

## Chapter 20

### Modeling Tools and Strategies for Developing Sustainable Feedstock Supplies

Presenting author and affiliation: Fernando E. Miguez, Department of Agronomy, Iowa State University, Ames, IA 50014

Presenter Email: femiguez@iastate.edu

Additional author(s) and affiliation(s)

Michael Dietze, Department of Plant Biology, University of Illinois at Urbana-Champaign, IL 61801

Armen R. Kemanian, Department of Crop and Soil Sciences, The Pennsylvania State University, State College, PA

#### Introduction

Assessing the feasibility of developing a strong biofuel industry around biomass feedstock requires a comprehensive evaluation of agronomic, environmental, social and economic factors. An encompassing assessment of the sustainability of biomass production as a feedstock for a developing bioenergy sector is complex due to the multiple dimensions involved in a complete evaluation of its social, technological and economic factors. The current trend of rising fossil fuel prices and observed climate change, and other adverse environmental and societal impacts of energy use make the exploration for more sustainable ways to use energy more important than ever (Kowalski et al., 2009). According to Hill et al. (2006) for biofuels derived from crops to be a viable alternative they should:

- provide a net energy gain,
- have environmental benefits,
- be economically competitive,
- be producible in large quantities and
- do not reduce food supplies.

Incorporating these multiple objectives into a single framework is challenging and requires tools and strategies to support decisions of stakeholders and policymakers. A fundamental component of such comprehensive assessments is the evaluation of the potential and attainable productivity of biofuel crops in different locations and growing conditions. Acquiring this type of information through field experimentation in herbaceous and woody crops, as well as in native forests and grasslands, is both expensive and time consuming, as it can take years of field trials to provide accurate estimates of potential production. An alternative science-based approach to estimate bioenergy crops productivity is to use biophysical or empirical simulation models. These models can provide estimates of average productivity and its inter-annual variability based on soil, weather, and bioenergy crops management databases that serve as inputs to the model.

To some extent the future of biofuels depends on technological breakthroughs which are difficult to predict, as technological advances might give an edge to particular renewable energy alternatives. Nonetheless, the current understanding is that transportation will continue to rely on liquid fuels in the coming decades and that a fraction of the liquid fuel supply will be based on oil, starch, and in particular ligno-cellulosic crops (Richard, 2010). Establishing a large scale biofuel industry requires a careful assessment of resources, logistic capabilities, and potential bottlenecks in the production chain before large investments are deployed in the field. Crops might play an important role supplying the feedstock for this demand of transportation fuels. Some of the more pressing questions are: Which crops to grow, where, and how to grow them? Also, what are the local and global consequences of growing crops for biofuel?

Approaching these questions can benefit greatly from modeling tools such as databases, computer simulation models and novel statistical approaches to integrate data and model inputs and outputs. Historically, crop research has focused on increasing seed yields of cereal and oilseed crops and much less attention has been given to improving yields of crops for total biomass. Recent interest in biomass crops has spurred research in developing annual grasses (e.g. sorghum), perennial rhizomatous grasses (e.g. switchgrass, *Miscanthus*, sugarcane, *Spartina*) and woody (e.g. willow, poplar) feedstocks that can be converted to liquid fuels using cellulose as the main substrate (Perlack et al., 2005). In this chapter we will briefly review some of the candidate feedstocks for which our modeling efforts are relevant, describe data requirements (databases), biophysical models, and statistical tools to connect data and models and assess model performance.

### Food-Based Biofuels

Currently, food crops are the main source of feedstock for biofuel. Grain maize is the main source of ethanol used mostly as an additive to conventional gasoline. However, it has been criticized mainly for competing with food production and having a low conversion efficiency to ethanol. This low conversion efficiency is in part a result of the large amounts of nitrogen (N) fertilizer needed to achieve high yields (Shapouri et al. 2002). Soybean oil is used for the production of biodiesel which seems to have a more favorable conversion efficiency and emissions reduction than ethanol production from maize grain (Hill et al., 2006). In addition to being food crops and having relatively low conversion efficiencies, the conversion of all U.S. maize grain and soybean oil into biofuels would only contribute to 12% and 6% of the U.S. gasoline and diesel demands, respectively, having even in that extreme case a low impact in the development of a significant alternative renewable energy (Hill et al., 2006).

### Perennial Grasses

Perennial rhizomatous grasses have been put forward as dedicated biomass crops because of their many benefits which include high productivity, high water and nutrient use efficiency, nutrient recycling, long canopy duration and reduced agronomic inputs (e.g. fertilization and tillage) (Heaton et al., 2004b). These characteristics make them more suited for sustainable production of biomass than traditional crops grown for food production. Some of the species with great potential as biomass producers are: switchgrass (*Panicum virgatum*), *Miscanthus × giganteus*, and energycane (sugarcane bred for biomass production) (Somerville et al., 2010). Sugarcane is currently successfully used in Brazil for the production of ethanol (Nass et al., 2007) but there are concerns about its sustainability and the impact on deforestation of the Amazon and the Cerrado regions (Sawyer, 2008).

### Woody Biomass

Worldwide 75% of current biofuel use is derived from wood and wood by-products (Food and Agriculture Organization (FAO), 2007). In many ways woody biomass is the oldest biofuel, having been burned directly or converted to charcoal for millennia. In more industrialized settings woody biomass is also utilized as a solid fuel for both on-site energy generation using from industrial waste (e.g. at sawmills and pulp plants) and in larger scale “cogeneration” electrical plants that use a mix of wood and fossil fuels.

The use of wood as a liquid biofuel is currently limited, yet wood has advantages as feedstock for cellulosic ethanol production due to its higher density than grass crops which can lead to greater transportation efficiency. Woody biofuels are also less sensitive to harvest time, potentially allowing a more stable fuel production that would buffer both the annual cycle of crop harvests and the inter-annual variability in crop yields. Worldwide there are large areas of marginal agricultural land that has been abandoned and allowed to regrow as forest. There are also large afforested areas where markets may favor liquid fuel production. Existing native and plantation forests could both be harvested directly for biofuel production and either regrown under their current land-use or converted to short-rotation coppice forestry. Coppice forestry is based on frequent harvesting and rapid regeneration by stump re-sprouting. Most research has focused on hybrid varieties of poplar (*Populus*) and willow (*Salix*) that have been selected for rapid regeneration. A survey of the scientific literature across all climates and clones suggests that poplar and willow can deliver mean annual yields in the range of 7.5 and 8.9 Mg ha<sup>-1</sup> respectively with maximum reported annual yields of 40 and 38 Mg ha<sup>-1</sup> respectively (Wang and Dietze *unpublished data*).

### Biophysical Models

Computer simulation models play a critical role in the evaluation of potential biofuel crops. Unlike first generation biofuel crops, such as maize and soybean, which have been planted over large areas for many decades, most second generation crops have only been evaluated in a handful of field trials and in a comparatively short time span. This leads to a number of questions about how different crops will yield in different areas and what the long-term impacts on ecosystem services will be that can only be answered through the use of models.

Process-based simulation models are a cost-effective tool to assess the productivity and environmental benefit or impact of biofuel, forage, grain, and other mixed production systems. The successful application of these models requires a correct parameterization of crop, soil, and landscape properties, as well as a well defined initialization procedure. The quantification of the uncertainties associated with model-based extrapolation can be complex, and requires careful attention and interpretation. Models vary in the detail with which crop, soil and landscape-scale processes are treated and in the fundamental principles driving mass and energy flux in the system. These differences are briefly discussed for biomass accretion and nutrient cycling in the soil.

### Biomass Accretion

There are two approaches used to simulate crop processes in cropping and ecosystem simulation models. Some modeling systems use a generic vegetation model (e.g. APEX-EPIC, C-Farm, CropSyst, DayCent, Ecosys, WIMOVAC), while others use a species-based model (e.g. APSIM, DSSAT). In the former a common framework is used to simulate all processes and different species or cultivars are represented by variations in the parameters. This confers substantial advantages in terms of algorithm development and re-use of code at run time, while facilitating the data collection for calibration and testing of the model. In the species-based approach, a different model is developed for each species and the parameters adjusted for each cultivar ("genetic coefficients").

