GLOBAL CHANGE BIOLOGY
BIOENERGY

GCB Bioenergy (2010) **2**, 248–257, doi: 10.1111/j.1757-1707.2010.01059.x

Empirical geographic modeling of switchgrass yields in the United States

HENRIETTE I. JAGER*, LATHA M. BASKARAN*, CRAIG C. BRANDT†, ETHAN B. DAVIS‡, CARLA A. GUNDERSON* and STAN D. WULLSCHLEGER*

**Oak Ridge National Laboratory, Environmental Sciences Division, Oak Ridge, TN, USA, †Oak Ridge National Laboratory, Biosciences Division, Oak Ridge, TN, USA, ‡Dartmouth College, Thayer School of Engineering, Hanover, NH, USA*

Abstract

Switchgrass (*Panicum virgatum* L.) is a perennial grass native to the United States that has been studied as a sustainable source of biomass fuel. Although many field-scale studies have examined the potential of this grass as a bioenergy crop, these studies have not been integrated. In this study, we present an empirical model for switchgrass yield and use this model to predict yield for the conterminous United States. We added environmental covariates to assembled yield data from field trials based on geographic location. We developed empirical models based on these data. The resulting empirical models, which account for spatial autocorrelation in the field data, provide the ability to estimate yield from factors associated with climate, soils, and management for both lowland and upland varieties of switchgrass. Yields of both ecotypes showed quadratic responses to temperature, increased with precipitation and minimum winter temperature, and decreased with stand age. Only the upland ecotype showed a positive response to our index of soil wetness and only the lowland ecotype showed a positive response to fertilizer. We view this empirical modeling effort, not as an alternative to mechanistic plant-growth modeling, but rather as a first step in the process of functional validation that will compare patterns produced by the models with those found in data. For the upland variety, the correlation between measured yields and yields predicted by empirical models was 0.62 for the training subset and 0.58 for the test subset. For the lowland variety, the correlation was 0.46 for the training subset and 0.19 for the test subset. Because considerable variation in yield remains unexplained, it will be important in the future to characterize spatial and local sources of uncertainty associated with empirical yield estimates.

Keywords: bioenergy, functional validation, mapping, mixed models, *Panicum virgatum*, spatial modeling, switchgrass

Received 10 May 2010 and accepted 11 June 2010

Introduction

Dedicated bioenergy crops are being promoted in the United States and abroad as renewable alternative feedstocks to conventional petroleum energy supplies (Lewandowski *et al*., 2003; Ragauskas *et al*., 2006). Transportation fuels, like ethanol, derived from cellulosic plant biomass could benefit economic growth, enhance energy security, reduce greenhouse gas emissions and mitigate the potential impacts of global climate change (Kheshgi *et al*., 2000; Smith *et al*., 2000).

Correspondence: Henriette I. Jager, e-mail: jagerhi@ornl.gov

Perennial bioenergy feedstocks, such as native grasses and trees, are considered one of the most sustainable sources of renewable transportation fuel because they produce large amounts of biomass, require limited input of water and nutrients, and minimize ecological damage to soils and rivers (Sanderson *et al*., 1996; McLaughlin & Walsh, 1998; Heaton *et al*., 2008). Switchgrass (*Panicum virgatum* L.), a native warm-season grass found in grasslands of the eastern United States (McLaughlin & Kszos, 2005), is one perennial plant under intensive study as a possible bioenergy feedstock. It is a widespread component of the native North American tall grass prairie with a range of adaptation from Nova Scotia, Ontario, and Maine to

North Dakota and Wyoming, south to Florida, Nevada, and Arizona, and into Mexico and Central America (Hitchcock, 1971). Across this range, switchgrass populations exist either as upland or lowland ecotypes that differ in habitat preference, morphology, and productivity.

There is great interest in predicting biological, environmental, and geographic variation in yields for perennial bioenergy crops (Heaton *et al*., 2004). Two types of models can be used to predict yields: mechanistic plant-growth models and empirical models based on field data. For switchgrass, only plant-growth models have historically been used. Various general purpose plant-growth models, such as EPIC (Brown *et al*., 2000, Thomson *et al*., 2009), ALMANAC (Kiniry *et al*., 1996, 2005), and SWAT (Nelson *et al*., 2006; Baskaran *et al*., 2009), have been used to predict switchgrass. Grassini *et al*. (2009) published a model specifically developed for switchgrass. Predictions from these models have been validated against field data collected from a limited geographic range under uniform management conditions. Plant-growth models are extremely valuable, particularly for applications that require extrapolating beyond climate conditions currently experienced by switchgrass.

Empirical models also play an important role. One extreme view advocates the exclusive use of empirical models based directly on field measurements (Peters, 1980). In our view, empirical models, based on data collected over a wide geographic area under diverse management conditions, are needed to understand what responses to environmental gradients mechanistic models should be expected to reproduce. In the functional validation approach developed by Jager *et al*. (2000), discrepancies between empirical and mechanistic model responses are used to suggest future improvements in mechanistic models. Empirical models are the starting point for a functional validation approach.

The purpose of this study was to develop empirical models to describe relationships between switchgrass yield and environmental covariates. A second role was to use the empirical models to predict switchgrass yield for the conterminous United States. Our empirical modeling efforts built on the wealth of field trials reported in the open literature from site-specific variety trials conducted across the United States over the past two decades (Davis, 2007; Gunderson *et al*., 2008). In this study, we described empirical responses of yield to environmental covariates and management practices and differences in responses of lowland and upland varieties of switchgrass. In addition, we characterized the residual unexplained variation in switchgrass yield. These empirical models can now be used for functional validation of mechanistic plant-growth models and as input to other models that require yield predictions. Our results are presented spatially for the eastern United States and can be used to assess the implications of our findings for regional and national biomass supply.

