Predicting Site Locations for Biomass-using Facilities with Bayesian Methods

TIMOTHY M. YOUNG
The University of Tennessee, Center for Renewable Carbon
2506 Jacob Drive
Knoxville, TN 37996-4570
USA
tmyoung1@utk.edu, www.spc4lean.com

JAMES H. PERDUE
USDA Forest Service, Southern Research Station
The University of Tennessee, Center for Renewable Carbon
2506 Jacob Drive
Knoxville, TN 37996-4570
USA
jperdue@fs.fed.us, https://www.srs.fs.usda.gov/index.php

XIA HUANG
The University of Tennessee, Center for Renewable Carbon
2506 Jacob Drive
Knoxville, TN 37996-4570
USA
xhuang8@vols.utk.edu, www.spc4lean.com

Abstract: - Logistic regression models combined with Bayesian inference were developed to predict locations and quantify factors that influence the siting of biomass-using facilities that use woody biomass in the Southeastern United States. Predictions were developed for two groups of mills, one representing larger capacity mills similar to pulp and paper mills (Group II), and another group of smaller capacity mills similar to the size of sawmills (Group I). "Median Family Income," "Road Density," "Slope," "Timberland Annual Growth-to-Removal Ratio," and "Forest Land-Area Ratio" were highly significant in influencing mill location for Group I. "Slope," "Urban Land Area Ratio," and "Number of Primary Wood Processing Mills" were highly significant in influencing mill location for Group II. In validation the sensitivity of the model for Group I was 86.8% and specificity was 79.3%. In validation the sensitivity for Group II was 80.9% and specificity was 84.1%. The higher probability locations (> 0.8) for Group I mills were clustered in the southern Alabama, southern Georgia, southeast Mississippi, southwest Virginia, western Louisiana, western Arkansas, and eastern Texas. The higher probability locations (> 0.8) for Group II mills were clustered in southeast Alabama, southern Georgia, eastern North Carolina, and along the Mississippi Delta.

Key-words: - Biomass-using facilities, woody biomass, site location, prediction, Bayesian logistic models

1 Introduction

The 20th century was marked by rapid growth and increased prosperity in the world. By 2040, the world's energy consumption is predicted to be 48% higher than it is today [1]. Since the 1970s, macroeconomists have viewed changes in the price of oil as an important source of economic fluctuations, as well as a paradigm for global economic shock, likely to affect many economies simultaneously [2]. The amalgamation of economic, environmental, social, and national security concerns for petroleum-based economies have created a renewed emphasis on alternative sources of energy which include biomass [3-9].

Biomass is a renewable resource procured from multiple sources which include agricultural residues, land clearings, landscaping, industrial byproducts, *etc.* [10]. However, developing a new bioeconomy will involve understanding and quantifying many economic relationships [11-15]. The objective of this study was to improve the assessment of site locations for biomass-using facilities that rely on woody biomass feedstocks. Decision support tools utilizing GIS-based multi-criteria, land-use, and socio-economic analyses can generate visual evidence of biomass supply/demand, land suitability, energy crop production potential, and ecological benefits. Composite indices visually summarizing information contained in an array of individual attributes would help the public, industry, media, and policy makers see an overall picture that is not so obvious from the component attributes themselves.

The present study enhances the study by Young *et al*. [16]. Prior research exists on identifying economically viable sites for biomass-using facilities [17-23]. However, the authors are not aware of previous published research that combines logistic regression with Bayesian inference to assess site locations for biomass-using facilities. The current study relies on existing mills that use woody biomass (*e.g.,* sawmills, OSB mills, and pulp and paper mills). These mills were used as surrogates for potential woody biomass-using bioenergy and biofuels plants for validation. The assumption is that such mill share similarities in feedstock requirements, supply chain, and processing/handling technologies. Support for this assumption comes from Patari [24], *i.e.,* complementary resources held by forest and energy companies make collaboration in the bioenergy business favorable. Leveraging the synergistic business relationships that exist in the feedstock supply chain between the two industries will be essential for reducing risk and minimizing capital investment in the emerging bioenergy and biofuels industries [25-27]. The study by Thorpe *et al.* [28] also indicates that the pulp and paper sector will rely on cellulosic byproducts as an important business strategy for product mix and energy self-sufficiency.

