Agriculture, Ecosystems and Environment 146 (2012) 121–129

Contents lists available at SciVerse ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Climatic and genetic controls of yields of switchgrass, a model bioenergy species

Mirela G. Tulbure[a,*], Michael C. Wimberly[a], Arvid Boe[b], Vance N. Owens[b]

[a] Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007-3510, USA
[b] Department of Plant Science, South Dakota State University, Brookings, SD 57007-3510, USA

ARTICLE INFO

Article history:
Received 20 June 2011
Received in revised form 16 August 2011
Accepted 18 October 2011
Available online 22 November 2011

Keywords:
Spatio-temporal variability
Marginal climates
GAMs
Biomass
Lignocellulosic biofuels

ABSTRACT

The U.S. Renewable Fuel Standard calls for 136 billion liters of renewable fuels production by 2022. Switchgrass (*Panicum virgatum* L.) has emerged as a leading candidate to be developed as a bioenergy feedstock. To reach biofuel production goals in a sustainable manner, more information is needed to characterize potential production rates of switchgrass. We used switchgrass yield data and general additive models (GAMs) to model lowland and upland switchgrass yield as nonlinear functions of climate and environmental variables. We used the GAMs and a 39-year climate dataset to assess the spatio-temporal variability in switchgrass yield due to climate variables alone. Variables associated with fertilizer application, genetics, precipitation, and management practices were the most important for explaining variability in switchgrass yield. The relationship of switchgrass yield with climate variables was different for upland than lowland cultivars. The spatio-temporal analysis showed that considerable variability in switchgrass yields can occur due to climate variables alone. The highest switchgrass yields with the lowest variability occurred primarily in the Corn Belt region, suggesting that prime cropland regions are the best suited for a constant and high switchgrass biomass yield. Given that much lignocellulosic feedstock production will likely occur in regions with less suitable climates for agriculture, interannual variability in yields should be expected and incorporated into operational planning.

1. Introduction

In recent years volatile oil prices have triggered great interest in biofuel production to reduce dependence on fossil fuels (Lynd et al., 1991). Biofuels, fuels produced from biomass, could not only reduce dependence on fossil fuels, enhance national energy security and provide an answer to increased energy demands, but also reduce greenhouse gas emissions and provide opportunities for rural economic development (Lynd et al., 1991; McLaughlin et al., 2002; Fargione et al., 2008; Charles, 2009). In the U.S., ethanol derived from starch-based feedstocks such as corn (*Zea mays* L.) grain is the most common biofuel, and corn acreages have expanded in recent years (Landis et al., 2008). However, competing food/feed and fuel demands have led to the development of "second generation biofuels", which are biofuels derived from lignocellulose of plant biomass (Antizar-Ladislao and Turrion-Gomez, 2008). Lignocellulosic biomass has long been recognized as a potential sustainable source of mixed sugars for fermentation to biofuels and other biomaterials (Lynd et al., 1991). Although concern has been expressed about the effects that biofuel production may have on indirect land use changes, food crop production, deforestation, water use patterns (Fargione et al., 2008; Searchinger et al., 2008) and biodiversity (Fletcher et al., 2011), the Energy Independence and Security Act of 2007 calls for the production of 136 billion liters of liquid biofuels in the United States by 2022, with 57 billion liters per year of corn-based ethanol and 80 billion liters of advanced biofuels derived from non-corn starch products (U.S. Congress, 2007).

Given the range of climate and soil types in the U.S., a combination of feedstock crops will be needed to provide a high amount of biomass that is constant over time (Wright, 1994). Switchgrass (*Panicum virgatum* L.), a warm-season perennial grass native to North America, has emerged as a leading candidate among herbaceous species to be developed as a bioenergy feedstock (McLaughlin and Walsh, 1998; Parrish and Fike, 2005). The choice of switchgrass was based on the fact that it produces high yields, has a wide geographic distribution from Mexico to Quebec and east of the Rocky Mountains, tolerates and grows well over a vast range of conditions, and has high nutrient and water use efficiency (McLaughlin and Walsh, 1998; Parrish and Fike, 2005). Moreover, switchgrass has several environmental benefits including soil erosion prevention, increased carbon sequestration, reduced water runoff and provision of habitat for wildlife (Parrish and Fike, 2005).

Throughout the geographic range of switchgrass there are two genetically and phenotypically distinct cytotypes: lowland, found in wetter and southern latitudes and upland, found in drier, higher

* Corresponding author. Present address: School of Environmental Science, Murdoch University, WA 6150, Australia. Tel.: +61 8 9360 6121.
E-mail address: M.Tulbure@murdoch.edu.au (M.G. Tulbure).

0167-8809/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.agee.2011.10.017

latitudes (Porter, 1966; Sanderson et al., 1996; Casler et al., 2004). Variation within cytotypes relates to the latitude of origin of cultivars (e.g., winter-hardiness), with lower latitude cultivars typically producing higher yields than upland cultivars, but not being able to tolerate winter cold if moved too far north (Hultquist et al., 1996; Parrish and Fike, 2005). Biomass yield across geographic locations depends on the ecotype and cultivar being planted, crop management, soil type, and climate (Parrish and Fike, 2005).

