJ⁻¹ (Clifton-Brown et al., 2000). MISCANFOR uses a variable RUE by reducing maximum potential RUE value of 3.9 g MJ⁻¹ to a lower value whenever crop experiences extremes of temperature, water, and nutrient stresses (Hastings et al., 2009). APSIM uses two different fixed values for RUE for ratoon (1.65 g MJ⁻¹) and plant crops (1.80 g MJ⁻¹) and also a RUE reduction factor under temperature, water, and nutrient stresses conditions: RUE is reduced if the mean daily temperature is below 15 °C or above 35 °C, and RUE = 0 (no biomass production) when mean temperature reaches 5 °C or 50 °C (Keating et al., 1999). In ALMANAC, RUE value of 3.9 g MJ⁻¹ is used, which is further reduced for each 1 kPa increase in vapor pressure deficit (VPD) above 1 kPa (Kiniry et al., 2008). It is important to note that in ALMANAC, RUE in later growth stages declines with a decline function similar to LAI (Kiniry et al., 1996). EPIC treats RUE as a function of VPD and atmospheric CO₂ concentration (Stockle et al., 1992).

In the PR–based approach, a separate function or set of fixed coefficients are used for representing photosynthesis and respiration. In general, the PR approach uses conversion efficiency or quantum efficiency for deriving gross photosynthesis per unit IPAR. CANEGRO and 3PG both use the PR approach to simulate biomass growth. The CANEGRO model calculates daily dry matter production using photosynthetically active radiation conversion efficiency (PARCE in g MJ⁻¹) and IPAR (MJ ha⁻¹) (Singels & Bezuidenhout, 2002). In 3PG, monthly potential biomass production is calculated from maximum possible canopy quantum efficiency, IPAR, and a constant respiration coefficient (Waring et al., 1998).

The biochemical approach is closely linked to the seminal work by Farquhar et al. (1980) on biochemical leaf photosynthesis model for C3 crops. The model was later extended by Collatz et al. (1991) to a comprehensive biochemical model for C3 plants (FVC). Based on this model, the potential rate of fixing carbon during CO₂ assimilation processes in C3 plants can be expressed by three limiting factors – light-limited photosynthesis, Rubisco-limited photosynthesis, and photosynthesis limited by transportation or utilization capacity of photosynthetic products. Collatz et al. (1992) adapted the FVC modeling approach to describe a coupled photosynthesis-stomatal conductance model for C4 plants. Thus, gross photosynthesis is calculated as a function of incident solar radiation, intercellular CO₂ partial pressure, and leaf temperature. WIMOVAC and Agro-IBIS use the Collatz et al. (1992) approach to calculate gross carbon assimilation (Kucharik, 2003; Miguez et al., 2009), whereas LPjmL follows an adapted version of the FVC method proposed by Haxeltine & Prentice (1996). In Agro-BGC, the sun/shade model proposed by De Pury & Farquhar (1997) with enzyme-driven bundle sheath CO₂ concentration (Chen et al., 1994) was used to...
simulate C4 photosynthesis (Di Vittorio et al., 2010). SECRETS uses a simpler version of sunlit/shade model proposed by De Pury & Farquhar (1997). In this approach, carboxylation capacity \( V_{\text{max}} \) and potential electron transport capacity \( I_{\text{max}} \) are represented as input variables in the model and assumed as constants for all canopy foliage.

In models that use a biochemical approach to estimate photosynthesis, autotrophic respiration consists of two components, growth respiration, which is a fixed fraction of carbon produced, and maintenance respiration of carbon produced, and maintenance respiration is a fixed fraction of potential energy production, and evaporation by soil) are explored in this section. Bioenergy crop models used very simple to detailed descriptions of crop water demand and soil water availability (Table 3).

Soil water, carbon, and nitrogen dynamics

Soil water components

Soil water availability is a critical parameter in determining biomass yield of a crop. Therefore, representation of major water balance components (canopy interception, runoff, transpiration by vegetation, and evaporation by soil) is explored in this section. Bioenergy crop models used very simple to detailed descriptions of crop water demand and soil water availability (Table 3).