Materials and methods

Data

Published field studies of switchgrass yield were compiled from numerous literature sources (Davis, 2007; Gunderson *et al*., 2008). Following Gunderson *et al*. (2008), we excluded field studies growing a mixture of ecotypes in order to estimate yields specific to switchgrass ecotypes. Studies of harvest frequency have produced contradictory results (Sanderson *et al*. 1996; Thomason *et al*., 2004; Fike *et al*., 2006), but they concur that yields are lower when harvest frequency exceeds three times per year. We excluded first-year harvests because these are typically lower than those in subsequent years (Fike *et al*., 2006; Gunderson *et al*., 2008) and include cases of failure during establishment. Similarly, we excluded trials that experienced catastrophic failures, as indicated by yields $<1\,\mathrm{Mg\,ha^{-1}}$ dry weight (Gunderson *et al*., 2008). Studies included both those that did and did not irrigate during establishment, as yield was measured during later years.

For the lowland ecotype, field trials were available at 28 locations ranging in latitude from Texas to New Jersey (Table 1). For the upland ecotype, data from more field trials were available in northern locations (Montreal, Canada, North, and South Dakota), and fewer trials were available at southern locations (Louisiana, Texas, and Oklahoma) (Table 1). Our approach to obtaining covariates was to rely on geospatial databases. This was necessary because climate and soils information were not consistently reported across studies. Climate variables used as predictors were obtained from the nearest orographically corrected PRISM climate gridpoint (Daly *et al*., 1994; Table 1). Soils data (depth to bedrock and % sand) were obtained from the State Soil Geographic Database (STATSGO, USDA Soil Conservation Service, 1992). For each field observation of switchgrass yield, we determined location-specific minimum winter temperature ($^{\circ}$C) (T_{min}), average temperature ($^{\circ}$C) for April–September of the year of harvest (T_{avg}), total April–September precipitation (cm) during the year of harvest (P_{tot}), total nitrogen fertilizer (kg $\mathrm{ha^{-1}}$) applied (N_{tot}), an indicator variable set to one if fertilizer was applied (*IsFert*) and zero otherwise, depth to bedrock (D_{rock}) in m, number of harvests per year (*HarvFreq*), stand age (*Age*) in years, and an index of soil

Table 1 Description of field trial locations

Location	Latitude (°N)	Longitude (°W)	Climate station	Lowland field trials	Upland field trails	References
Arlington, WI	43.33	89.38	476 718	1 (1)	22(2)	Casler & Boe (2003)
Arlington, WI	43.33	96.48	391 392	1 (1)	2	Casler *et al.* (2004)
Athens, GA	33.87	83.41	092 318	8 (2)	18(2)	Bouton (2002)
Beeville, TX	28.44	97.80	410 639	23 (2)	1(1)	Kiniry *et al.* (1996), Sanderson *et al.* (1999a), Muir *et al.* (2001)
Blacksburg, VA	37.18	80.42	440 766	22 (2)	22(2)	Fike *et al.* (2006)
Brookings, SD	44.30	97.00	398 932	0 (0)	22(2)	Casler & Boe (2003)
Chariton, IA	40.97	93.43	134063	10 (2)	46(2)	Lemus *et al.* (2002)
Chickasha, OK	35.05; 35.04	97.91	341 504	114 (2)	26(2)	Fuentes & Taliaferro (2002), Thomason *et al.* (2004)
Clinton, LA	30.85	90.05	162 151	4 (2)	0	Cassida *et al.* (2005)
College Station, TX	30.60; 30.60; 30.67	96.35	411 048	22 (2)	15(2)	Kiniry *et al.* (1996), Sanderson *et al.* (1999a), Cassida *et al.* (2005)
Dallas, TX	32.75; 32.97	97.27	419 532	70 (2)	8(2)	Kiniry *et al.* (1996), Sanderson *et al.* (1999a)
Dickinson, ND	46.88	102.80	322 188	0	22 (2)	Berdahl *et al.* (2005)
Haskell, OK	35.75	95.64	346 130	26 (2)	40 (2)	Fuentes & Taliaferro (2002)
Hope, AR	33.67	93.58	035 908	4 (2)	1 (1)	Cassida *et al.* (2005)
Jackson, TN	35.88	88.83	404 561	10 (2)	10 (2)	Fike *et al.* (2006)
Knoxville, TN	35.88	83.99	406 534	10 (2)	10 (2)	Fike *et al.* (2006)
Mandan, ND	46.80	100.92	325479	0	46 (2)	Berdahl *et al.* (2005)
Mead, NE	41.22	96.48	250 375	1 (1)	1 (1)	Casler *et al.* (2004)
Montreal, Canada	45.42	73.88	301 966	0	31 (2)	Madakadze *et al.* (1998), Madakadze *et al.* (1999)
Morgantown, WV	39.62	79.95	369 050	10 (2)	10 (2)	Fike *et al.* (2006)
Orange, VA	38.22	78.12	446 712	10 (2)	9 (2)	Fike *et al.* (2006)
Perkins, OK	35.99	97.05	348 501	44 (2)	0	Thomason *et al.* (2004)
Princeton, KY	37.10	87.82	153 994	10 (2)	10 (2)	Fike *et al.* (2006)
Raleigh, NC	35.72	78.67	317 994	10 (2)	10 (2)	Fike *et al.* (2006)
Rock Springs, PA	40.72	77.94	368 449	0	8 (2)	Sanderson *et al.* (2004)
Shorter, AL	32.66	85.56	099 291	4 (2)	8 (2)	Sladden *et al.* (1991)
Spooner, WI	45.22	89.38	470 991	1 (1)	1 (1)	Casler *et al.* (2004)
Stephenville, TX	32.22	98.20	412 598	107 (2)	23 (2)	Kiniry *et al.* (1996), Sanderson *et al.* (1999a), Cassida *et al.* (2005)
Stillwater, OK	36.12	96.58	348 501	1 (1)	1 (1)	Casler *et al.* (2004)
Temple, TX	31.05	97.34	418 910	54 (1)	9 (2)	Sanderson *et al.* (1999a)
Tifton, GA	31.47	83.49	098 703	8 (2)	18 (2)	Bouton (2002)