Another dimension in which vegetation models vary is in the treatment of plant and population properties, with some models simulating growth and development of an individual plant (some species in the APSIM and DSSAT) and other simulating these processes on a unit-area basis (most models). Most models mentioned in this chapter use a "top-down" approach for modeling crop processes, which means that the underlying mechanisms are modeled only one or two levels of resolution "below" the response variable of interest. The appropriateness of each approach is more related to the objective in the model application than with the approach itself. Large-scale or country-wide simulations that respond to climate and soil variables are likely more robust based on generic crop models (e.g. applications of EPIC in the Conservation Effects Assessment Project) while system biology studies may require a greater level of de-aggregation of physiological processes. The number of parameters of a model grows dramatically as the level of resolution increase, making the calibration difficult.

The algorithms to simulate growth vary for different models. Some models used a detailed, multi-layered canopy approach in which photosynthesis is simulated at multiple heights through the plant canopy on a sub-daily basis (typically hourly) and aggregated for the entire canopy (e.g. WIMOVAC). Some models fully couple photosynthesis, transpiration, and the other component of the energy balance (Grant, 1995; Kremer et al., 2008), while others simulate these processes somewhat independently (Sadras et al., 2005).

One approach that has been used in models to predict biomass production (Clifton-brown et al., 2004) to simulate and analyze crop growth is to express biomass accumulation as the product of a resource captured and the efficiency with which it is converted biomass. When the resources are radiation, water, or nutrients in general, the expression can be formalized as follows:

$$B = RUE \times f_{is} \times S_t$$

$$B = TUE \times f_{is} \times ET$$

$$B = X_n UE \times X_n$$

where  $B$  is biomass produced ( $\text{g m}^{-2}$ ),  $RUE$  is the radiation-use efficiency which is a crop / cultivar specific parameter ( $\text{g MJ}^{-1}$ ),  $f_{is}$  is the fraction of the incident solar radiation intercepted by the canopy,  $S_t$  is total incoming solar radiation ( $\text{MJ m}^{-2}$ ) in a given time interval,  $TUE$  is transpiration use efficiency ( $\text{g B kg}^{-1} \text{H}_2\text{O}$ ),  $ET$  is the evapotranspiration,  $f_{is}$  is the fraction of  $ET$  which is crop transpiration ( $\text{kg H}_2\text{O m}^{-2}$ ), and  $X_n UE$  is the use efficiency ( $\text{kg B kg}^{-1} X_n$ ) of nutrient  $X_n$  ( $\text{kg m}^{-2}$ ). The subject has been discussed and reviewed extensively for the radiation-based approach (Monteith, 1977; Sinclair and Muchow, 1999; Stöckle and Kemanian, 2009) and the transpiration based approach (Tanner, 1981; Tanner and Sinclair, 1983; Kemanian et al., 2005). As opposed to the original crop growth analysis proposed by Watson (1952), this framework targets the canopy instead of a representative leaf area section, and offers a robust framework for hypothesis-driven research (Sadras et al., 2005). Most simulation models using this “big leaf” approach for simulating growth apply the radiation-based approach (e.g. EPIC) while a more sophisticated dual approach is used in APSIM, C-Farm, and CropSyst in which the minimum of two estimations of growth is used, one based on transpiration and the other based on radiation interception. Stöckle and Kemanian (2009) have shown that the transpiration based approach is robust in most circumstances, being applicable without any calibration in different environments provided that transpiration is correctly simulated.

The alternative to the “efficiency” based models are enzyme-kinetic models that calculate photosynthesis and transpiration based on a semi-mechanistic understanding of the effects of light,  $\text{CO}_2$ , temperature, humidity, and nitrogen on leaf-level photosynthetic rates and stomatal conductance (Farquhar et al 1980, Collatz et al 1992, Leuning 1995). Multi-layered coupled photosynthesis and transpiration models as those used in the Ecosys model (Grant, 1995), the model WIMOVAC (Humphries and Long, 1995) and that presented by Kremer et al. (2008). A recent study suggested that these multi-layered models perform better than efficiency based models, especially at short measurement timescales (Alton and Bodin, 2010).

### Soil Carbon and Nutrient Cycling

One of the advantages of developing a bioenergy industry is the possibility of producing fuel while reducing the GHG emissions through direct reduction in emission and by offsetting fossil fuel usage. Therefore, simulating the components of the global warming potential of feedstock production systems is critical for a comprehensive assessment of the benefits and impact of bioenergy cropping systems.

Soil carbon cycling is an essential component of comprehensive agricultural and ecological models. Different approaches for simulation the soil carbon balance and its linkages with other nutrients have been discussed extensively elsewhere (Stewart et al., 2008) and a brief summary presented in Kemanian and Stöckle (2010) is used here to present examples of different models. Soil organic carbon is composed of an array of organo-mineral complexes whose turnover rates vary along a continuum from labile or fast turnover fractions to highly recalcitrant fractions. Representing this continuum has been a challenge for soil scientists and biological systems modelers. Early models of soil carbon (Cs) cycling consisted of one Cs pool and one residue pool (Henin and Dupuis, 1945). As basic knowledge on Cs dynamics expanded, new multi-compartment models represented explicitly the microbial pool and separated residues and Cs in several pools (Jenkinson and Rayner, 1977; McGill et al., 1981; Paul and N.G. Juma, 1981; Parton et al., 1988; Verberne et al., 1990; Coleman and Jenkinson, 2005). Other models represented mathematically the Cs turnover rate continuum (Ågren and Bosatta, 1987).

Multi-compartment models separate Cs in pools with different turnover rates. Each pool decomposes due to microbial attack at different rates assumed to depend on the chemical recalcitrance and physical protection of the organic matter fraction: the higher the recalcitrance and physical protection the lower the turnover rate. The carbon lost by a pool can have as destiny the atmosphere ( $\text{CO}_2$  from microbial respiration), the microbial biomass pool, or another carbon pool through chemical reactions or physical aggregation. The transfer of carbon from one pool to another is accompanied by fluxes of other elements such as nitrogen and phosphorus. Six et al. (2002) concluded after an extensive literature review that the success at matching measurable and modelable Cs pools has been minimal. Multi-compartment models such as the Century model (Parton et al., 1988) and Daycent (Del Grosso et al., 2005) have been widely used for assessing Cs evolution and variations of multi-compartment models have been incorporated in comprehensive cropping systems models (e.g. EPIC, Izaurralde et al., 2006; CropSyst, Stockle et al., 2003).

Another approach to accommodate the continuum of turnover rates of soil organic matter is to simulate a single pool of soil organic matter whose turnover rate varies with the size of the carbon pool. This approach is followed in the C-Farm model (Kemanian and Stöckle, 2010). In addition, the size of the organic carbon pool in relation to an assumed maximum carbon carrying capacity or carbon saturation level (Hassink and Whitmore, 1997; Six et al., 2002; Stewart et al., 2008). While this approach requires further testing the number of core parameters of the model is lower than that of multi-compartment models, the spin-up period for equilibrating organic matters pools is not needed, and the interpretation of outputs is straightforward.

### **Nitrous Oxide Emissions**

The high temporal and spatial variability of nitrous oxide emissions from soil under agricultural management makes measurements at regional or national scales impractical (Giltrap et al., 2010). For this reason, there is an opportunity to use process-based models to assess nitrous oxide which are important components of improving the efficiency of cropping systems (minimizing N losses) and reducing their impact on greenhouse gases emissions. However, the variability of  $\text{N}_2\text{O}$  emissions makes modeling this process difficult in various ways. First, it requires an accurate spatial and temporal simulation of nitrate and oxygen content and heterotrophic respiration in soil. Second, there is large spatial variation in this process and the correct “average” condition for a field can be difficult to predict for different landscapes.