Predictive models were developed to quantify significant factors that influence the siting of such biomass-using facilities. Economic factors, transportation related influences, and biomass availability were studied as predictor variables. The study region consisted of 13 states in the Southeastern United States.[1]

2 Materials and Methods

2.1 Date Set

This study involved organizing large volumes of data collected from various sources, including the U.S. Census Bureau [29, 30], U.S. Forest Service [31], U.S. National Land Cover Database [32], U.S. National Elevation Dataset [33], U.S. Department of Agriculture National Agricultural Statistic Service [34], U.S. Environmental Protection Agency [35], and state-level mill location directories of 2010. Cost data from the BioSAT model were also used [36]. All records in this study were organized at the U.S. Census Bureau [30] 5-digit ZIP Code Tabulation Area (ZCTA) level. There were 10,016 ZCTAs in the study region which corresponded to 10,016 potential analytical polygons or potential sites for woody biomass-using facilities. The average area size for 5-digit ZCTAs in the 13-state study regions was 209.84 km^2.

2.2 Biomass Estimation using GIS

Forest biomass annual growth and removal quantity data were collected at the county level from the Forest Inventory and Analysis Database (FIADB) version 3.0 and reallocation was done for each of the 10,016, 5-digit ZCTAs using geographic information system (GIS) technology (Figure 1a). National land cover data [34] and digital raster map data were used to identify forestland. In the digital raster map, each pixel represents one particular land cover class, *i.e.,* water, urban, forest, or cropland, *etc.* (Figure 1b). Forest biomass annual growth and removal quantities were proportionally allocated with GIS spatial overlay techniques to each 5-digit ZCTA using the county boundary of the 5-digit ZCTA, and the land cover image data. Due to some misalignments of county boundaries with 5-digit ZCTA boundaries each forest biomass county was split into multiple area parts via the 5-digit ZCTA area shape file, and assigned a unique 5-digit ZCTA identifier. By overlaying each area part with the land cover image layer, the numbers of pixels in all land cover classes within each area of the 5-digit ZCTA were estimated (Figure 1c). By summing up the pixels of deciduous, coniferous,

[1]Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia.

and mixed deciduous- coniferous forests, which together represented total forestland, a forestland pixel ratio for each 5-digit ZCTA within a county boundary was calculated. The forest biomass quantity in every 5-digit ZCTA was derived using this pixel ratio technique (Figure 1d). A summed quantity value was then calculated as the forest biomass quantity in a 5-digit ZCTA. Allocation of biomass supply using GIS land cover data and proportionality allocating it at the 5-digit ZCTA level while incorporating socio-economic factors such as urban areas, road network density, etc., as well as other geographic phenomenon such as park and preserve boundaries, waterways, etc., maintains the integrity of the U.S. Census data with the overlays and expands upon prior research [36, 3, 9].

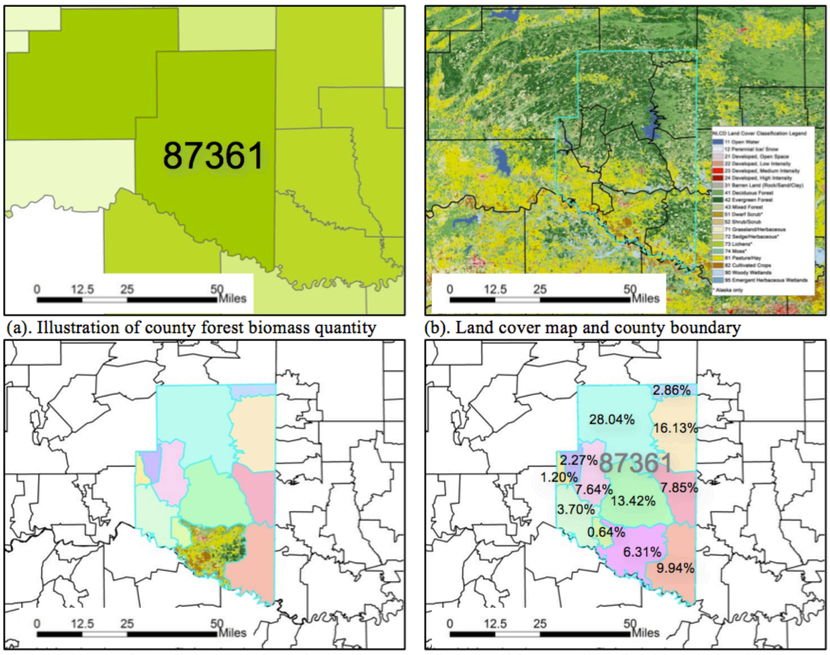

Figure 1. Illustration of forest biomass allocation at the level of 5-digit ZCTA.

2.3 Logistic Regression

In Young *et al.* [16] logistic regression was used to identify site locations for biomass-using facilities and significant factors associated with these site locations. This study applied Bayesian inference for estimation of the parameters in the logistic regression models. Bayesian inference specifies the probability distribution for the underlying categorical or continuous variables and estimates parameters β. Bayesian inference allows for incorporation of prior beliefs and the combination of such beliefs with statistical data which are well suited for representing the uncertainties in the value of independent variables [37, 38]. The data set used to develop the Bayesian logistic regression (10,016 observations associated with the 5-digit ZCTAs) were partitioned into two parts using a stratified random sampling technique for each state which ensured a spatially proportionate data allocation across the study region: 80% for training and 20% for validation. The training data were used to develop the models while the validation data were used to evaluate the model performance.