Latitude and longitude are in decimal degrees. PRISM climate stations identifiers are given for each location. We list the number of average yield values reported for each of the two ecotypes, with the number of observations in the test subset in parentheses.

wetness (*WetSoil*) calculated as (100-% sand) $\times P_{tot}$, Our soil wetness index represents an interaction between temporally variable precipitation and the percentage of sand (constant for each location). Soils with a lower percentage of sand have a higher water holding capacity, which has implications for yield even at the same level of precipitation (Evers & Parsons, 2003, Parrish & Fike, 2005).

Empirical models

We estimated average yield for each ecotype using generalized logistic regression. We applied a logit transform to average yield, $LYield = \log(\mathrm{Yield}/\mathrm{Yield}_{\max})/[1-\log(\mathrm{Yield}/\mathrm{Yield}_{\max})]$, to ensure that mapped values would not exceed those represented in the data. The maximum yield ($\mathrm{Yield}_{\max}$) for the upland ecotype was 28 and 40 Mg ha^{-1} dry weight for the lowland ecotype. The full model included both climatic and nonclimatic covariates [Eqn (1)]. *LYield* is expressed as a linear function of variables defined in 'Data', with coefficients v_1 to v_{11} and intercept, v_0. The model for residual error, å, indicates that it is assumed to be normally distributed with variance–covariance matrix, **C**.

Fullmodel

$$LYield = v_0 + v_1\,T + v_2T^2 + v_3T_{\min} + v_4P + v_5TP + v_6\,WetSoil + v_7Age + v_8HarvFreq + v_9N_{\mathrm{tot}} + v_{10}D_{\mathrm{rock}} + v_{11}IsFert + \varepsilon(i,j)$$

$$\varepsilon \sim N(\mathbf{0},\mathbf{C}), \quad \text{where } c_{ij} = \begin{cases} r+r, & i=j \\ r, & L(i)=L(j) \\ 0, & L(i)L(j) \end{cases},$$

for $L(i)$, the location *of field trial i*.

Because several of the field trials provided multiple estimates of switchgrass yield at a given location, it was important to account for within-location correlation. Our model assumes independence between locations but nonzero correlations within yield measurements taken at the same location, *i*. This error structure is described by a compound symmetric variance–covariance model, which has block-diagonal variance–covariance matrix ***C*** of the errors, å [Eqns (2) and (3)]. Within-location correlation, ñ, is estimated for nonzero blocks. Data limitations prevented us from estimating location-specific fixed effects.

$$\varepsilon \sim N\{\mathbf{0},\mathbf{C}\},\ \mathbf{C} = \begin{pmatrix} \mathbf{C_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C_n} \end{pmatrix}, \qquad (2)$$

for location $l = 1, n$,

$$C_l = \begin{pmatrix} r+\rho & \rho & \rho & \cdots & \rho \\ \rho & r+\rho & \rho & \cdots & \rho \\ \rho & \rho & r+\rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & r+\rho \end{pmatrix} \qquad (3)$$

(diagonal block for location l).

We first fitted a full model including all predictors using the 'nlme' package in R (Pinheiro & Bates, 2000). We then fitted reduced models (models with fewer predictor variables) for upland and lowland varieties to be used in mapping. For each ecotype, we selected a reduced model to include only those predictors that were significant at $\propto = 0.1$.

Model selection

We divided our data into two parts: a subset for parameter estimation and a test subset for evaluating goodness-of-fit. Because we wished to consider correlations among measurements from the same location, we stratified the sample by ecotype and location. We selected two measurements for the test subset at random from each stratum, except in cases where only one was available. We used the estimation subset of data to estimate parameters. We then assessed goodness-of-fit of each model by fitting each to the test dataset, which represents approximately 10% of the total data available. Predicted yields were obtained for the test subset by back-transforming logit-transformed estimates based on reduced models in Table 2. The training (test) subset of data in the reduced models included 600 (48) lowland and 459 (55) upland observations.

We evaluated alternative models using both goodness-of-fit and information-theoretic. We reported two goodness-of-fit criteria: residual standard error and Pearson's correlations between predicted and observed yields for both the training and test data subsets. We also reported Akaike's information criterion (AIC). A model with a lower AIC should be preferred over alternative models with higher AIC, even if its goodness-of-fit is poorer. This is because AIC penalizes for over fitting to a particular dataset by including excessive number of predictors and favors models more likely to perform well with new datasets (Burnham & Anderson, 2002).