Nonetheless, a number of applications of simulation models to estimate nitrous oxide emission rates are presented in the literature. For example, Del Grosso et al. (2005) used the DAYCENT ecosystem model to estimate the nitrous oxide emissions for the main crops in the U.S. arguing that the combination of a process-based model that accounts for cropping system, soil type, climate and tillage and provide more informed decisions than a simple methodology which only considers an emission factor based on N applications. In the emission factor model, nitrous oxide emissions from cropping systems are mainly driven by fertilization events and there is no consideration to other processes that affect emissions such as fertilizer timing or application method. These authors suggest that converting the cropland area to no tillage can reduce, at the national scale, 20 percent of agricultural emissions of this greenhouse gas.

Another model that has been frequently used for simulation of nitrous oxide emissions is DNDC (Denitrification-Decomposition) (Li et al., 1992). Giltrap et al. (2010) reviewed the status of the model and the ability of the model to simulate GHG emissions under different ecosystems. They recognized that the model is a useful tool for modeling the environmental impact of agricultural practices and for improving our understanding of the underlying processes. Hsieh et al. (2005) used DNDC to simulate N<sub>2</sub>O emissions from a fertilized humid grassland in Ireland and found that major emission events followed nitrogen applications and heavy rainfall. The measured annual emissions were 11.6 kg N ha<sup>-1</sup> and the modeled prediction 15.4 kg N ha<sup>-1</sup>, showing that the modeled captured the major emission events reasonably well. This study also indicated that emissions are predicted to increase up to 22.4 kg N ha<sup>-1</sup> under the future climate scenario of the Hadley Center model output, holding other factors constant. Although this model was used here in a grazing system (not a biomass crop) it shows how biophysical models can be applied to better assess the long-term sustainability of cropping systems. Clearly, biomass crops that reduce or minimize external inputs such as N fertilizer will be both energetically more favorable as well as more likely to cause a smaller impact on future climate. In addition, reduced use of N fertilizer will make biomass crops more competitive economically with other alternative sources of energy.

### Sustainability of Biomass Production

There have been several efforts at developing and testing biophysical models with the objective of simulating *M. × giganteus* and *P. virgatum* biomass production and evaluating the sustainability and economic feasibility of bioenergy crops. A recent study by Jain et al. (2010) integrated a biogeochemical model, a simple crop model (based on RUE and light interception) and an economic analysis to evaluate the feasibility and competitiveness of biomass crops *M. × giganteus* and *P. virgatum* with alternative row crops building upon the work of Khanna et al. (2008). In terms of productivity their model estimated that yields of *M. × giganteus* are largely driven by temperature and radiation in the Midwest with maximum peak yields of 7-48 Mg ha<sup>-1</sup> (maximum of 62 Mg ha<sup>-1</sup>). For switchgrass a similar pattern was found but average yields were about 3 times lower (10-16 Mg ha<sup>-1</sup>-maximum of 40 Mg ha<sup>-1</sup>). Under a low-cost scenario, *M. × giganteus* biomass was estimated to have a farm-gate cost between 34 and 80 \$ Mg<sup>-1</sup> (58-131 under the high-cost scenario). The combination of predicted yields and economic considerations identified Missouri as a more competitive state for biomass crops.

A similar modeling approach was used by Heaton et al. (2004a) where a model based on RUE previously calibrated for Ireland (Clifton-brown et al., 2000) was used to predict potential biomass production for *M. × giganteus* in Illinois. As in the model used by Jain et al. (2010), these results are primarily driven by radiation and temperature and they suggested peak average yields between 27-44 Mg ha<sup>-1</sup> for Illinois.

A different approach taken by Wullschlegel et al. (2010) developed a database of *P. virgatum* productivity based on 39 field trials and estimated potential harvestable biomass based on a regression approach with maximum biomass yields projected in a corridor westward from the mid-Atlantic coast region to Kansas and Oklahoma. As opposed to Jain et al. (2010) who concentrated on the *P. virgatum* cultivar Cave-in-Rock, they evaluated a variety of lowland (southern and wetter habitats) and upland (mid and northern latitudes and drier habitats) *P. virgatum* cultivars.

Models that contain biogeochemical routines are suited for evaluating the potential for soil carbon sequestration and the fate of agricultural nitrogen. A subset of these (e.g. Century, DayCENT, CropSyst, C-Farm) are further able to evaluate trace gas emissions. As an example, Davis et al. (2009), using DayCENT, evaluated the greenhouse gas emissions of *M. × giganteus*, corn, *P. virgatum* and native mixed species prairie. All of the perennial crops had lower net greenhouse gas emissions than corn. These authors found *M. × giganteus* to be a sink for GHG emissions in contrast to the net positive GHG emissions from corn, *P. virgatum* and mixed prairie. *M. × giganteus* also had a higher potential for building soil organic carbon than the other feedstocks. In addition, this study suggested that *M. × giganteus* is capable of fixing substantial amounts of atmospheric N, since this was a requirement for balancing the N budget in the DayCENT models and potential N-fixing activity was measured in the rhizomes and rhizosphere of *M. × giganteus* in Illinois (Davis et al., 2009). Further research is needed to confirm the potential of biomass crops with substantial N fixing potential that can reduce the need for external fertilizer inputs.

One of the main concerns of the use of highly productive grasses for biofuel production is their accompanied increase in water use and its effects on the hydrologic cycle. Models that have hydrology sub-models are able to address questions about the potential impacts of biofuel crops on stream flow and nutrient run-off. Vanloocke et al. (2010) used Agro-IBIS to study the potential impact of growing *M. × giganteus* in the Midwestern U.S. Their simulations suggested that if *M. × giganteus* were to be grown in 10% of the land as suggested by Heaton et al. (2008) little impact will occur to the hydrological cycle. Only when simulating a replacement of current vegetation with 50% (or greater) of *M. × giganteus* noticeable changes were detected in the overall hydrological cycle of the Midwestern U.S. with an increase of 40-160 mm per year in total evapotranspiration. This higher ET under *M. × giganteus* is mainly a result of the longer growing season of *M. × giganteus* compared to annual crops such as corn and soybean. However, this small impact on the hydrological cycle can have major effects on climate as the area devoted to highly productive biomass crops is expanded.

Models that have a land surface model are designed to capture the full energy and mass balance of the ecosystem at a fast time scale. This enables these models to be coupled with atmospheric models and thus address questions about the potential atmospheric feedbacks that could result from large-scale biofuel crop deployment. These feedbacks could include changes in air temperature and precipitation patterns. This is an active area of research and integrated models capable of producing robust forecasts are under development.

The Ecosystem Demography model (ED) is a physiologically-based plant growth model that was originally formulated to model forest ecosystem dynamics (Medvigy et al., 2009). ED is being applied to evaluate woody biofuel crops such as hybrid poplar as well as to evaluate the potential use of native forest and other novel tree species (Wang and Dietze, *in prep*). ED has also been reformulated to represent perennial grasses and in particular is leveraging its representation of community dynamics to address the use of native grasslands and polycultures.

## Databases

There are a number of datasets that play a critical role as drivers of biofuel crop models as well as in their parameterization, calibration, and validation. Below we highlight some of these resources. For drivers we focus on the availability of data related to weather and soils, while for model testing we focus on databases that compile site-level yield data and species-level ecophysiological data. There are a number of other resources that are commonly used to test plant and ecosystem models in other contexts but which are not yet utilized extensively by biofuel modelers, generally because there is a limitation of data due to the small scales at short histories for many second generation crops. These include remote sensing, eddy-covariance, and USDA county-level data on crop and forest production. As research matures, and biofuel crops are planted on larger scales, modelers are encouraged to look more broadly to these and other emerging data sets.