The U.S. Department of Energy's Bioenegy Feedstock Development Program (BFDP) has supported a wealth of field trial research to understand the best suited cultivars and crop management for sites at different locations (McLaughlin et al., 1999; McLaughlin and Kszos, 2005). This information at different geographic locations is invaluable, and work to compile and synthesize existing yield data and to develop techniques for modeling and mapping the capacity for biofuels feedstock production at a national level is ongoing (e.g., the Sun Grant Initiative). However, few studies have used existing data to evaluate switchgrass productivity across a broad geographic region. Heaton et al. (2004) used 77 observations to assess how productivity of switchgrass compares to that of *Miscanthus* × *giganteus* in relation to growing season precipitation, N fertilizer, and growing degree days. Switchgrass yields were positively influenced by N fertilizer and less by growing season precipitation, while temperature did not affect yield (Heaton et al., 2004). More recently, Wullschleger et al. (2010) compiled 1190 switchgrass yield observations and found that 30% of the variation in switchgrass yield was explained by growing season precipitation, annual temperature, N fertilization, and cytotype, all four factors having equal contribution. Wullschleger et al. (2010) recommended that future studies should extend the geographic distribution of field trials for a better understanding of switchgrass productivity as a function of genetic variability, climate, and crop management factors.

In order for the biofuels industry to be environmentally sustainable and economically viable, a high and constant supply of biomass is needed. It is important to quantify the spatial and temporal variability in biomass yield to ensure a reliable feedstock supply for biorefineries (Schmer et al., 2010) Thus, an understating of how environmental factors influence patterns of switchgrass biomass yields across space, and stability of yields through time, is essential. Therefore, the overall aim of this study was to ascertain the major drivers of variation in switchgrass yield at a broad geographic scale, focusing on climatic, genetic, edaphic, and management variables. Specific questions addressed were: (1) What are the most important variables influencing switchgrass yields (e.g., climate, other environmental variables, genetic factors)? (2) Do lowland cultivars have a more distinctive niche than upland cultivars (i.e., does the relationship of switchgrass yield with climate variables differ between lowland and upland cultivars?) (3) What spatial and temporal patterns of switchgrass yield are attributable to climatic variability?

2. Methods

2.1. Data

Switchgrass yield data from 1167 observations associated with 45 field trial locations where switchgrass was grown for biofuels production across fifteen U.S. states (ND, SD, IA, OK, TX, AK, LA, WI, WV, KY, TN, NC, AL, VA, NE) were used (Fig. 1). Yield data were gathered from 19 reference papers as well as personal communication with project investigators (Table 1). For every yield data point, we extracted information on site location (latitude, longitude), stand age, genetic variables (cultivar, cytotype, and origin of cultivar), month and season of harvest, amount of nitrogen fertilizer applied, soil texture, and climatic variables (temperature and precipitation). Because our focus was on bioenergy production, we only used switchgrass yield data from field trials used for biofuel production rather than studies where switchgrass was grown for forage, as practices between switchgrass grown for forage and bioenergy production can be very different (e.g., harvest frequency). Here we limited our analysis to cultivars that had fifteen or more observations. Soil texture variables (% sand, % silt, and % clay) were derived by overlaying the latitude and longitude of field trials on to soil texture data from the Soil Survey Geographic Database (SSURGO). Climatic variables for each field trial location were obtained using the PRISM data set (PRISM Climate Group, accessed March 2009).

2.2. Data analysis

In a first step, to gain insight into the drivers of switchgrass yield and to understand the relative importance of genetic, climatic, and cultural practice factors for explaining variability in switchgrass yield, 22 explanatory variables thought to affect switchgrass yield were examined (Table 2). Only variables identified in the literature as having an influence on switchgrass yield were included. These variables included climate, soil texture, stand age, and crop management variables hypothesized in the literature to influence switchgrass yields. To assess the importance of explanatory variables, we computed variable importance by applying the random Forest (RF) algorithm (Breiman, 2001) using the randomForest package (Liaw and Wiener, 2002) in R (R Development Core Team, 2008). RF is an ensemble classifier that builds a large number of regression trees, hence a "forest" of trees, and grows each tree with a randomized subset of predictors, hence "random" forests. Regression trees are nonparametric, hierarchical models that consist of a set of decision rules on the predictor variables that recursively partition the data based on binary splits. The trees in RF are grown to maximum size without pruning and are aggregated by averaging the trees (Breiman, 2001). The RF algorithm was designed to produce accurate predictions that do not overfit the data (Prasad et al., 2006). We built 1000 trees with six randomly sampled candidate variables evaluated at each split. Based on results of the RF analysis which showed that genetic variability is an important factor in explaining variability in switchgrass yields, we further split the data set and conducted the analyses separately for the two cytotypes.

To test for collinearity among explanatory variables, we did a pairplot of each group of variables (e.g., temperature, precipitation, soil variables) and computed correlation coefficients between each two variables. When all variables within a group were collinear with one another as indicated by a correlation coefficient greater than 0.5, only one variable was kept. Therefore models were built with the remaining 10 variables (Table 2), which included early and late growing season precipitation, average annual temperature, cultivar, germplasm origin, month of harvest, percent sand, percent clay, stand age, and nitrogen fertilizer. During data exploration all six continuous variables (% silt, % clay, April–May precipitation, June–September precipitation, average growing season temperature, nitrogen fertilizer, and stand age) showed non-linear relationships with switchgrass yield. Therefore we used general additive modeling (GAM) to model switchgrass yields as a smooth, nonlinear function of explanatory variables. GAMs use smoothing curves to model the relationships of explanatory variables with the dependent variable, allowing for non-linear relationships and thus revealing structure in the data that might otherwise be missed with linear models (Hastie and Tibshirani, 1990; Faraway, 2004). We applied a forward model selection using the Akaike information criteria (AIC). The general formula of an additive model with multiple explanatory variables is given below:

$$y_i = \alpha + \sum f_i(x_{ij}) + \varepsilon_i,\ \varepsilon_i \approx N(0, \delta^2)$$