Residual analysis

For each model, we examined the distribution of residuals to determine whether the mean was significantly different from zero. We also regressed the predicted

Table 2 Parameter estimates including coefficients and two parameters describing compound symmetry in residual error for the full models on the left

Parameter		Full generalized least square models				Reduced parameter models			
		Upland		Lowland		Upland		Lowland	
		Estimate	*P*-value	Estimate	*P*-value	Estimate	*P*-value	Estimate	*P*-value
v_0	Intercept	−5.5235	0.0560	−23.7305	0.0001	−7.4184	0.0045	−23.2001	0.0002
v_1	T_{avg}	0.4647	0.0855	1.9827	<0.0001	0.5906	0.0152	1.9960	<0.0001
v_2	T^2_{avg}	−0.0107	0.0906	−0.0435	<0.0001	−0.0136	0.0173	−0.0437	<0.0001
v_3	T_{min}	0.0626	0.0007	0.0617	0.0485	0.0660	0.0006	0.0519	0.0863
v_4	P_{tot}	0.0625	0.0001	0.1021	0.0031	0.0653	0.0000	0.1102	0.0013
v_5	$T_{avg} \times P_{tot}$	−0.00336	0.0007	−0.0045	0.0015	−0.0035	0.0000	−0.0047	0.0011
v_6	WetSoil	0.00019	0.0061	0.00007	0.2634	0.0002	0.0044		
v_7	Age	−0.0655	0.0191	−0.0642	0.0009	−0.0504	0.0019	−0.0616	0.0012
v_8	HarvFreq	0.4634	<0.0001	−0.0650	0.1150	0.4400	<0.0001	−0.0698	0.0897
v_9	N_{tot}	0.0007	0.7194	0.0006	0.0003			0.0007	<0.0001
v_{10}	D_{rock}	−0.0003	0.9492	0.0044	0.4207				
v_{11}	IsFert	−0.6199	0.4194	0.1758	0.1952				
r	MSE	0.7556		0.7931		0.6975		0.7971	
$\tilde{n}$	Location	0.5787		0.3963		0.4843		0.4026	
	Total df	451		585		458		599	
	Residual df	439		573		448		576	
	AIC	822.4		1261.8		823.1		1230.8	

Estimates for the reduced models used in predicting potential switchgrass yields are shown on the right. Predictors are location-specific average temperature (T_{avg}) for April–September in the year of harvest, minimum (T_{min}) winter temperature (°C), total April–September precipitation (cm) during the year of harvest (P_{tot}), an index of soil wetness (WetSoil), total nitrogen fertilizer (kg/ha) applied (N_{tot}), an indicator variable for fertilizer application (IsFert), depth (m) to bedrock (D_{rock}), number of harvests per year (HarvFreq), and stand age (Age) in years.

values against observed and compared these visually for each switchgrass ecotype.

Mapping analysis

The purpose of our mapping analysis was to use the empirical model developed (i.e., reduced models) to estimate potential switchgrass yields in geographic locations where no data were collected, without extrapolating beyond the range of climate values represented. This is not meant to imply that the current crop- or land-cover would in actuality be supplanted by switchgrass, but rather indicates expected yields, according to the empirical models, based on climate and soils. We will refer to these models as 'mapping versions'. Our data included field trials conducted at winter temperatures between −17 and 8 °C, mean growing season temperatures between 13.8 and 27 °C, and >310 mm total growing season precipitation. In the mapping analysis, we masked out regions of the United States with more extreme values. For management variables, which are not intrinsically spatial, we assumed fixed values. For the lowland ecotype model, which included N_{tot}, we assumed switchgrass would be fertilized with 80 kg N ha^{-1}. For both ecotypes, we used a stand age of 4 years. Fike *et al*. (2006) found that upland varieties produce higher yields with two harvests than with one. We therefore set harvest frequency in a way that is optimal for each ecotype: one harvest per year for lowland and two harvests per year for upland varieties.

Results

The full and reduced models explained a significant amount of the variability in switchgrass yield for both the upland and lowland varieties. Yield showed the expected uni-modal response to average growing season temperature, with a significant positive coefficient for T_{avg} and a negative coefficient for the quadratic temperature term to lower yields at high temperatures (positive v_1 and negative v_2 in Table 2). Both ecotypes showed a positive response to minimum winter temperature. Both ecotypes showed a positive response to precipitation and both had a significant negative interaction between precipitation and temperature (v_4 and v_5 in Table 2). Lowland varieties showed stronger responses to average temperature than upland varieties.

We considered two soil-related variables (v_6 and v_{10} in Table 2). Yield showed a significant positive response to our soil moisture index (*WetSoil*) for the upland, but not the lowland, variety. Depth to bedrock (D_{rock}) was not a significant predictor of yield for either ecotype, and was excluded from the reduced models.

The full models included four management-related variables: stand age, number of harvests per year, an indicator variable for fertilization, and total nitrogen. Of these, only the lowland ecotype showed a positive response to total nitrogen. The remaining predictors were not significant and were excluded from the reduced models (Table 2).

Correlations between yields from field trials in the same location, *c*, were significantly greater than zero in the final, reduced models (Table 2). Note that the number of observations increased slightly (total degrees of freedom $+1$ in Table 2) in the reduced models because observations that had missing values for predictors were removed could be used in the analysis.

Model selection

All Pearson's correlations between predicted and observed values (back-transformed to $Mg\,ha^{-1}$) were highly significant. For the upland variety, the correlation was 0.6190 (95% CI = [0.5591, 0.6725], df = 456, $P<0.0001$) for the training subset and 0.5795 (95% CI = [0.3690, 0.7335], df = 52, $P<0.0001$) for the test subset. For the lowland variety, the correlation between predicted and observed yield was 0.4596 (95% CI = [0.3932, 0.5213], df = 583, $P<0.0001$) for the training subset and 0.1851 (95% CI = [−0.1111, 0.4511], df = 44, $P = 0.22$) for the test subset. Correlations are usually lower for the test subset than for the data used to develop the model.