### Biofuel Trait and Yield Database

The Biofuel Ecophysiological Trait and Yield Database (BETY-db, <http://ebi-forecast.igb.uiuc.edu/>) was created in order to compile the available field data about proposed “second generation” biofuel crops. There are two categories of data currently represented in the database: information on the productivity of different species and cultivars at different sites and “trait” information on the characteristics of different species. These data are also associated with detailed information on treatments that have been applied (e.g. different levels of N addition) and different management operations (e.g. dates of planting and harvest). Both types of data can be queried in a number of ways, for example by species or by location using a Google map interface. In the context of modeling biofuel crops the trait database is intended primarily to produce initial estimates for model parameters. Existing utilities in BETY-db have been designed to estimate the probability distributions of each trait based on a meta-analytical model (LeBauer et al *in prep*). Yield data across many sites are also critical for model validation. Beyond model applications, the database is intended to promote data sharing and cross-site syntheses. For example, a meta-analysis of the switchgrass data from this database suggested that perennial grasses grown with legumes may have comparable yields and lower inputs than fertilized monocultures (Wang et al., 2010). Similarly, analyses of trait data may be useful for pre-screening potential species or cultivars based on comparison to the traits of current crops. Finally, the spatial query in the database is intended to allow land managers and extension agents evaluate what yields have actually been achieved in a given region by different crops.

### Meteorological Data

A crucial component needed for evaluating which crops to grow for bioenergy and where and how to grow them is the weather and climate data for a particular region. In order to make regional-scale projections of biofuel crops all models require estimates of climate that reflect the differences among regions. Furthermore, most models are dynamic and thus need detailed weather data with high temporal and spatial resolution. The critical variables are precipitation, temperature, yet most models also require or render better results when humidity, atmospheric pressure, and wind speed are also available. Carbon dioxide concentration is also needed by enzyme-kinetic photosynthesis models. Land surface sub-models, which explicitly calculate the overall energy balance, will typically need to be able to resolve the sub-daily cycles of these variables. A number of models that run at hourly time intervals are capable of using daily meteorological records and simulate hourly conditions based on typical patterns of temperature and radiation daily fluctuations (Campbell and Norman, 1998). The hourly fluctuations of precipitation, wind speed and relative humidity are harder to simulate realistically based on daily summaries and these are often considered uniform or simulated stochastically using a “weather generator” algorithm.

Closely related to weather is solar radiation, with models varying from those that just require an overall light level to those that need radiation broken up by different spectral bands (e.g. photo-synthetically active radiation, near infra-red, and long-wave infra-red) or into direct “beam” radiation versus indirect or diffuse radiation. Since meteorological stations are not laid out on a well-defined grid, modelers rely on data products that have been interpolated either statistically or, more often, via data assimilation in atmospheric models. Weather databases can generally be divided by their spatial and temporal resolution. Below we will describe some of the data products available at a state-by-state level, nationwide, and globally.



## Statewide

The Iowa Environmental Mesonet (<http://mesonet.agron.iastate.edu/>) is an example of weather data that are synthesized from 7 different observing networks and represents an outstanding effort at integrating meteorological variables for different purposes. Hourly (or even every minute) data from ASOS (Automated Surface Observing Systems) and AWOS (Automated Weather Observing System) can be obtained from the Iowa Environmental Mesonet from a convenient interface (<http://mesonet.agron.iastate.edu/request/asos/1min.phtml>). For Illinois, there is another weather database managed by the Illinois State Water Survey (<http://www.isws.illinois.edu/data/climatedb/> and <http://www.isws.illinois.edu/warm/datatype.asp>). These weather databases are suitable for use in most computer simulation models that typically run at daily or hourly time intervals.

## U.S. Nationwide

At the national scale there are a number of data products available, however there is a strong trade-off among data products in terms of spatial vs. temporal resolution. The PRISM database (Parameter-elevation Regressions on Independent Slopes Model) has the greatest spatial resolution (a grid of 800-m) but has the coarsest temporal resolution (monthly). At the other extreme, NARR, the product with the highest temporal resolution (3 hrs) also has the coarsest spatial resolution (32-km). This trade-off in part reflects the fact that there is only a finite amount of information in the network of weather stations. It also reflects a switch between statistical and atmospheric models, the latter possessing computational constraints in reducing their spatial resolution but inherently operating at high temporal resolution.

### PRISM-<http://www.prism.oregonstate.edu/>

PRISM uses meteorological station “point” data and a digital elevation model (DEM) to generate fine-scale (800m) gridded estimates of climate parameters on a month-by-month basis (Daly et al., 1994). PRISM is designed specifically to capture the small-scale topographic variability in climate, using a DEM and a windowing technique to group stations onto individual topographic facets. PRISM develops a weighted precipitation/elevation (P/E) regression function to predict precipitation at the elevation of each cell using data from nearby stations, with greater weight given to stations with location, elevation, and topographic positioning (e.g. aspect) similar to that of the grid cell. In a model comparison, PRISM exhibited superior performance to various methods of kriging, and has been successfully applied to the entire United States with excellent results.

### Daymet-<http://www.daymet.org>

Daymet is a semi-mechanistic statistical model conceptually similar to PRISM that generates daily surfaces of seven variables: daily mean, minimum, and maximum temperature, precipitation, humidity, radiation, and day length (Thornton and Running, 1999). The Daymet data set spans 1980-2003 and has a 1km resolution. Data are downloadable either as time-series at point locations or climatological maps.

### NARR-<http://nomads.ncdc.noaa.gov/>

The North American Regional Reanalysis (NARR) is an atmospheric-model data-assimilation product from NOAA that covers all of North America and parts of the Atlantic Ocean, Pacific Ocean, Central America and the Eurasian arctic. Historical climate data that has been assimilated through atmospheric models is typically referred to as “reanalysis” products and a number of other reanalysis data sets are available on a global scale and will be discussed below. The NARR has a spatial resolution of approximately 32 km and a 3 hour temporal resolution and spans the time period from 1979 to the present. Because the NARR is processed through an atmospheric model there are a large number of output variables available that include both the state of the land surface and the atmosphere.

## Global

At a global scale there is a diversity of different products available. In terms of raw weather station data and statistically interpolated products there are three worth mentioning: CRU, LocClim, and WorldClim. The CRU dataset is a product of the Climate Research Unit at the University of East Anglia (<http://www.cru.uea.ac.uk/cru/data/availability/>) which provides gridded surface temperature datasets over the past 150 years and has played a critical role in diagnosing spatial patterns of climate change. LocClim (<http://www.fao.org/sd/locclim/srv/locclim.home>) is a UN FAO tool used to estimate eight different climate variables: Average, minimum, and maximum temperatures, precipitation, light, humidity, wind speed, and potential evapotranspiration. Estimates are available at monthly, 10-day, and daily time intervals. The grid resolution in LocClim is not predetermined; the utility performs interpolation on-the-fly based on latitude, longitude, and elevation. The underlying dataset in LocClim is the FAOCLIM data set of 28800 met stations. WorldClim is a high-resolution (1km) global gridded data set of average climate for 1950-2000 (Hijmans et al., 2005) for 23 climate variables: mean, minimum, and maximum temperature, precipitation, and 19 bioclimatic indicators. The same algorithm has also been used to produce climate maps for IPCC climate change scenarios (2020, 2050, and 2080 under the A2A and B2A emissions scenarios) and for the mid-Holocene (6000BP), last glacial maximum (21,000BP), and last interglacial (130,000BP).

In addition to statistically gridded data sets, there are also a few key global “reanalysis” data sets. The most commonly used are the ECMWF (European Centre for Medium-Range Weather Forecasts) “ERA-40” (Uppala et al 2005, <http://data.ecmwf.int/data/>) and the NCEP (National Center for Environmental Prediction) “Reanalysis 2” (Kanamitsu et al. (2002), <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html>). Both these data products have a 2.5 degree resolution and a 6 hour time step. The ERA-40 covers 1957-2001 with a newer ERA-Interim product covering 1989-2009 while the NCEP covers 1979-2008 with a newer “Twentieth Century” product covering 1871-2008 (Compo et al., 2010). There is also a reanalysis from the Princeton Land Surface Hydrology Research Group (LSHRG, Sheffield et al 2006)) that attempts to correct biases in the NCEP reanalysis based on a number of satellite and surface data compilations, such as CRU, and which appears to have the least biased radiation (Ricciuto pers com). This data set is available at 3-hr and monthly time steps and a 1.0 degree resolution.

## Soil Databases

Another important component for estimating biomass productivity and ecosystem services of biomass production are soil characteristics. For a specific location, soil properties can be measured directly, but soil sampling and analysis is typically time consuming and costly; and for large regions prohibitive. Assessing sustainability of biomass production at a regional level requires incorporating soil information and here we describe the main sources of soil data on a national and global scale.