Canopy interception. Many bioenergy crop models consider canopy interception (CI) of rainfall in their respective hydrology sub-models but follow a range of approaches. 3PG treats CI as a fixed fraction of the rainfall (Sands & Landsberg, 2002), whereas in Agro-IBIS, APSIM, and CANEGRO, CI is estimated as a function of LAI. Agro-BGC calculates CI using a user-defined canopy interception coefficient and LAI (Golinkoff, 2010). EPIC estimates CI using LAI, aboveground production, and a maximum CI per unit rainfall event while calculating CI (Williams et al., 2008). LPJmL adopted an approach proposed by Kergoat (1998) in which CI is treated as a fraction of potential evapotranspiration (ET) (Gerten et al., 2004). The SECRETS model calculates stem and foliage components of CI separately while accounting for total interception. In addition, SECRETS also assumes an instant evaporation from rainfall intercepted by foliage (Meiresonne et al., 2003).

Runoff. Net precipitation \( P_{\text{CN}} = P - \text{CI} \) is used by many models for calculating runoff volume. In WIMOVAC, Agro-BGC, and 3PG, the runoff process starts when the soil is saturated with water. In contrast, EPIC, ALMANAC, APSIM, and CANEGRO use a modified version of the Soil Conservation Service’s (SCS) curve number (CN) approach to simulate runoff (USDA-SCS, 1972). Agro-IBIS simulates surface runoff by following the instantaneous water-retaining capacity concept in the upper soil layer including soil surface, sum of infiltration rate in upper soil layer, and maximum surface puddle depth (Soylu et al., 2010). In LPJmL, runoff is derived from three components, excess water over field capacity in upper and lower soil layers, and water percolating through the lower soil layer.

Evapotranspiration. Most of the selected bioenergy crop models rely on some variant of the Penman-Monteith
method for calculating evapotranspiration (ET), but this varies widely across the models. In EPIC, potential evapotranspiration ($ET_p$) is calculated based on a modified Penman-Monteith method taking into account atmospheric CO$_2$ levels (Stockle et al., 1992). Evaporation from soil and transpiration from a plant are separately calculated in EPIC from $ET_p$ and LAI using the method of Ritchie (1972). EPIC calculates total soil water evaporation (after snow and water collected in litter storage is evaporated) from the depth distribution of water within the soil profile (Williams et al., 2008). Agro-BGC uses a modified Penman-Monteith equation for estimating evaporation of canopy intercepted water, transpiration during photosynthesis, and soil evaporation (Waring & Running, 2007). In contrast to the widespread use of the Penman-Monteith approach to estimating ET, Agro-IBIS uses an approach proposed by Pollard & Thompson (1995) for calculating evapotranspiration, while MISCANFOR uses the Thornthwaite equation (Hastings et al., 2009) and LPJmL uses an approach proposed by Prentice et al. (1993).

In APSIM, actual evaporation estimation is adjusted for plant residues and growing plants on the soil surface, and transpiration demand is modeled as a function of the crop growth rate and transpiration-use efficiency (Monteith, 1986; Sinclair, 1986). $ET_p$ is partitioned into soil and plant components in CANEGRO using LAI and information on soil moisture. In WIMOVAC, transpiration is linked to soil water content and uses information about soil water content, user-provided critical soil water content, and soil water at wilting point to calculate actual ET$_p$ with variations in soil water (Humphries & Long, 1995). Transpiration in 3PG is calculated using the Penman-Monteith equation with an adjustment factor for canopy conductance (Dye & Olbrich, 1993; Leuning, 1995; Landsberg & Waring, 1997). Transpiration in Agro-IBIS is calculated separately for upper and lower canopy leaves using the model proposed by Pollard & Thompson (1995) (Foley et al., 1996; Kucharik et al., 2000). LPJmL uses the approach by Federer (1982) to simulate transpiration where a minimum of two supply and demand functions determines rates of plant transpiration (Gerten et al., 2004).
Soil carbon and nitrogen components