Residual analysis

The median difference between measured and predicted switchgrass yield was $0.081\,Mg\,ha^{-1}$ (range -2.9758 to $3.734\,Mg\,ha^{-1}$) for lowland and $0.0718\,Mg\,ha^{-1}$ (range -2.941 to $3.678\,Mg\,ha^{-1}$) for upland varieties. For the upland variety, the reduced model produced a mean residual standard error of 0.6975, with standardized residuals between −2.98 and 3.73 SD and an interquartile range of (−0.56 to 0.71). For the lowland variety, the reduced model had a mean residual standard error of 0.7971, with standardized residuals between −2.74 and 5.48 SD and an interquartile range of (−0.59 to 0.55). Lowland values with magnitudes greater than three were evaluated as potential outliers.

A simple least-squares regression showed significant positive relationships between measured and predicted switchgrass yields (Fig. 1), although a great deal of scatter remained. The largest deviations were predictions of the highest lowland yields, which were underpredicted by the reduced model (Fig. 1a). We had no other reason to remove these observations as putative outliers.

Mapping analysis

The mapping version of the reduced models above showed the expected gradient of higher yields in the eastern United States and lower yields in the western United States (Fig. 2a). Note that we excluded grey areas from prediction because the predictors fell outside the observed range in Fig. 2. The highest predicted lowland yields were centered on the three-state junction of Tennessee, North Carolina, and Georgia, with lower predictions moving outward from this junction (Fig. 2a). High yields were also predicted throughout the states of Illinois, Kentucky, and Virginia. Low yields were predicted in the far west, the Gulf coast, and at higher latitudes of New York and Michigan (Fig. 2a). Interestingly, moderately high yields were predicted in

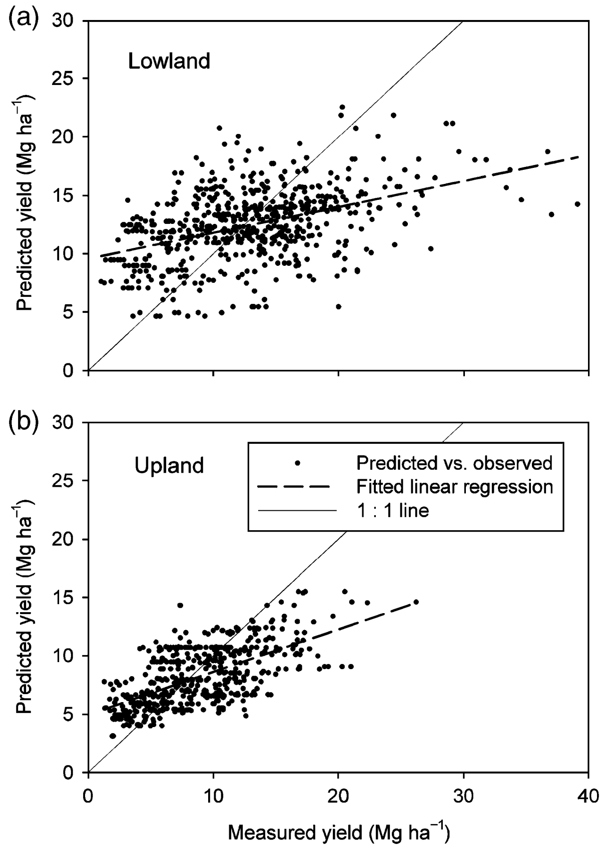

Fig. 1 Relationships between yields predicted by the reduced models and measured yield for the (a) lowland and (b) upland ecotypes of switchgrass.

Published 2010

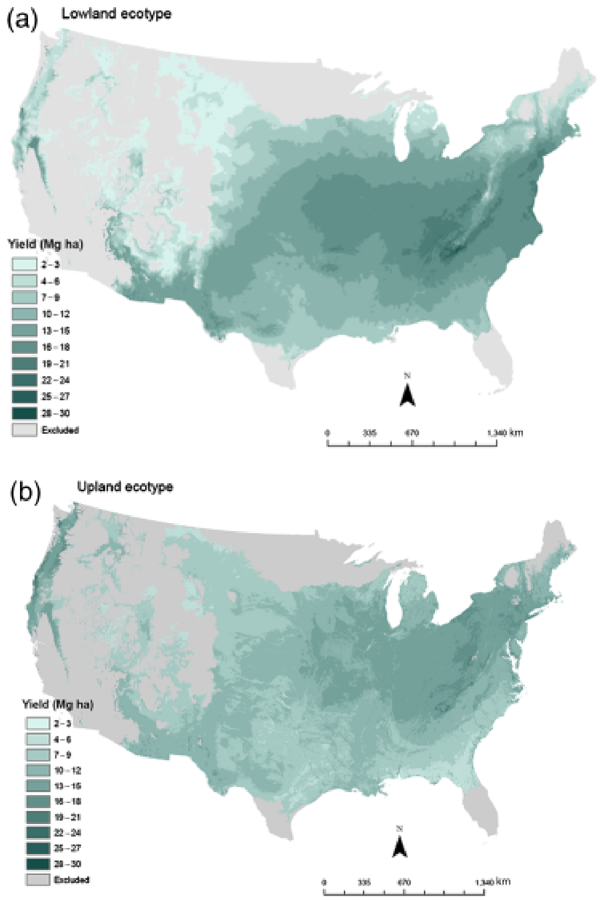

Fig. 2 Maps of predicted potential switchgrass yield for (a) lowland and (b) upland ecotypes using the mapping versions of the reduced empirical models. Areas with climate outside the range represented by field trials were excluded.

some isolated pockets of the Sierra-Nevada Mountains, areas outside the natural range for switchgrass (Fig. 2a). Maps of estimates represent potential yield on lands available for planting switchgrass and do not suggest that switchgrass will replace existing land cover.