## SSURGO

The Soil Survey Geographic (SSURGO) database is available for selected counties and areas throughout the United States and its territories. In SSURGO mapping scales generally range from 1:12,000 to 1:63,360 and this is the most detailed level of soil mapping done by the Natural Resources Conservation Service (NRCS). Maps are derived from point observation and conceptual models of soil formation (Soil Survey Staff, 2009). This database is linked to a National Soil Information System (NASIS) attribute database which provides the relative extent of the component soils and their properties for each map unit. The SSURGO map units consist of 1 to 3 components each (Figure 1). The database consists of two main components, a GIS polygon map of different soil map units and a set of attribute tables that describe different soil properties for those map units, often with attributes varying with depth. For the purpose of biomass production modeling, examples of information that can be queried from the database are: soil texture, soil organic matter, pH, available water capacity, soil reaction, and electrical conductivity.

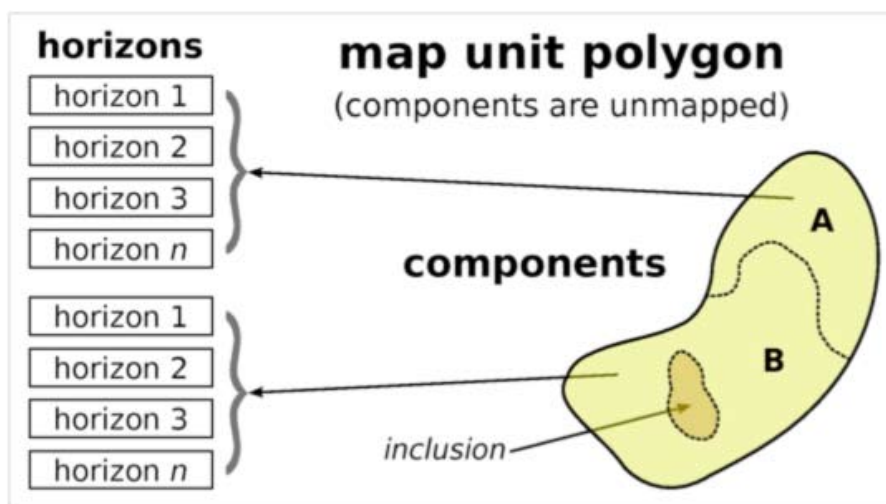


Figure 1: Structural diagram of USDA-NRCS digital soil survey data. Spatial data represent map unit polygons, usually consisting of multiple un-mapped components. The complex hierarchy of map unit → component → horizon data is encoded through a series of 1-to-many tabular relationships. Reproduced with permission from Beaudette, 2008.

The database provides basic information from where the soil profile input required for the model has to be derived. This is not a simple task as the input, for instance the layering of the soil profile, is more detailed than the original information and the correlation between variables has to be conserved. Soil organic matter estimates for the profile and the distribution with depth has to be scrutinized carefully as using the raw data carelessly will most likely result in poor outputs. Pedotransfer functions are customarily used to predict soil properties from basic textural data (e.g. Saxton and Rawls, 2006).

## STATSGO2

For larger scale simulations (i.e. national scale) the U.S. General Soil Map, known as STATSGO2, consists of general soil association units, which is generalized soil information interpreted from detailed soil survey data and inferred from natural conditions where soil information is absent. It was developed by the National Cooperative Soil Survey and it consists of a broad-based inventory of soils and non-soil areas that occur in a repeatable pattern on the landscape and that can be cartographically shown at the approximate scale of 1:250,000.

The design of STATSGO is very similar to SURGO. The tabular data contain estimated data on the physical and chemical soil properties, soil interpretations, and static and dynamic metadata. Most tabular data exist in the database as a range of soil properties, depicting the range for the geographic extent of the map unit. In addition to low and high values for most data, a representative value is also included for these soil properties. This indicates that working at this scale there is a source of uncertainty that has to be taken into account, since the magnitude of the variability in soil variables of interest can be substantial.

## Using the Soil Databases

The simplest way to access the data from the soil databases is the Web Soil Survey (<http://websoilsurvey.nrcs.usda.gov/app/>) which is an interactive web application that allows access to maps and soil characteristics and attributes.

Data for soil survey contains a tabular and spatial component. The spatial component is a vector file (ESRI shape file) with the “map unit” key as the main information. The tabular data contains four general classes of information: 1) chemical and physical data (pH, CEC, particle size distribution, etc.), 2) morphologic data (horizonation, etc.), 3) taxonomic data and 4) interpretations for land use and engineering. The vast number of decisions made based on soil surveys reflect the inherent value of this information (Beaudette, 2008).

An example application of the STATSGO2 database is the rasterized calculation of available water capacity, using a 32 by 32 km grid over the conterminous U.S. (Figure 2). The available water capacity of a soil is a crucial variable in estimating the potential biomass productivity of different regions.

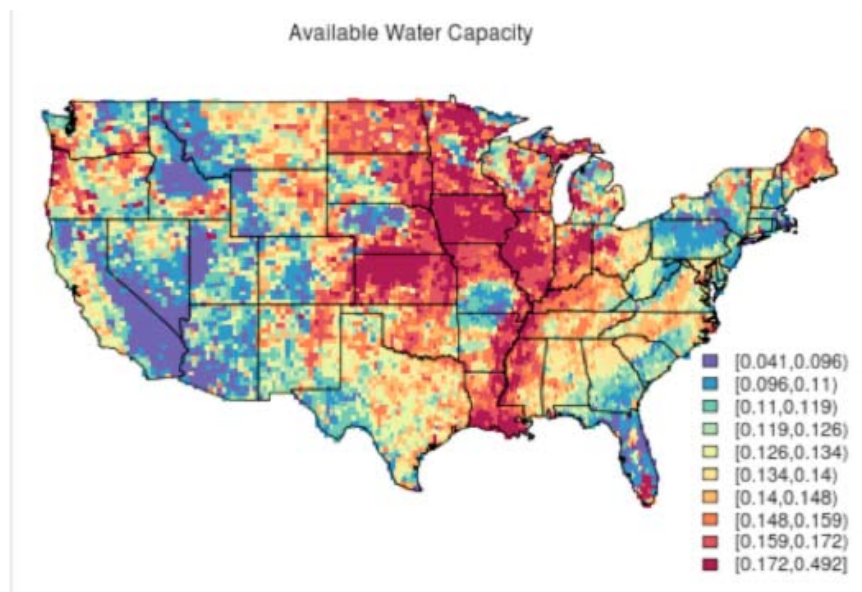


Figure 2: Available water capacity (proportion) based on a 32 by 32km grid over the conterminous U.S.. This map was produced by rasterizing the STATSGO2 and performing a weighted average over different horizon depths and the proportionate contribution of soil components.

### Global Scale Soils Data

At a global scale it is difficult to compile all the different national soils maps that use different resolutions, classifications, and sampling methods. Fortunately the U.N. Food and Agriculture Organization does provide a global-scale soils map (<http://www.fao.org/nr/land/soils/en/>). This map is fairly coarse in resolution, but does provide information on soil texture and soil depth that is required to drive the soil moisture sub-models of most vegetation models. To our knowledge there is not a global scale map of soil carbon stores, soil nutrients, or other soil biogeochemical rates or properties, though model-based estimates of some of these do exist as part of climate change research (IPCC, <http://www.ipcc-data.org/>).

### Land Use Databases

Other databases of importance are those providing information about land cover. This is useful when performing detailed landscape-level assessments of the impact of crops, trees or other large scale practices. One example is the national land cover database from the Multi-Resolution Land Characteristics Consortium (MRLC, [www.mrlc.gov](http://www.mrlc.gov)). This is available for the 50 U.S. states and it provides classification of land on a 30 by 30m resolution that can be used to plan where biomass crops might be deployed at a more detailed level. One disadvantage is that the latest version is from 2001 and many changes might have occurred to land use since then. Examples of land classification are: open water, grassland, cropland, mixed forest, etc. Another useful database is the USDA-NASS Cropland Data Layer (CDL) which contains crop specific information. The CDL Program annually focuses on producing digital categorized geo-referenced output products using imagery from the Resourcesat-1 AWIFS and the Landsat 5 TM satellites (<http://www.nass.usda.gov/research/Cropland/SARS1a.htm>). At a global scale the MODIS satellites provide an annual 500m land cover estimate from 2001 to the present (<http://modis-land.gsfc.nasa.gov/>). These maps provide up-to-date land cover information and can be useful for both modeling outside the U.S. and for assessing land cover change.