Some of the selected bioenergy crop models such as EPIC (Izaurralde et al., 2006), WIMOVAC (Long et al., 1998), Agro-IBIS (Kucharik et al., 2000), APSIM (Probert et al., 1998), LPjmL (Sitch et al., 2003), Agro-BGC, and SECRETS (Sampson et al., 2001) are capable of modeling soil organic matter (SOM) dynamics (Table 3). Many of the bioenergy crop models have adopted or adapted SOM and soil nitrogen modules from the CENTURY model (Parton et al., 1988). SECRETS uses routines of the GRASSLAND DYNAMICS model for simulating SOM (Sampson et al., 2001). In both approaches, SOM is categorized into different types/pools based on the rate of decomposition. In EPIC, SOM is divided into three pools such as microbial biomass and slow and passive pools, with turnover time increasing from microbial biomass to passive pool (Izaurralde et al., 2006). WIMOVAC also uses three SOM pools: active (microbial biomass and microbial products), slow, and passive pools (Long et al., 1998). In Agro-IBIS, surface plant material is categorized into three residue types: decomposable plant matter, structural plant matter, and resistant plant matter (Kucharik et al., 2000), and the belowground soil carbon is divided into four pools: active (microbial biomass), protected, unprotected, and stabilized pools. Specific microbial efficiencies are associated with the transformation of organic carbon into different SOM pools. Agro-BGC has four soil carbon pools, (e.g., fast, medium, slow, and recalcitrant). In addition, there are three litter pools and a coarse woody debris pool. Dead coarse roots and stems enter into a coarse woody debris pool, which then enter into the various litter pools over time depending on soil moisture and temperature. These litter pools decompose and enter into the SOM pools. SOM decomposition is constrained by soil water and temperature (Thornton, 1998; Golinkoff, 2010). In APSIM, single fresh organic matter pool is used to represent surface plant materials, and belowground SOM is divided into three different pools: BIOM (microbial biomass), HUM (other belowground organic matter), and INERT (part of HUM) pools (Probert et al., 1998; Thorburn et al., 2005). In LPjmL, SOM dynamics is described through one litter pool and two SOM pools, such as intermediate and slow pools (Sitch et al., 2003). Surface/soil litter is partitioned into metabolic, cellulosic, and lignin pools in SECRETS (Sampson et al. 2001; Thornley, 1998). Depending on the soil clay content, litter with high lignin content is further divided into protected and unprotected SOM pools, and unprotected SOM is transformed into stabilized SOM based on nitrogen concentrations in the mineral pools. Loss of soil carbon during transformation across different pools is also addressed in the model (Thornley, 1998). APSIM, Agro-BGC, and LPjmL allow gaseous loss of carbon, while Agro-IBIS, WIMOVAC, and SECRETS account for the gaseous and leaching losses. EPIC considers erosion loss of soil carbon along with gaseous and leaching loss (Probert et al., 1998; Sitch et al., 2003; Golinkoff, 2010). The model EPIC, ALMANAC, Agro-BGC, AGRO-IBIS, APSIM, and SECRETS represent relevant soil nitrogen processes (Table 3).