Predicted upland yields were generally lower than lowland yields. Upland yields were higher than lowland yields in many areas of the western United States and at high latitudes, including northern Michigan, Wisconsin, and Maine (Fig. 2b). The highest upland yields were centered near the three-state junction of West Virginia, Kentucky, and Ohio (Fig. 2b).

Discussion

The generalized logistic model presented here provides a means of estimating switchgrass yields in different locations based on local climate, soil conditions and management choices. In this study, we found that yields in field trials of the lowland ecotype were generally higher than yields of the upland ecotype. Both ecotypes showed a quadratic response to average temperature. The lowland ecotype showed a more-significant positive response to minimum winter temperature. This is expected since this ecotype does not do as well at high latitudes (Casler *et al*., 2004). Precipitation was strongly correlated with yields of both ecotypes. Lee & Boe (2005) noted a strong precipitation response for upland varieties in North Dakota. Only the upland ecotype showed a significant response to our soil moisture index (Table 2). It has been suggested that replacing SSURGO-derived or locally measured soil water holding capacity for % sand in our soil wetness might improve the skill of this predictor.

Despite removing data for field trials during the first year of establishment, we found a negative response to stand age that was significant for both ecotypes, suggesting a decline in yield with age after several years of harvest, as noted by Lee & Boe (2005) for upland varieties. Fertilizer application had a positive effect on lowland, but not upland, yield. Other studies have also shown a positive effect of nitrogen for upland (Madakadze *et al*., 1999) and lowland varieties (Sanderson & Reed, 2000), but with diminishing returns. Sanderson & Reed (2000) reported that fertilizer was not beneficial during the establishment year. Bedrock depth was not identified as an important predictor of yield, perhaps because few field trials were conducted in shallow soils.

Spatial patterns predicted by the mapping versions of the empirical models seem to deviate most from expectations on the western and northern margins of the natural range for switchgrass. This highlights how important it is to collect field data from sites with marginal conditions, which provide more information than data from sites with ideal conditions for use in both empirical and process yield models. In drier western areas, predictions for lowland yield based on the empirical model appear higher than expected. For example, our results indicate lowland yields of 10–15 Mg ha^{-1} in the Big Bend region along the United States–Mexico border in western Texas. According to Sanderson *et al*. (1999b), the high-yielding lowland 'Alamo' variety would likely not perform well in western Texas where annual rainfall is $<$50 cm. Likewise, predicted yields of 5–10 Mg ha^{-1} in the semi-arid rangelands of SD, WY, and CO are higher than expected. Baskaran *et al*. (2009) found the largest deviations between SWAT-model predictions for Alamo switchgrass and those of the lowland mapping model in the southwest and between the latitudes of 41° and 43° and east of the Dakota's. Additional trials in these western areas are needed to better define productivity in more arid environments. Although this study made a special effort to identify and include sites as far north as Montreal, Canada in order to better represent yields at high latitudes, trials in more northern locations are needed to better define yields for lowland and for upland varieties at higher latitudes (Casler *et al*., 2004). In these areas, where it is necessary to extrapolate to new conditions, estimating yields using process-based models is probably a better alternative.

Previous studies have used mechanistic plant-growth models to predict switchgrass yield. Kiniry (1996) was able to explain 76% of variation in yield at five sites in one state (Texas) using the ALMANAC model. However, in a later comparison, Kiniry *et al*. (2005) was able to explain only 47% of variation among five locations in the south. ALMANAC performed well in explaining variation among locations, but not as well in explaining year-to-year variation within yield. We also found that temporal variation within-location were the most difficult to predict. This suggests to us that attributes shared by trials at the same location, such as soils properties, are unlikely to improve predictions. Grassini *et al*. (2009) also developed a plant-growth model for switchgrass and compared predictions for 10 years at six sites both in the far northern and southern range of the Midwestern United States. Aboveground biomass predictions were within 15% of reported values. The EPIC model was used by Thomson *et al*. (2009) to simulate switchgrass yields over a larger region (for the conterminous United States). Spatially, their predictions showed some similarities with results presented here, with both predicting low values in the west. But the two studies also showed some differences in geographic patterns. EPIC predicted high yields in Florida, along the Gulf coast of Texas and Louisiana and the coast of North Carolina. Our empirical model predicted lower yields in these areas. Calibration was conducted for seven locations in the southeast and overall validation statistics were not reported. We caution that comparing R^2 values obtained by comparing observed and predicted values from different plant-growth models or from empirical models is a questionable practice, due to differences in the numbers of parameters involved. Generally speaking, it would be best to report such statistics for new 'test' data (locations, years) not used in calibration.

In our view, the most important contribution of the empirical relationships identified here is to serve as a basis for evaluating and improving mechanistic plant-growth models for switchgrass. Understanding where relationships between mechanistic models and their drivers fail to reproduce those observed in nature is a more constructive approach to validation than simply comparing the values themselves (Jager *et al*., 2000). Baskaran *et al*. (2009) compared SWAT-predicted yields for Alamo switchgrass, a lowland variety, with those predicted by mapping version of the lowland empirical model. A regression between SWAT-predicted and empirical model yields gave an R^2 of 0.51. However, on average, lowland yields predicted by the empirical model (Fig. 2a) tended to be higher than those of the SWAT model. As discussed earlier, the empirical model for the lowland ecotype predicts much higher yields on the southwestern and northern margins.