## Model Assessment

Models are only as good as the data that go into building them and thus model assessment is a critical activity. We can conceptually break assessment down into two phases, training and testing. Activities in the training phase are focused on using data to estimate model parameters while the testing phase is focused on confronting the model with independent data.

There are a number of different approaches used during the training phase and we will conceptually break them down into what we call parameterization and calibration, though these labels are not universally used and not all techniques fit nicely into these definitions. By parameterization we refer to the process of setting model parameters where there is a direct mapping of field or experimental data to a specific parameter or set of parameters. This definition is distinct from usages found in other fields, such as atmospheric science, where most model parameters are known physical constants and parameterization instead refers to the choice of a functional form for modeling a process statistically rather than mechanistically. Examples of this could range from 1:1 mappings between parameters and data, such as the C:N ratio of a tissue or the specific leaf area of a leaf, to parameters that are fit statistically but still have a direct link to data, such as the estimation of photosynthetic parameters from an A/Ci curve or an exponential decay rate from a litter bag experiment. Parameterization has traditionally occurred by reference to the scientific literature or using expert opinion to fix parameter values. In the past it has often been difficult for the non-expert to see where specific model parameters have come from, which has been known to engender distrust of models. Some of the disadvantages of traditional parameterization are that error distributions associated with parameters have rarely been reported and there has been a bit of subjectivity in choices about why parameter values from one study were chosen over another. Newer meta-analytical techniques aim to get around this because they allow parameters to be constrained based on the combined weight of multiple studies and provide a formal estimate of parameter uncertainty that can be used for error propagation (LeBauer et al in prep).

In contrast to parameterization, where there is a direct mapping between data and parameters, we use the term calibration to deal with the situation where the connection between data and parameters is often less direct but more holistic. In general during calibration we are comparing a model output to data, for example the comparison between predicted and observed yield. Yield is not determined by a single parameter but is influenced by many different parameters in many different processes. Another important distinction between parameterization and calibration is that the whole model has to be run in calibration while in parameterization we only need to know the biological meaning of a parameter or a single functional relationship. Because of this, calibration methods end up being much more computationally intensive. However, there are a few advantages of calibration. First, it allows the estimation the overall error variance of the model. Second, it potentially allows for the estimation of covariances between parameters, which can often be substantial and tend to reduce the overall model uncertainty. Third, calibration allows one to estimate model parameters that are difficult or impossible to measure directly in the field, for example, carbon allocation (Miguez, 2009).

There are a number of statistical methods available that can be used during calibration. In general it is best to base calibration on objective criteria rather than simply “tuning” the model-manually adjusting free parameters to make the model match the data. The statistical approaches to calibration have sometimes been referred to as “inverse modeling” because it is the reverse of “forward” modeling where a model is run forward given a set of known parameters in order to produce an unknown output. Instead in inverse modeling the desired output is known (i.e. data) and the goal is to figure out what parameters produce the required outputs. We will discuss three approaches to calibration: minimization of an objective function, maximization of a likelihood, and estimation of the posterior parameter distribution. In the first approach the modeler must specify some function that they would like to minimize. Traditionally, the mean squared error (MSE), the sum of squares error (SSE) or other function that expresses the mismatch between the model and the data, which will be minimize.

$$SSE = \sum_i (O_i - S_i)^2$$

$$MSE = \frac{1}{n} \sum_i (O_i - S_i)^2$$

Where  $O_i$  is the observed data,  $S_i$  is the simulated data and  $n$  is the total number of observations.

Given the complexity of vegetation models analytical solutions to these minimizations typically do not exist and one uses a numerical optimization algorithm (Bolker, 2008). The second approach, maximum likelihood, is similar to the objective function approach except that instead of minimizing an objective function one is instead calculating the probability that a certain parameter set would have produced the observed data. This probability statement is referred to as the likelihood function and the goal is usually to find the most likely parameter values, i.e. those that maximize the likelihood function. As with objective functions, likelihood functions are usually evaluated using numerical optimization. The most common choice of probability distributions is to assume that error is normally distributed, in which case the maximum likelihood solution is equivalent to the sum of squares objective function, which is likewise the most commonly chosen objective function (Givens and Hoeting, 2005).

An example in the context of biomass crops where the objective was to produce reliable estimates of switchgrass productivity used a combination of parameters derived from the literature and optimization using a numerical algorithm minimizing the mean sum of squares of the error function (Di Vittorio, et al. 2010). While the authors were able to obtain several parameters directly from the literature, they identified 5 parameters which needed to be optimized based on data. These parameters were mostly related to root and carbon dynamic processes which are seldom measured in detail in individual studies. This effort at identifying uncertainty in parameters and evaluating the robustness of model simulations is crucial for the generation of robust forecasts of feedstock availability.

The third alternative for calibration, estimation of the posterior parameter distribution, is also based on probability theory, just like maximum likelihood, but instead employs Bayes' Theorem in order to estimate the full probability distribution of a parameter (Gelman et al., 2004). Bayesian methods are popular because most often what we are actually interested in is the probability of the model parameters not the probability of the data, which is calculated in maximum likelihood. Furthermore, because these methods provide a whole probability distribution for the parameter, rather than a single optimum value, they more directly capture and propagate model uncertainty. Bayesian posterior parameter distributions are usually estimated by Markov Chain-Monte Carlo (MCMC) numerical techniques, which tend to be more computationally demanding than numerical optimization (Brooks, 1998).

Before proceeding on to model testing we also wanted to briefly touch on data assimilation methods, which have received a lot of attention in the modeling literature lately. The exact definition of data assimilation varies from discipline to discipline and many modelers refer to techniques that we would lump under calibration as data assimilation. Traditionally in atmospheric science, where data assimilation has seen the greatest use, the technique referred strictly to methods for estimating the value of a model's state variables from data, rather than estimating model parameters. Data assimilation can further be broken down into off-line methods, where all the data are available, and on-line methods, where data assimilation is being performed in real time and each new data point arrives in order with analyses being updated at each time point. Wikle and Berliner (2007) give a good overview of both classical and Bayesian approaches to data assimilation while Lewis et al. (2006) provided a detailed treatment of these methods.

Finally, after the model training phase models then often undergo a testing phase, which is sometimes also referred to as model validation or model verification. Typically some portion of the data collected is withheld during the training phase for use in the testing phase, since the aim is to provide an independent test of model performance rather than testing the model against the same data that was used for calibration. During testing model parameters are either fixed to the values estimated during the training phase or, if Bayesian methods were used, are sampled from their posterior distributions. In the latter it is customary to run the model many times to generate an “ensemble” estimate of model uncertainty. While modelers often refer to this phase a model validation, technically we can never assess if the model is valid in all situations, and indeed all models will be wrong under some conditions (Oreskes et al 1994). Rather we are attempting to discern under which conditions the model is reliable and which it is not.

A major challenge in modeling efforts is to integrate databases, field experiments, biophysical models while using optimization and sensitivity analysis techniques. A strategy for simulating productivity and assessing sustainability of biomass feedstocks is to integrate databases, biophysical models and statistical approaches. Miguez et al (2009) developed *M. × giganteus* harvestable biomass projections by integrating weather data from North American Regional Reanalysis, the U.S. general soil map (STATSGO2) and a biophysical model (Figure 3). A recent example of data and model integration, specifically targeted to evaluating the sustainability and productivity of biofuel crop systems was presented by Zhang et al. (2010). Within their spatially explicit framework, they integrated weather data from NARR, the EPIC biophysical model, the SSURGO soil database, Land use, hydrological unit and political boundaries into a homogeneous spatial modeling unit. Using an optimization algorithm they were able to develop a set of optimal solutions that represents a compromise between N losses, energy production and greenhouse gas emissions.