Abiotic stresses

The impact and accurate representation of environmental stresses on various plant growth and developmental processes are critical components that need to be addressed in a crop model. EPIC and ALMANAC include stresses due to water, temperature, nutrient, and aeration on bioenergy crop growth (Table 4). For example, water stress is calculated as a ratio of the daily water uptake to the daily potential transpiration. Thus, the value of stress factors ranges from 0–1, with 0 indicating complete stress and 1 indicating no stress. Similarly, EPIC uses a ratio-based approach to derive stress factor values for temperature, nutrient availability, and aeration. The daily growth limiting factor (REG) is determined as the minimum of these stress factors, and daily potential LAI and biomass production are adjusted using REG (Williams et al., 2008). In the CANEGRO model, water stress is quantified using a soil-water stress factor (Singels & Bezuidenhout, 2002), which is similar to EPIC’s water stress function. APSIM follows a similar approach but uses two soil-water deficit factors, one that reduces RUE and another that reduces the rate of daily leaf expansion (Keating et al., 1999). In both CANEGRO and APSIM, carbon assimilation and leaf expansion are affected when leaf nitrogen concentration is below a critical level (Lisson et al., 2005). In Agro-IBIS, leaf temperature and moisture stress functions are applied to modify the gross primary production (GPP) and stomatal conductance. A soil nitrogen stress function is used to modify the crop parameters to account for the impact of nitrogen availability. Additionally, temperature, water stress, and nitrogen stress determine the leaf respiration rate (Kucharik & Byre, 2003). In Agro-BGC, water stress is implemented via the influence of leaf water potential and vapor pressure deficit on stomatal conductance (Mu et al., 2007). WIMOVAC uses a simple linear function to reduce stomatal conductance under water stress condition and adjusts growth stage-specific biomass partitioning factors with changes in water stress conditions. WIMOVAC uses a simple empirical water stress response function to monitor changes in average daily plant water potential against the pre-fixed growth stage-specific threshold water potential (Long et al., 1998). Whenever the average daily plant water potential is
below the corresponding threshold value, more biomass is allocated to the root by changing the biomass partition coefficient. In MISCANFOR, RUE is reduced by water, temperature, and nutrient stresses (Hastings et al., 2009). In this model, the value of the water stress factor is increased from 0 to 1 as the soil water content changes from wilting point to field capacity. Additionally, a temperature variation factor is applied to account for the impact of temperature impact on RUE. Potential biomass production in 3PG model is modified by environmental factors such as temperature, soil water, vapor pressure deficit, and nutrition. Each factor is calculated as a fractional value ranging between 0 and 1, which is then multiplied with potential biomass production to calculate actual biomass production (Landsberg & Waring, 1997).

### Databases for calibration and validation of bioenergy crop models

The Biofuel Ecophysiological Trait and Yield Database (BETY-db), which is maintained at the University of Illinois, was created in order to compile the available field data on ‘second-generation’ biofuel crops and provide information on the productivity and ‘trait’ information of different species and cultivars at different sites. Globally, the database currently contains 3950 yield observations and 20,896 observations on plant traits and ecosystems services. These data were extracted from 455 publications covering 647 study sites. Uncertainty estimates and sample size are attached to observations when they are present. The database is fully searchable by species, trait, and geographic location using an intuitive Google Maps based interface, and also provides interactive maps of model-based yield estimates for the conterminous US for Miscanthus, switchgrass, and hybrid poplar. Yield data are focused on temperate perennial grasses (*Panicum n = 1897, Miscanthus n = 624, Poa n = 209*), tropical canes (*Saccharum n = 244*), and temperate trees (*Populus n = 509, Salix n = 288*) with most observations from North America and Europe, but a few are from other continents as well. Detailed information on treatments (e.g., different levels of N addition) and crop management operations (e.g., dates of planting and harvest) are also available. Trait and ecosystem service data span a broad range of properties commonly used in ecosystem models for calibration and validation, such as photosynthetic parameters, leaf mass per unit area, LAI, and tissue-specific stoichiometries, turnover times, and respiration rates. Trait data also encompass a broader array of species (39 genera have over 50 observations) and biomes. Ongoing development within the database is focused on expanding functionality to serve as a recognized public repository for biofuel data that are accepted by funding agencies and journals requiring data deposition and data management plans. More information about BETY-db can be found at [https://ebi-forecast.igb.illinois.edu/bety/](https://ebi-forecast.igb.illinois.edu/bety/).