We have several suggestions for future data collection to facilitate regional assessments. First, seasonal timing of harvest has a well-known effect on yield. It would be useful if future studies could report local measurements of temperature and rainfall. Reporting yields by year, instead of reporting averages across multiple years, would also increase the usefulness of data reported in the literature by allowing matching to the relevant local conditions. Reporting relevant local soil attributes, such as water holding capacity, depth to bedrock, slope, and elevation would be useful. Reporting precise field locations is important as it can improve associations of yields with available geospatial data. It would be helpful to include future trials from a much wider range of locations and conditions. For example, yield data are needed for sites farther west, at higher elevations and slopes, shallower soils, and under less-than-ideal conditions for growth.

The empirical estimates provided by this study can be used to facilitate functional validation of plant-growth models. Results from the best-available yield models, whether empirical or mechanistic, are needed as input to other regional models used in bioenergy assessments. For example, economic models that estimate changes in land use require estimates of the relative profitability of growing switchgrass instead of other crops. Best-available regional yield estimates are also needed by models to identify optimal locations for siting biorefineries (e.g., Graham *et al*., 2000).

In future, we hope to have the opportunity to quantify the uncertainty associated with our model predictions. Because the uncertainty varies spatially, quantifying spatial uncertainties associated with yield estimates is important to any decision-making process that relies on the models presented here. Estimation of prediction errors for generalized least squares models are not provided as part of existing statistical software such as RS nlme package (Pinheiro & Bates, 2000) or SAS© Proc Mixed (Littell *et al*., 1996), but can be accomplished by resampling of the residuals. In situations such as this, where a fair amount of unexplained variance in yields remains, presenting visual maps of spatial uncertainty along with predictions is especially important.

Acknowledgements

This research was funded, in part, by the Department of Energy Office of Biomass Programs. ORNL is managed by UT-Battelle, LLC for the USDOE under contract DE-AC05-00OR22725. We thank Nadia Ally for spending a summer perusing the literature for switchgrass data from higher latitudes. Bob Perlack deserves a great deal of credit for supporting and shepherding this research and sharing his expertise on switchgrass yields. We appreciate Robin Graham for her support and review of this manuscript. Finally, Tris West is responsible for significant improvements in this manuscript and we thank him for organizing this special symposium.

References

Baskaran LM, Jager HI, Schweizer PE, Srinivasan R (2009) Use of the SWAT model to evaluate the sustainability of bioenergy production at a National scale. Proceedings of the 2009 5th International SWAT Conference, August 5–7, 2009, Boulder, Colorado.

Berdahl JD, Frank AB, Krupinsky JM, Carr PM, Hanson JD, Johnson HA (2005) Biomass yield, phenology and survival of diverse cultivars and experimental strains in western North Dakota. *Agronomy Journal*, **97**, 549–555.

Bouton JH (2002) *Bioenergy crop breeding and production research in the southeast*. Final report for 1996 to 2001. ORNL/SUB-02-19XSV810C/01. Available at http://www.osti.gov/bridge/ (accessed 7 January 2010).

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York, NY, USA.

Brown RA, Rosenberg NJ, Hays CJ, Easterling WE, Mearns LO (2000) Potential production and environmental effects of switchgrass and traditional crops under current and greenhouse-altered climate in the central United States: a simulation study. *Agricultural Ecosystems and Environment*, **78**, 31–47.

Casler MD, Boe AR (2003) Cultivar × environment interactions in switchgrass. *Crop Science*, **43**, 2226–2233.

Casler MD, Vogel KP, Taliaferro CM, Wynia RL (2004) Latitudinal adaptation of switchgrass populations. *Crop Science*, **44**, 293–303.

Cassida KA, Muir JP, Hussey MA, Read JC, Venuto BC, Ocumpaugh WR (2005) Biomass yield and stand characteristics of switchgrass in south central U.S. environments. *Crop Science*, **45**, 673–681.

Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountanious terrain. *Journal of Applied Meteorology*, **33**, 140–158.

Davis E (2007) *Evaluating potential U.S. cellulosic feedstocks and ethanol production as a substitute for petroleum and gasoline*. Master of Science Thesis. Sustainable Energy Planning and Management, Aalborg University, Denmark.

Evers EW, Parsons MJ (2003) Soil type and moisture level influence on Alamo switchgrass emergence and seedling growth. *Crop Science*, **43**, 288–294.

Fike JH, Parrish DJ, Wolf DD, Balasko JA, Green JT, Rasnake M, Reynolds JH (2006) Switchgrass production for the upper southeastern USA: influence of cultivar and cutting frequency on biomass yields. *Biomass and Bioenergy*, **30**, 207–213.

Fuentes RC, Taliaferro CM (2002) Biomass yield stability of switchgrass cultivars. In: *Trends in New Crops and New Uses* (eds Janick J, Whipkey A), pp. 276–282. American Society for Horticultural Sciences (ASHS) Press, Alexandria, VA. Available at http://www.hort.purdue.edu/newcrop/ncnu02/v5-toc.html#biomass

Graham RL, English BC, Noon CE (2000) A geographic information system-based modeling system for evaluating the cost of delivered energy crop feedstock. *Biomass and Bioenergy*, **18**, 309–329.