2.4 Group I and Group II Subsets

Bioenergy and biofuel plants are defined as facilities that integrate woody biomass conversion processes, and equipment to produce wood pellets for energy, biofuels, biopower, or value-added biochemicals [39]. Only 60 such facilities are known to exist in the study region. Given the large amount of ZCTAs that did not contain bioenergy or biofuels mills (which is problematic for logistic regression) more traditional wood-using facilities in the study area were used as surrogates (*e.g.*, sawmills, OSB mills, and pulp and paper mills). The assumption is that similar factors will influence site preference and suitability given the commonality in feedstocks and procurement systems [40-43, 16]. The mills were partitioned into groups based on capacity:

- ***Group I:*** Sawmills (Figure 2);
- ***Group II:*** Pulp and paper mills, OSB mills and wood pellets mills (Figure 3).

2.5 Response and Explanatory Variables

Two separate response variables were considered for modeling and ranking potential sites. For Group I, the response variable, y_{i1} equals *1* if *ith* ZCTA had at least one woody biomass-using facility, and y_{i2} equals *1* was defined similarly for Group II mills. In logistic regression the dependent variable is either a *1* or *0* depending on the desired outcome of y, in this case *1* is a desired outcome of mill location. Thirteen explanatory variables for data available in the public domain were examined in the Bayesian logistic models (Table 1). The 13-explanatory variables were selected given the ability to use data sources in the public domain that are periodically updated and that such data could be organized at the resolution of the 5-digit ZCTA. These variables were selected given the findings of previous research, *e.g.*, population density [42], median income [44], etc.

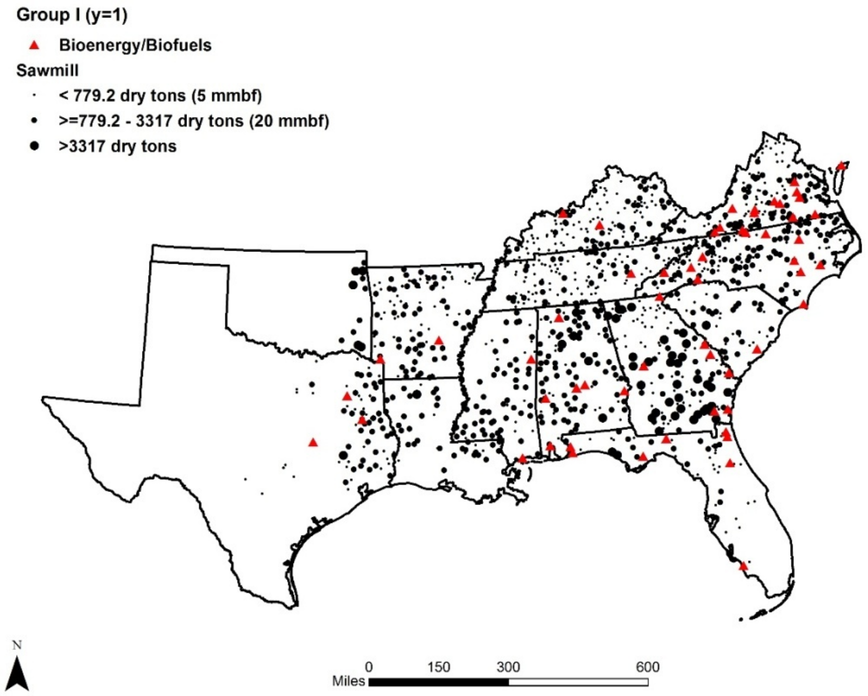

Figure 2. Illustration of Group I woody biomass-using mills.