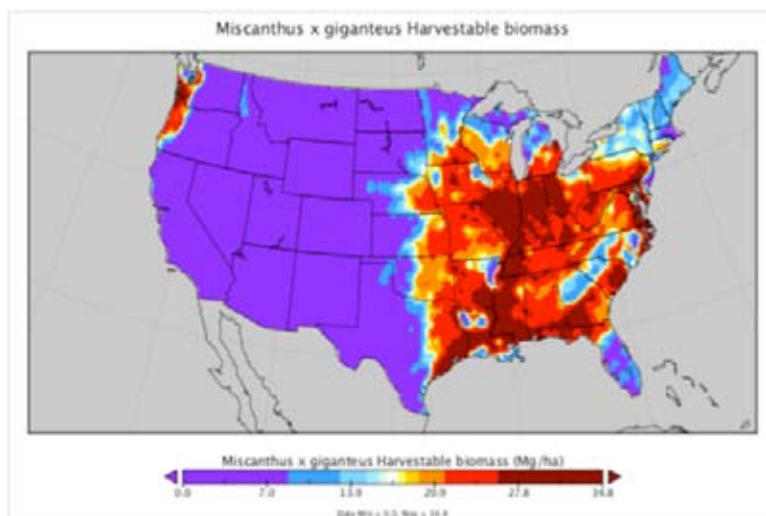


Figure 3: *M. × giganteus* simulated harvestable biomass production for the U.S. integrating weather (NARR) and soil (STATSGO2) databases.

## Conclusions

The main tools for regional assessment of productivity and sustainability of biomass feedstocks were presented in this chapter and these are databases, biophysical models of crops, forests and ecosystems and statistical approaches to training and testing of the models.

An ideal framework would be to have easy-to-use (and access) databases that can inform models in their development or adaptation process. It is of utmost importance that these databases are maintained and quality control criteria are used. The databases can later be used in testing the model simulations as well and this does not need to be a static process, but rather a continuous process in which models are developed, tested and refined.



Biophysical models are useful but the objectives and applications typically determine the set of models which are more appropriate in a given situation and suggest the correct balance in terms of model complexity, data available and desired outcome.

## References

- Ågren, G., Bosatta, N., 1987. Theoretical Analysis of the Long-Term Dynamics of Carbon and Nitrogen in Soils. *Ecology* 68 (5), 1181-1189.
- Alton, P, Bodin P. 2010. A comparative study of a multilayer and a productivity (light-use) efficiency land-surface model over different temporal scales. *Agricultural and Forest Meteorology*. 150(2):182-195.
- Beaudette, D., 2008. New Technologies to Construct and Present Soil Surveys. Ph.D. thesis, University of California Davis.
- Bolker, B., 2008. *Ecological Models and Data in R*. Princeton University Press, Ch. Optimization and All that, pp. 222-262.
- Brooks, S. P., 1998. Markov Chain Monte Carlo and Its Application. *The Statistician* 47, 69-100.
- Campbell, G. S. and Norman, J. M. 1998. *Introduction to Environmental Biophysics*. Springer. Second Edition. Pg. 286.
- Clifton-Brown, J. C., Neilson, B., Lewandowski, I., Jones, M. B., 2000. The modeled productivity of *Miscanthus x giganteus* (GREEF et DEU) in Ireland. *Industrial Crops and Products* 12, 97-109.
- Clifton-Brown, J. C., Stampfl, P. F., Jones, M. B., Apr. 2004. *Miscanthus* biomass production for energy in Europe and its potential contribution to decreasing fossil fuel carbon emissions. *Global Change Biology* 10 (4), 509-518.
- Collatz, G., Ribas-Carbo, M. & Berry, J. (1992) Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants. *Australian Journal of Plant Physiology*, 19, 519.
- Coleman, K., Jenkinson, D., 2005. A model for the turnover of carbon in soil. <http://www.rothamsted.bbsrc.ac.uk/aen/carbon/download.htm>.
- Compo, G., Whitaker, J., Sardeshmukh, P., Matsu, N., Allan, R., Yin, X., B.E. Gleason, J., Vose, R., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R., Grant, A., Groisman, P., Jones, P. D., Kruk, M., Kruger, A., Marshall, G., Maugeri, M., Mok, H., Nordli, O., Ross, T., Trigo, R., Wang, X., Woodruff, S., Worley, S., 2010. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*.
- Daly, C., Neilson, R., Phillips, D. L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33 (2), 140-158.
- Davis, S. C., Parton, W. J., Dohleman, F. G., Smith, C. M., Grosso, S. D., Kent, A. D., DeLucia, E. H., 2009. Comparative Biogeochemical Cycles of Bioenergy Crops Reveal Nitrogen-Fixation and Low Greenhouse Gas Emissions in a *Miscanthus x giganteus* Agro-Ecosystem. *Ecosystems* 13 (December 2009), 144-156.
- Del Grosso, S., Mosier, A., Parton, W., Ojima, D., 2005. DAYCENT model analysis of past and contemporary soil NO and net greenhouse gas flux for major crops in the U.S.A. *Soil and Tillage Research* 83 (1), 9-24.



- Di Vittorio AV, Anderson RS, White JD, Miller NL, Running SW. 2010. Development and optimization of an Agro-BGC ecosystem model for C4 perennial grasses. *Ecological Modelling*. 2010; 221(17):2038-2053.
- Farquhar, G., Caemmerer, S. & Berry, J. (1980) A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta*, 149, 78-90.
- Food And Agriculture Organization (FAO), 2007. State of the World's Forests.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2004. Bayesian Data Analysis. Chapman and Hall CRC.
- Giltrap, D. L., Li, C., Saggar, S., Mar. 2010. DNDC: A process-based model of greenhouse gas fluxes from agricultural soils. *Agriculture, Ecosystems & Environment* 136 (3-4), 292-300.
- Givens, G., Hoeting, J., 2005. Computational Statistics. Wiley.
- Grant, R., 1995. Salinity, water use and yield of maize: Testing of the mathematical model ecosys. *Plant and Soil* 172, 309-322.
- Hassink, J., Whitmore, A. P., 1997. A model of the physical protection of organic matter in soils. *Soil. Sci. Soc. Am. J* 61, 131-139.
- Heaton, E., Voigt, T., Long, S., Anonymous, 2004a. A quantitative review comparing the yields of two candidate C4 perennial biomass crops in relation to nitrogen, temperature and water. *Biomass and Bioenergy* 27 (1), 21-30.
- Heaton, E. a., Clifton-Brown, J., Voigt, T. B., Jones, M. B., Long, S. P., Oct. 2004b. Miscanthus for Renewable Energy Generation: European Union Experience and Projections for Illinois. *Mitigation and Adaptation Strategies for Global Change* 9 (4), 433-451.
- Heaton, E. a., Dohleman, F. G., Long, S. P., Sep. 2008. Meeting U.S. biofuel goals with less land: the potential of Miscanthus. *Global Change Biology* 14 (9), 2000-2014.
- Hénin, S., Dupuis, M., 1945. Essai de bilan de la matière organique du sol. *Ann. Agron.* 15, 17-29.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., Jarvis, A., Dec. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25 (15), 1965-1978.
- Hill, J., Nelson, E., Tilman, D., Polasky, S., Tiffany, D., Jun. 2006. From the Cover: Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences* 103 (30), 11206-11210.
- Hsieh, C.-I., Leahy, P., Kiely, G., Li, C., Sep. 2005. The Effect of Future Climate Perturbations on N<sub>2</sub>O Emissions from a Fertilized Humid Grassland. *Nutrient Cycling in Agroecosystems* 73 (1), 15-23.
- Humphries, S., Long, S., 1995. WIMOVAC-a software package for modeling the dynamics of the plant leaf and canopy photosynthesis. *Computer Applications in the Bioscience* 11 (4), 361-371.
- Izaurrealde, R., Williams, J., McGill, W., Rosenberg, N., Jakas, M., Feb. 2006. Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecological Modelling* 192 (3-4), 362-384.