### Needs, opportunities, and future challenges of bioenergy crop simulations

Similar to other process-based models, bioenergy crop models such as those summarized here require quality input data (Table 5) and data for parameterization,
<table>
<thead>
<tr>
<th>Model</th>
<th>Input data</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPIC</td>
<td>Daily minimum and maximum temperature, precipitation, solar radiation, maximum rooting depth, heat units to maturity, base temperature, radiation use efficiency, and crop management practices.</td>
</tr>
<tr>
<td>ALMANAC</td>
<td>Daily minimum and maximum temperature, precipitation, solar radiation, maximum leaf area index, maximum rooting depth, heat units to maturity, base temperature, and crop management practices.</td>
</tr>
<tr>
<td>MISCANMOD</td>
<td>Daily or monthly temperatures and precipitation, solar radiation, important soil properties (Soil moisture holding capacity and plant available water), radiation use efficiency, and cumulative degree days for ending crop season.</td>
</tr>
<tr>
<td>MISCANFOR</td>
<td>Daily or monthly mean temperature, precipitation, solar radiation, soil data also required including soil water holding capacity, clay content, wilting point, field capacity, and bulk density. Radiation use efficiency, leaf expansion index and base temperature, length of growing season for photosynthesis expressed in degree days.</td>
</tr>
<tr>
<td>WIMOVAC</td>
<td>Temperature, solar radiation, relative humidity, wind speed, precipitation. If daily data are available, hourly data can be generated internally, maximum rate of carboxylation, quantum efficiency, dark respiration. Dry matter partitioning coefficients, thermal periods in degree days for six growth stages; soil properties (maximum rooting depth, field capacity, wilting point, etc.) and photosynthetic parameters including quantum efficiency maximum rate of assimilation, and dark respiration.</td>
</tr>
<tr>
<td>Agro-IBIS</td>
<td>Temperature, solar radiation, relative humidity, wind speed, and precipitation. If daily data are available, hourly data can be generated internally, maximum rate of carboxylation, quantum efficiency, dark respiration, soil properties, initial carbon pools, and management.</td>
</tr>
<tr>
<td>Agro-BGC</td>
<td>Requires up to 54 static vegetation parameters, nine location and soil parameters, daily climate data, and annual atmospheric CO2 concentrations and nitrogen deposition and fixation inputs. The six daily climate variables required to run Agro-BGC are maximum and minimum temperature, precipitation, vapor pressure deficit, net downward shortwave radiation, and day length.</td>
</tr>
<tr>
<td>APSIM</td>
<td>Temperature, solar radiation, relative humidity, wind speed, precipitation, radiation use efficiency, leaf area index, and thermal time in degree days, soil depth, water holding capacity, and nitrogen status, crop management practices.</td>
</tr>
<tr>
<td>AUSCANE</td>
<td>Maximum and minimum daily air temperature, leaf area index, biomass partitioning, harvest index, soil albedo, bulk density, texture, nutrient status, organic carbon content, and crop management practices.</td>
</tr>
<tr>
<td>LPJmL</td>
<td>Monthly data for mean temperature, precipitation, number of wet days, and sunshine hours. Soil texture, atmospheric CO2 concentration, and management practices are also required.</td>
</tr>
<tr>
<td>CANEGRO</td>
<td>Maximum and minimum temperature, rainfall, solar radiation, maximum relative humidity, wind speed, dew point temperature, soil water parameter (drained upper limit, lower limit, saturated water capacity, root distribution weighting) cultivars (23 parameters), crop management practices.</td>
</tr>
<tr>
<td>3PG</td>
<td>Daily or monthly minimum and maximum air temperature, monthly rainfall, number of rain and frost days per month, and monthly average day-time vapor pressure deficit. Initial biomass in foliage, stems, and roots, estimates of quantum-use efficiency, soil type parameters.</td>
</tr>
<tr>
<td>SECRETS</td>
<td>Maximum rates of Rubisco carboxylation and potential electron transport rate are required. Coefficients are necessary to determine carbon allocation to shoot and roots. Site data on soil texture and rooting depth must be specified.</td>
</tr>
<tr>
<td>EPI</td>
<td>Daily global radiation, percent of maximum sunshine hours, maximum and minimum nighttime temperature, maximum and minimum relative humidity, and total rainfall.</td>
</tr>
</tbody>
</table>

Validation, and uncertainty quantification. Research on bioenergy crops began only recently and thus compared with other traditional crops, detailed agronomic information of growth, development, and management of bioenergy crops is scarce. This limits effective parameterization, model improvement, and efficient application of bioenergy crop models (Thomson et al., 2009; Heaton et al., 2010). Fortunately, several databases are being developed and this will ensure the summarization and sharing of quality data for model evaluation and model inter-comparisons.