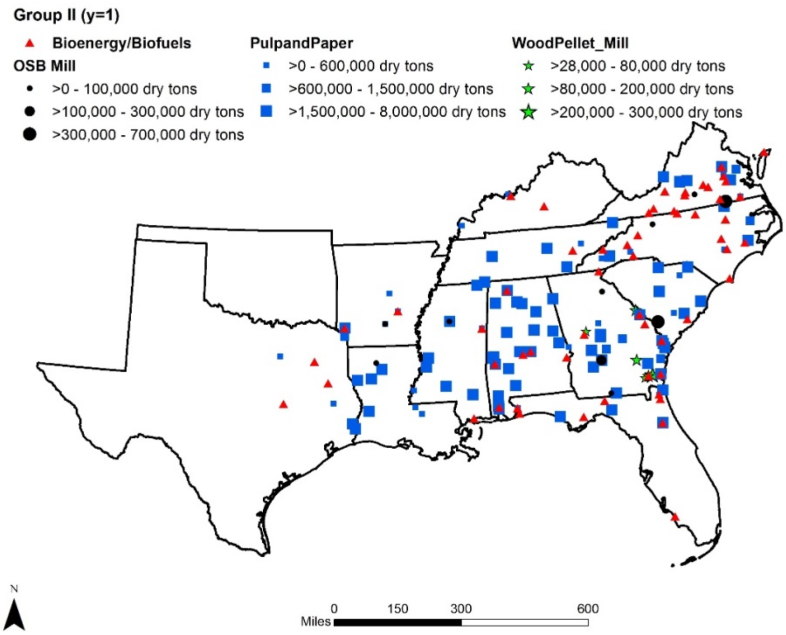

Figure 3. Illustration of Group II woody biomass-using mills.

Table 1. Explanatory Variables Organized by ZCTA

Variable	Original Data Resolution	Unit	Data Sources
Population Density	5-digit ZCTA	People/mile^2	U.S. Census Bureau (2010) population density in each 5-digit ZCTA.
Household Density	5-digit ZCTA	Household/mile^2	U.S. Census Bureau (2010) household density in each 5-digit ZCTA.
Household Unit Density	5-digit ZCTA	Household unit/mile^2	U.S. Census Bureau (2010) household unit density in each 5-digit ZCTA.
Median Family Income	County	Dollar	U.S. Census Bureau (2010) median family income in each county
Farm Net Income	County	Dollar	USDA NASS Census Agriculture (2007) farm net income in each county.
Road Density	5-digit ZCTA	km/km^2	U.S. Census Bureau (2010) road length
Crop Cultivated Land Area Ratio			
Forest Land Area Ratio			
Urban Land Area Ratio			
Water Area Ratio	5-digit ZCTA	percent	U.S. National Land Cover Database (2006)
Slope	5-digit ZCTA	percent	U.S. National Elevation Dataset (2010) NED 1arc second
Timberland Annual Growth-to-Removal Ratio	County	-	Forest Inventory and Analysis – The Timber Products Tools (TPO) (2009)
Number of Primary Wood Processing Mills in Each ZCTA	5-digit ZCTA	-	U.S. Forest Service (2009) and state mill directories

2.6 Modeling Scoring and Interpretation

Given a specific response variable and set of predictor variables, the fitted Bayesian logistic regression model provided an estimated probability that a ZCTA will contain a woody biomass-using facility, *i.e.*, it is the probability (not the *odds ratio*) of exactly one biomass using facility. The probability was used in the validation data set to compare ZCTAs with actual mill locations.

3 Results

3.1 Bayesian Logistic Regression Estimates

3.1.1 Predicted Mill Locations for Group I

Eight out of the possible 12 predictor variables were statistically significant (p-value < 0.05), see Table 3. To compare the maximum likelihood estimator (MLE) and Bayesian inference estimation methods for parameter coefficients, Classification Tables were used for the training and validation data sets. The Classification Tables confirmed that the Bayesian logistic regression inference assuming a uniform prior had good predictive power for the siting locations for Group I mills (Tables 4 and 5). In the validation data set, the sensitivity of this model was 86.78% (*i.e.*, the model predicted a current mill location correctly 86.78%), and specificity (*i.e.*, predicting no mill location) was 79.26%. The sensitivity rates in training and validation data sets exceeded those of the study by Young *et al.* [16] and were considered to be acceptable when compared to the stringent 75% criteria of medical radiology screening [45].

Table 3. Significant Variables for Group I mills

Significant variables	Mean Estimates	p-value
Intercept	-4.54	
Median Family Income	-0.000037	<0.0001
Road Density	-0.33	<0.0001
Slope	-0.075	<0.0001
Timberland Annual Growth-to-Removal Ratio	2.09	<0.0001
Forest Land Area Ratio	2.66	<0.0001
Urban Land Area Ratio	-0.33	<0.0012
Water Area Ratio	5.90	<0.0020
Household Density	-0.0002	<0.0073

Table 4. Summary of Classification Table for Training Dataset for Group I mills

Parameter Estimation Method		Training Data Set (y = Prediction Value |Actual Value)					
		y=0|0	*y=1|0*	*y=0|1*	*y=1|1*	**Specificity** $P(\hat{y}=0 \mid y=0)$	**Sensitivity** $P(\hat{y}=1 \mid y=1)$
Maximum Likelihood Estimation (MLE)		2670	658	132	778	79.99%	85.49%
Bayesian Inference	*Uniform*	2673	655	124	786	80.08%	86.37%
	Gaussian	2670	658	133	777	79.99%	85.38%