- Jain, A. K., Khanna, M., Erickson, M., Huang, H., May 2010. An integrated biogeochemical and economic analysis of bioenergy crops in the Midwestern United States. *GCB Bioenergy*, no-no.
- Jenkinson, D., Rayner, J., 1977. The turnover of soil organic matter in some of the Rothamsted classical experiments. *Soil Science* 123, 293-305.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., Potter, G., 2002. NCEPâ€œDOE AMIP-II reanalysis (R-2). *Climate Research* (November), 1631-1643.
- Kemanian, a., Stockle, C., Huggins, D., May 2005. Transpiration-use efficiency of barley. *Agricultural and Forest Meteorology* 130 (1-2), 1-11.
- Kemanian, A. R., Stöckle, C. O., Jan. 2010. C-Farm: A simple model to evaluate the carbon balance of soil profiles. *European Journal of Agronomy* 32 (1), 22-29.
- Khanna, M., Dhungana, B., Clifton-Brown, J., 2008. Costs of Producing Miscanthus and Switchgrass for Bioenergy in Illinois. *Biomass and Bioenergy*, 482-493.
- Kowalski, K., Stagl, S., Madlener, R., Omann, I., Sep. 2009. Sustainable energy futures: Methodological challenges in combining scenarios and participatory multi-criteria analysisâ€†. *European Journal of Operational Research* 197 (3), 1063-1074.
- Kremer, C., Stöckle, C., Kemanian, A., Howell, T., 2008. Response of crops to limited water. *Advances in Agricultural Systems Modeling* (L.R. Ahuja, V.R. Reddy, S.A. Saseendran, Q. Yu Eds), Ch. A reference canopy transpiration and photosynthesis model for the evaluation of simple models of crop productivity, pp. 165-190.
- Leuning, R. (1995) A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. *Plant, Cell and Environment*, 18, 339-355.
- Lewis, J. M., Lakshmivarahan, S., Dhall, S., 2006. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, New York.
- Li, C., Frolking, S., Frolking, T. A., 1992. A Model of Nitrous Oxide Evolution From Soil Driven by Rainfall Events: 1. Model Structure and Sensitivity. *Journal of Geophysical Research*, 9759-9776.
- McGill, W., Hunt, H., Woodmansee, R., Reuss, J., 1981. *Terrestrial Nitrogen Cycles: Processes, Ecosystem, Strategy and Management Impacts*. Clark F.E. and Rosswall T. (Eds.). *Ecol. Bull.* (Stockholm), Ch. PHOENIX: a model of the dynamics of carbon and nitrogen in grassland soils, pp. 49-115.
- Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., Moorcroft, P. R., 2009. Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *Journal of Geophysical Research* 114 (G1), 1-21.
- Miguez, F., 2009. Estimating dry biomass partitioning coefficients in a dynamic vegetation model. In: *ASA-CSSA-SSSA*. ASA, Maison, WI, pp. CD-ROM.
- Miguez, F. E., Zhu, X., Humphries, S., Bollero, G. a., Long, S. P., 2009. A semimechanistic model predicting the growth and production of the bioenergy crop *Miscanthus Ægiganteus*: description, parameterization and validation. *GCB Bioenergy* 1 (4), 282-296.
- Monteith, J., 1977. Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. London, Ser. B. Biol. Sci* 281 (980), 277-294.

- Nass, L. L., Pereira, P. A. A., Ellis, D., Nov. 2007. Biofuels in Brazil: An Overview. *Crop Science* 47 (6), 2228-2237.
- Oreskes, N., Shrader-Frechette, K. & Belitz, K. (1994) Verification, validation, and confirmation of numerical models in the Earth sciences. *Science* (New York, N.Y.), 263, 641-6.
- Parton, W., Stewart, J., Cole, C., 1988. Dynamics of c, n, p and s in grassland soils: a model. *Biogeochemistry* 5, 109-131.
- Paul, E., N.G. Juma, ., 1981. Terrestrial Nitrogen Cycles: Processes, Ecosystem, Strategy and Management Impacts. Clark F.E. and Rosswall T. (Eds.). *Ecol. Bull.* (Stockholm), Ch. Mineralization and immobilization of soil nitrogen by microorganism, pp. 179-195.
- Perlack, R., Wright, L., Turhollow, A., Graham, R., Stokes, B., Erbach, D., 2005. Biomass as Feedstock for a Bioenergy and Bioproducts Industry : The Technical Feasibility of a Billion-Ton Annual Supply. No. April. Oak Ridge National Laboratory, Oak Ridge, TN, Oak Ridge, TN.
- Richard TL. Challenges in Scaling Up Biofuels Infrastructure. *Science*. 2010;329(5993):793-796.
- Sadras, V., Oleary, G., Roget, D., Feb. 2005. Crop responses to compacted soil: capture and efficiency in the use of water and radiation. *Field Crops Research* 91 (2-3), 131-148.
- Sawyer, D., May 2008. Climate change, biofuels and eco-social impacts in the Brazilian Amazon and Cerrado. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363 (1498), 1747-52.
- Shapouri H, Duffield JA, Wang M. 2002. The Energy Balance of Corn Ethanol: An Update. *Agricultural economic report N 814*. Pg 19.
- Sinclair, T. R., Muchow, R. C., 1999. Radiation Use Efficiency. *Advances in Agronomy* 65, 215-265.
- Six, J., Conant, R. T., Paul, E. A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter : Implications for C-saturation of soils. *Plant and Soil*, 155-176.
- Soil Survey Staff, Nov 2009. Soil Survey Geographic (SSURGO) Database for [U.S.]. United States Department of Agriculture accessed [05/13/2009], 1-22.
- Somerville, C., Youngs, H., Taylor, C., Davis, S. C., Long, S. P., Aug. 2010. Feedstocks for Lignocellulosic Biofuels. *Science* 329 (5993), 790-792.
- Stewart, C. E., Plante, A. F., Paustian, K., Conant, R. T., Six, J., 2008. Soil Carbon Saturation: Linking Concept and Measurable Carbon Pools. *Soil Science Society of America Journal* 72 (2), 379.
- Stöckle, C., Donatelli, M., Nelson, R., Jan. 2003. CropSyst, a cropping systems simulation model. *European Journal of Agronomy* 18 (3-4), 289-307.
- Stöckle, C., Kemanian, A., 2009. Crop Physiology (V. Sadras and D. Calderini Eds). Academic Press, Elsevier Inc., Ch. Crop radiation capture and use efficiency: A framework for crop growth analysis., pp. 145-170.
- Tanner, C., 1981. Transpiration efficiency of potato. *Agronomy Journal* 73, 59-64.

- Tanner, C., Sinclair, T., 1983. H.M. Taylor et al. (ed) Limitations to efficient water use in crop production. ASA, Madison, WI., Ch. Efficient water use in crop production: research or re-search., pp. 1-27.
- Thornton, P., Running, S. W., Mar. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology* 93 (4), 211-228.
- Vanloocke, A., Bernacchi, C. J., Twine, T. E., Jul. 2010. The impacts of *Miscanthus* production on the Midwest U.S. hydrologic cycle. *GCB Bioenergy*, no-no.
- Verberne, E., Hassink, J., de Willigen, P., Groot, J., , Van Veen, J., 1990. Modelling organic matter dynamics in different soils. *Netherlands J. Agric. Sci* 38, 221-238.
- Wang, D., Lebauer, D. S., Dietze, M. C., Feb. 2010. A quantitative review comparing the yield of switchgrass in monocultures and mixtures in relation to climate and management factors. *GCB Bioenergy* 2 (1), 16-25.
- Wikle, C., Berliner, L., 2007. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena* 230 (1-2), 1-16.
- Wulschleger, S. D., Davis, E. B., Borsuk, M. E., Gunderson, C. a., Lynd, L. R., Jul. 2010. Biomass Production in Switchgrass across the United States: Database Description and Determinants of Yield. *Agronomy Journal* 102 (4), 1158-1168.