An integrated framework for the efficient execution of bioenergy crop models would be useful as the community moves from site to larger regional to continental scale simulations. A computational framework is especially important for high-resolution modeling, starting with preparation of spatially explicit input data, execution of model runs, analysis of results, and visualization. Such a framework should have (1) a geographic information system (GIS) for the preprocessing of spatial datasets, (2) an efficient computational platform for high-performance simulations, and (3) powerful post-processing and analysis of model output. Detailed information about local geographic features and spatial patterns of land use/land cover, soil, topography, and climate data, which is critical for accurate assessment of sustainability of biomass cultivation (Hellmann & Verburg, 2011), can be efficiently processed in support of...
high-resolution biomass simulations using GIS software (ArcGIS by ESRI, Redlands, CA, USA). While numerous examples are available that show large scale, spatial simulations of biomass yields and thus the utility of GIS, Zhang et al. (2010) used GIS to facilitate preprocessing of terabytes of input data, define homogeneous spatial modeling units, and extract input information for use in their biophysical and biogeochemical model (e.g., EPIC) at regional scales. In addition, the spatially explicit integrative modeling framework (SEIMF) developed by Zhang et al. (2010) contains components for (1) importing millions of text files generated by EPIC into PostgreSQL relational database for online sharing, post analysis, and visualization, and (2) optimizing spatial configurations of biofuel cropping systems by simultaneously considering multiple, often conflicting, objectives (e.g., productivity, nitrogen leaching, and GHGs emission). Such an approach coupled with the high-resolution simulations of bioenergy crop yields placed a heavy computational burden on their analysis. In this context, high-performance computing systems have been used to address this challenge. Nichols et al. (2011) recently described a high performance computing (HPC)-EPIC application capable of executing in parallel the millions of simulations required for high-resolution regional studies. By using 32 CPUs on an SGI Altix cluster system, the parallel computing capacity of HPC-EPIC has been used to address this challenge. Nichols et al. (2011) recently described a high performance computing (HPC)-EPIC application capable of executing in parallel the millions of simulations required for high-resolution regional studies. By using 32 CPUs on an SGI Altix cluster system, the parallel computing capacity of HPC-EPIC allows reducing the total execution wall time from 300 to 8 h for an EPIC execution problem with a total of 1 048 358 simulations (Nichols et al., 2011). Quantification of uncertainty of bioenergy crop simulations is another challenge as it is seldom systemically conducted for most crop modeling studies. Like other process-based models, uncertainty in bioenergy crop modeling is typically from three major sources – input uncertainty, model uncertainty (structure and parameter), and observation uncertainty. In the presence of input and structural uncertainty, traditional calibration, and validation procedures would fail to guarantee reliable parameter estimation. Some of these uncertainties may be propagated during simulation. Therefore, a comprehensive uncertainty analysis framework, including examination of uncertainties associated with observed data, model structure, and model parameters needs to be established to analyze and understand current modeling variability and limitations.

Conclusions

Fourteen models used to simulate herbaceous and woody bioenergy crops, as well as crops with CAM metabolism were reviewed. These models vary in their degree of sophistication. Field trials that address the influence of genetic, environmental, and crop management on biomass production will provide valuable data for the development and calibration of bioenergy crop models. Field data are, however, available for only a few countries and a few bioenergy crops, which limits model validation and application. Nonetheless, new energy crop models continue to be published (Cuadra et al., 2012; Lee et al., 2012) thus documenting widespread interest in this area. Future research should address the development of (1) specific models for emerging bioenergy crops (e.g., energy cane and CAM crops), (2) platforms that facilitate acquisition and sharing of high-quality field experimental data for model development and testing, and (3) an integrated framework for efficient execution of large-scale simulations and processing of input and output data. Advances in these areas will enable the scientific community to further evaluate sustainable bioenergy production systems.

Acknowledgements

This research was sponsored by the US Department of Energy (DOE), Office of Science, Biological and Environmental Research (BER) program. Oak Ridge National Laboratory is managed by UT- Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725, and was carried out in coordination with the Global Sustainable Bioenergy Project. Individuals working on this project also received support from the DOE Great Lakes Bioenergy Research Center and the Energy Biosciences Institute, and the DOE Office of Biomass Program Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830).

References