Table 5. Summary of Classification Table for Validation Dataset for Group I mills

Parameter Estimation Method		Validation Data Set (y = Prediction Value |Actual Value)					
		y=0|0	**y=1|0**	**y=0|1**	**y=1|1**	**Specificity**	**Sensitivity**
Maximum Likelihood Estimation (MLE)		640	194	37	190	76.74%	83.70%
Bayesian Inference	*Uniform*	661	173	30	197	79.26%	86.78%
	Gaussian	640	194	37	190	76.74%	83.70%

Median Family Income, Road Density, Slope, Timberland Annual Growth-to-Removal Ratio, and Forest Land Area Ratio were highly significant in influencing mill location (recall Table 3). Other statistically significant variables were *Urban Land Area Ratio, Water Area Ratio*, and *Household Density*. A higher family income, higher household density, higher road density, and land area ratio classified as *Urban* had negative coefficients. *Slope* had a negative coefficient, and *Timberland Annual Growth-to-Removal Ratio, Forest Land Area Ratio*, and *Water Area Ratio* had positive coefficients. This indicates that landscape with a lower slope but abundant forestland, high water area ratios, and high forest land ratio are preferred.

The study results related to median family income supports the study by White and Mazza [43]. As White and Mazza [43] note in citing other studies *key determinants in land conversion are increasing human populations, rising personal incomes, and changing societal preferences* [46, 47]. The signs of the coefficients in the Bayesian logistic regression models support the findings of White and Mazza [43] and other authors as cited in their study, *i.e.*, higher family incomes, household density and road densities have a negative influence on forestation which would deter any future wood processing mill location. See Kimsey *et al.* [48] for the limitations of slope on timber harvesting. The study also supports the findings of Luppold and Baumgras [49] and the interaction between forest industry and the proximity of the forest resource.

This result suggests the importance of *landscape suitability* on mill location [44]. Four categories of probability were developed from the Bayesian logistic model (Figure 4). The higher probability locations (> 0.8) for Group I mills were clustered in the southern Alabama, southern Georgia, southern Mississippi, southern Virginia, western Louisiana, southwest Arkansas, and eastern Texas.

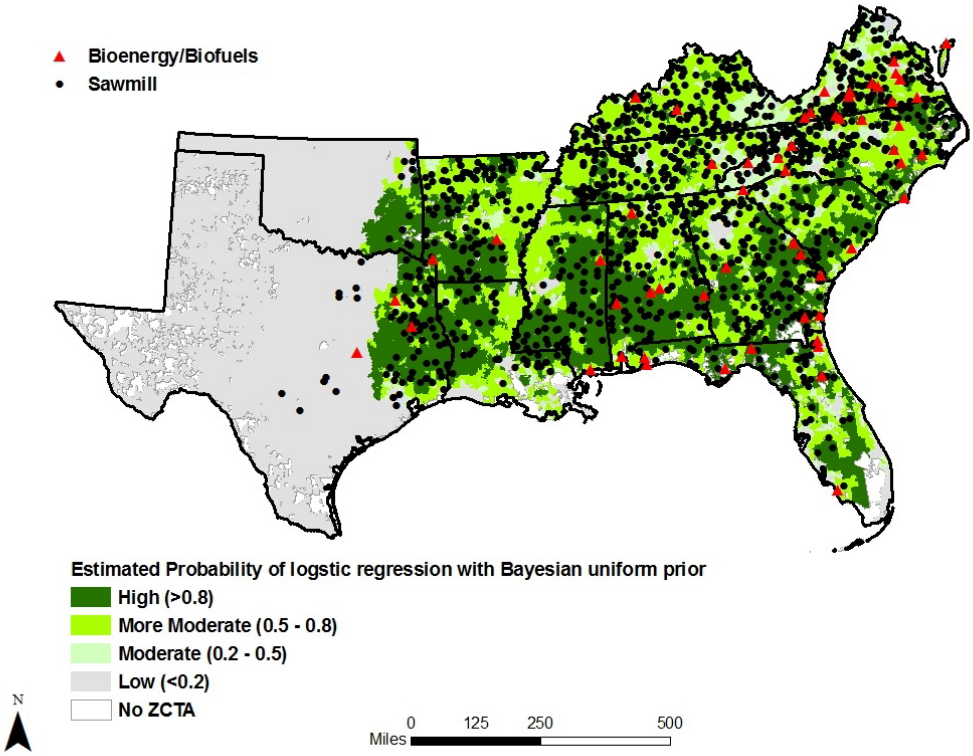

Figure 4. Estimated probability locations for Group I.