Grassini P, Hunt E, Mitchell RB, Weiss A (2009) Simulating switchgrass growth and development under potential and water-limiting conditions. *Journal of Agronomy*, **101**, 564–571.

Gunderson CA, Davis EB, Jager HI *et al*. (2008) *Exploring potential U.S. switchgrass production for cellulosic ethanol*, ORNL/TM-2007/183. Available at OSTI #936551.

Heaton E, Voigt T, Long SP (2004) A quantitative review comparing the yields of two candidates C-4 perennial biomass crops in relation to nitrogen, temperature and water. *Biomass and Bioenergy*, **27**, 21–30.

Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of *Miscanthus*. *Global Change Biology*, **14**, 2000–2014.

Hitchcock AS (1971) *Manual of Grasses of the United States*, Vol. II. Dover Publications Inc., New York, NY.

Jager H, Hargrove WW, King AW, Olson RJ, Brandt CC, Scurlock JMO, Rose KA (2000) Constructive contrasts between modeled and measured climate responses over a regional scale. *Ecosystems*, **3**, 396–411.

Kheshgi H.S, Princes RC, Marland G (2000) The potential of biomass fuels in the context of global climate change: focus on transportation fuels. *Annual Review of Energy and the Environment*, **25**, 199–244.

Kiniry JR, Cassida KA, Hussey MA (2005) Switchgrass simulation by the ALMANAC model at diverse sites in the southern US. *Biomass and Bioenergy*, **29**, 419–425.

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

Lee DK, Boe A (2005) Biomass production of switchgrass in central South Dakota. *Crop Science*, **45**, 2583–2590.

Lemus R, Brummer EC, Moore KJ, Molstad NE, Burras CL, Barker MF (2002) Biomass yield and quality of 20 switchgrass populations in southern Iowa, USA. *Biomass and Bioenergy*, **23**, 433–442.

Lewandowski I, Scurlock JMO, Lindvall E, Christou M (2003) The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass and Bioenergy*, **25**, 335–361.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) *SAS System for Mixed Models*. SAS Institute Inc., Cary, NC.

Madakadze IC, Coulman BE, Peterson P, Stewart KA, Samson R, Smith DL (1998) Leaf area development, light interception, and yield among switchgrass populations in a short-season area. *Crop Science*, **38**, 827–834.

Madakadze IC, Stewart KA, Peterson PR, Coulman BE, Smith DL (1999) Cutting frequency and nitrogen fertilization effects on yield and nitrogen concentration of switchgrass in a short season area. *Crop Science*, **39**, 552–557.

McLaughlin SB, Kszos LA (2005) Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*, **28**, 515–535.

McLaughlin SB, Walsh ME (1998) Evaluating environmental consequences of producing herbaceous crops for bioenergy. *Biomass and Bioenergy*, **14**, 317–324.

Muir JP, Sanderson MA, Ocumpaugh WR, Jones RM, Reed RL (2001) Biomass production of 'Alamo' switchgrass in response to nitrogen, phosphorus, and row spacing. *Agronomy Journal*, **93**, 896–901.

Nelson RG, Ascough JC, Langemeier MR (2006) Environmental and economic analysis of switchgrass production for water quality improvement in northeast Kansas. *Journal of Environmental Management*, **79**, 336–347.

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

Peters RH (1980) Useful concepts for predictive ecology. *Synthese*, **43**, 257–269.

Pinheiro JC, Bates DM (2000) *Mixed-Effect Models in S and S-PLUS*. Springer-Verlag, New York, NY.

Ragauskas AJ, Williams CK, Davidson BH (2006) The path forward for biofuels and biomaterials. *Science*, **311**, 484–489.

Sanderson MA, Read JC, Reed RL (1999a) Harvest management of switchgrass for biomass feedstock and forage production. *Agronomy Journal*, **91**, 5–10.

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

Sanderson MA, Reed RL, Ocumpaugh WR (1999b) Switchgrass cultivars and germplasm for biomass feedstock production in Texas. *Bioresource Technology*, **67**, 209–219.

Sanderson MA, Reed RL (2000) Switchgrass growth and development: water, nitrogen, and plant density effects. *Journal of Range Management*, **53**, 221–227.

Sanderson MA, Schnabel RR, Curran WS, Stout WL, Genito D, Tracy B F (2004) Switchgrass and big bluestem hay, biomass, and seed yield response to fire and glyphosate treatment. *Agronomy Journal*, **96**, 1688–1692.

Sladden SE, Bransby DI, Aiken GE (1991) Biomass yield, composition and production costs for 8 switchgrass varieties in Alabama. *Biomass and Bioenergy*, **1**, 119–122.

Smith P, Powlson DS, Smith JU, Falloon P, Coleman K (2000) Meeting Europe's climate change commitments: quantitative estimates of the potential for carbon mitigation by agriculture. *Global Change Biology*, **6**, 525–539.

Thomason WE, Raun WR, Johnson GV, Taliaferro CM, Freeman KW, Wynn KJ, Mullen RW (2004) Switchgrass response to harvest frequency and time and rate of applied nitrogen. *Journal of Plant Nutrition*, **27**, 1199–1226.

Thomson AM, Izarrualde RC, West TO, Parrish DJ, Tyler DD, Williams JR (2009) *Simulating potential switchgrass production in the United States*, PNNL-19072, 20 pp.

USDA (United States Department of Agriculture, soil conservation service) (1992), STATSGO-State Soils Geographic Database, Soil Conservation Service, Washington, DC.

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

Published 2010