3.1.2 Predicted Mill Locations for Group II

Four predictor variables were statistically significant (Table 6). The classification tables confirmed that the logistic regression with Bayesian inference assuming a uniform prior had good predictive power for the siting locations at the 5-digit ZCTA resolution for Group II (Tables 7 and 8). Bayesian Inference with a uniform prior had a sensitivity of 80.95% and specificity of 84.10% for Group II training data (Table 7). In the validation data set, the sensitivity of this model was 90.48% and specificity was 80.14% (Table 8). Both sensitivity and specificity were greater than those of Young *et al.* [16].

Table 6. Significant Variables for Group II mills

Significant variables	Mean Estimates	p-value
Intercept	-0.511	<0.0001
Slope	-0.511	<0.0001
Forest Land Area Ratio	-0.077	0.0075
Urban Land Area Ratio	4.848	<0.0001
Number of Primary Wood Processing Mills in Each ZCTA	5.251	<0.0001

Table 7. Summary of classification table for Training Dataset for Group II mills

Parameter Estimation Method		Training Data Set (y = Prediction Value |Actual Value)					
		y=0|0	*y=1|0*	*y=0|1*	*y=1|1*	Specificity $P(\hat{y}=0 \mid y=0)$	Sensitivity $P(\hat{y}=1 \mid y=1)$
Maximum Likelihood Estimation (MLE)		489	96	17	67	83.59%	79.76%
Bayesian Inference	*Uniform*	492	93	16	68	84.10%	80.95%
	Gaussian	489	96	17	67	83.59%	79.76%

Table 8. Summary of classification table for Validation Dataset for Group II mills

Parameter Estimation Method		Validation Data Set (y = Prediction Value |Actual Value)					
		y=0|0	*y=1|0*	*y=0|1*	*y=1|1*	**Specificity**	**Sensitivity**
Maximum Likelihood Estimation (MLE)		117	29	4	17	80.14%	80.95%
Bayesian Inference	*Uniform*	117	29	2	19	80.14%	90.48%
	Gaussian	117	29	4	17	80.14%	90.48%

Slope, *Urban Land Area Ratio*, and *Number of Primary Wood Processing Mills in Each ZCTA* were highly significant in influencing Group II mill locations. Another less statistically significant variable was *Forest Land Area Ratio*. *Slope* had negative influences on mill locations for Group II. F*orest Land Area Ratio*, *Urban Land Area Ratio*, and *Number of Primary Wood Processing Mills in Each ZCTA* had positive influences on Group II mill locations. This may reinforce the preposition of the synergistic relationship that exists between Group I and Group II facilities which are dependent on feedstocks from Group I mills. Many pulp and paper mills rely on wood chips from sawmills and the statistical significance of Group I mills (primarily sawmills) as an independent variable for the larger Group II mills highlight another strength of the models. Four categories of probabilities were developed from the Bayesian logistic model (Figure 5). The higher probability locations for Group II mills were clustered in southeast Alabama, southern Georgia, eastern North Carolina, and the Mississippi Delta.

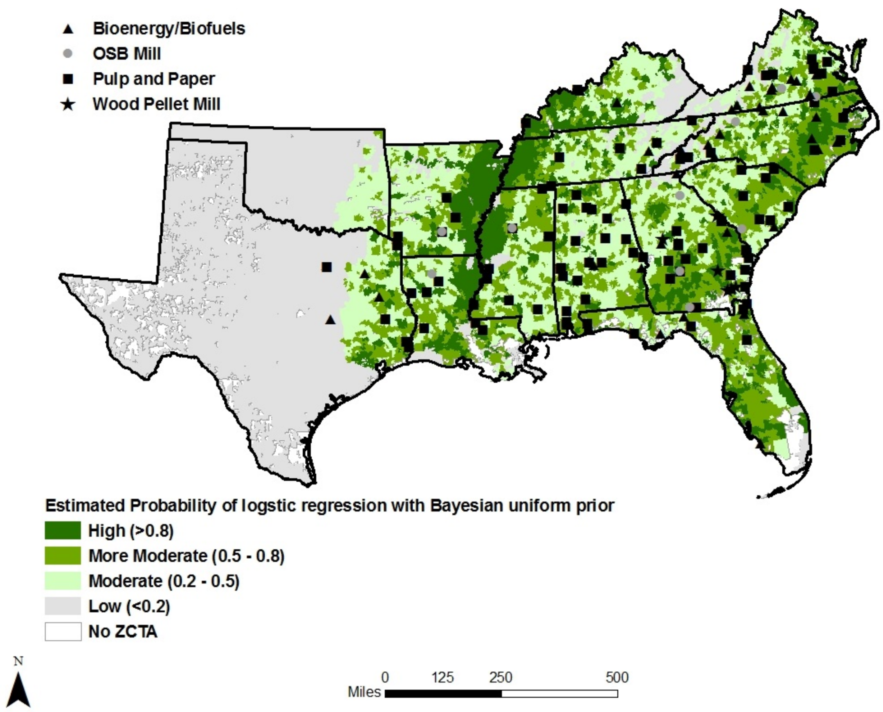

Figure 5. Estimated probability locations for Group II.

3.3 Validation with newly start-up plants in the year of 2011-2013

There were 21 bioenergy/biofuels, four pulp and paper mills, and 12 wood pellets mills start-ups between 2011-2013 (Table 9). Considering the medium-to-large operating volume of these mills, the siting model for Group II was used for this additional validation study. Fourteen of the 21 bioenergy/biofuels had estimated probabilities over 0.5. Three of the four pulp and paper mills fell in the *probable locations* where the estimated probabilities were greater than 0.5, and all wood pellets mills had the estimated probabilities greater than 0.5 (Table 10). The overall correct rate in the use of newly startup mills for Group II siting model was 78.4%. Although there were eight mills not directly falling into the *probable locations* of the Group II model, they were very close to the nearby preferred probable locations (Figure 6).

Table 9. Newly startup plants in the year of 2011-2013

Plant Type	Plant Name	County	State	Zip Code	Startup	Capacity
Biofuels	Savannah River Site	Aiken	SC	29802	2011	200,000 tons
Biofuels	Aspen Power	Angelina	TX	75901	2011	500,000 tons
Biofuels	Rentech/Biomass G&E	Gulf	FL	32456	2011	550,000 tons
Biofuels	Dominion Virginia	Wise	VA	24293	2012	537,000 tons
Biofuels	WGS Energy	Twiggs	GA	31044	2012	350,000 tons
Biofuels	Southern Co	Nacogdoches	TX	75788	2012	1,000,000 tons
Biofuels	Green Power Solutions	Laurens	GA	31040	2012	560,000 tons
Biofuels	Grow Green Power	Anson	NC	28170	2012	370,000 tons
Biofuels	Weyerhaeuser/KGRA	Pitt	NC	28513	2012	45,000 tons
Biofuels	ClearFuels	Wayne	TN	38450	2013	70,000 tons
Biofuels	North Star Renewable	Jackson	GA	30549	2013	240,000 tons
Biofuels	Dominion Power	Franklin	VA	23851	2013	500,000 tons
Biofuels	Dominion Power	Hopewell	VA	23860	2013	500,000 tons
Biofuels	South Boston Energy	South Boston	VA	24592	2013	500,000 tons
Biofuels	MeadWestvaco	Covington	VA	24426	2013	750,000 tons
Biopower	Piedmont Green Power	Lamar	GA	30204	2012	50 Megawatt
Biopower	Southern Power	Nacogdoches	TX	75961	2012	100 Megawatt
Biopower	Gainesville Renewable Energy Center	Alachua	FL	32653	2013	100 Megawatt
Biorefinery	KiOR	Lowndes	MS	39701	2012	182,500 tons
Biorefinery	ClearFuels	Wayne	TN	38450	2013	200,000 tons
Biorefinery	HCL Clean Tech	Grenada	MS	38901	2013	1,000,000 tons
Pulp and paper	Packaging Corporation of America	Lowndes	GA	31601	2012	457,000 tons
Pulp and paper	International Paper Franklin VA Mill	Franklin	VA	23851	2012	300,000 tons
Pulp and paper	International Paper Pine Hill Mill	Wilcox	AL	36769	2012	450,000 tons
Pulp and paper	MeadWestvaco Evadale Mill	Hardin	TX	77656	2012	696,000 tons
Wood pellet	Georgia Biomass	Ware	GA	31503	2011	1,500,000 tons
Wood pellet	Enviva	Hertford	NC	27910	2011	770,000 tons
Wood pellet	Biomass Energy/Enviva	Louisa	VA	23024	2011	260,000 tons
Wood pellet	Fram Renewable Fuels	Appling	GA	31513	2012	400,000 tons
Wood pellet	Westervelt Renewable En.	Pickens	AL	35442	2012	620,000 tons
Wood pellet	Varn Wood Products	Brantley	GA	31542	2012	160,000 tons
Wood pellet	German Pellets	Tyler	TX	75979	2012	1,000,000 tons
Wood pellet	Equustock	Chesterfield	VA	23836	2012	88,000 tons
Wood pellet	Enviva	Northampton	NC	27832	2012	880,000 tons
Wood pellet	Enviva Southampton	Southampton	VA	23837	2013	550,000 tons
Wood pellet	Point Bio Energy	Baton Rouge	LA	70767	2013	880,000 tons
Wood pellet	Enviva	Southampton	VA	23837	2013	1,000,000 tons

Table 10. Validation Group II model with newly startup plants in the year of 2011-2013

Plant Type	Total Number	Estimated Probability for Group II (logistic with the competition index)				Correct Rate for "Probable Location" (probability>0.5)
		>0.8	0.5-0.8	0.2-0.5	<0.2	
Bioenergy/Biofuels	21	1	13	6	1	66.67%
Pulp and paper	4	0	3	1	0	75%
Wood pellet	12	1	11	0	0	100%

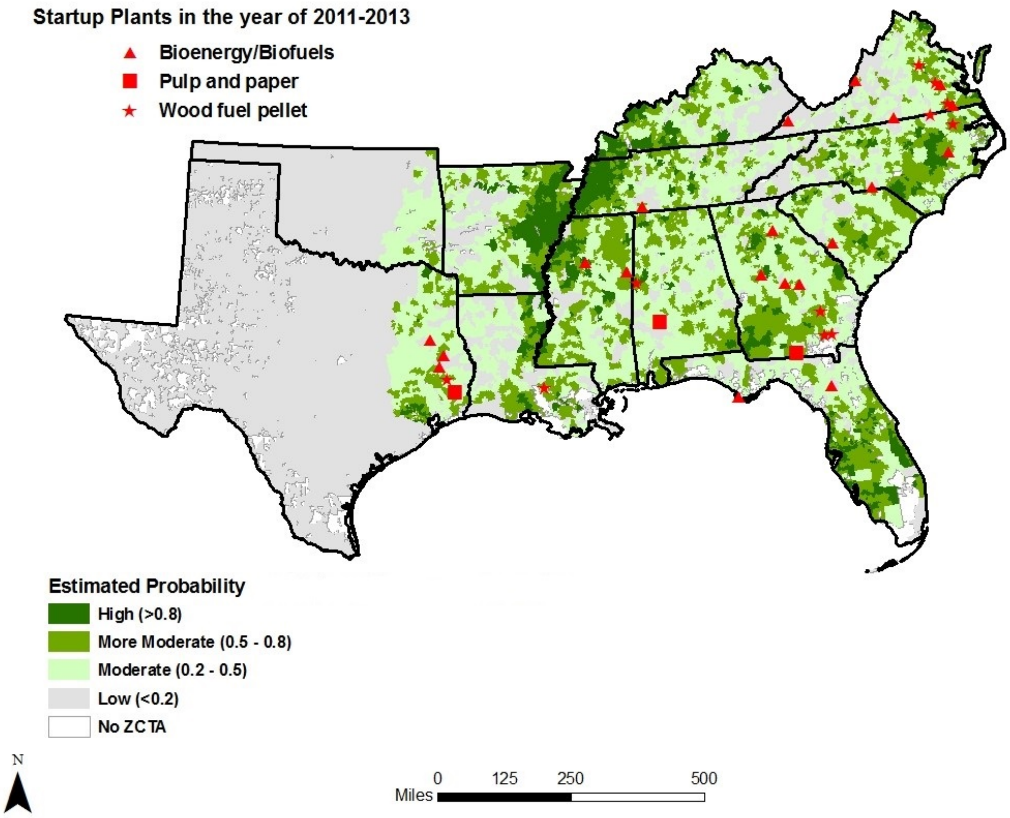

Figure 6. Newly startup mills overlay of estimated probability locations for Group II facilities.

5 Conclusions

Logistic regression models combined with Bayesian inference were developed to quantify factors that influence the siting of biomass-using facilities that use woody biomass and predict potential biorefinery locations in the Southeastern United States. *Median Family Income, Road Density, Slope, Timberland Annual Growth-to-Removal Ratio,* and *Forest Land Area Ratio* were statistically significant in influencing mill location for smaller capacity mills similar to sawmills (Group I). *Slope, Urban Land Area Ratio,* and *Number of Primary Wood Processing Mills* were statistically significant in influencing mill location for large capacity mills like pulp and paper mills (Group II).

References:

[1] Energy Information Administration. 2016. International energy outlook. DOE/EIA 0484(2016). https://www.eia.gov/outlooks/ieo/pdf/0484(2016).pdf

[2] Blanchard, O.J. and J. Gali. 2007. The macroeconomic effects of oil price shocks: why are the 2000s so different from the 1970s? *In* Proc. of NBER ME Conference on International Dimensions of Monetary Policy, S'Agaro, Catalonia, Spain, 2007.

[3] Perlack, R., L. Wright, A. Turhollow, R. Graham, B.J. Stokes and D. Erbach. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility

of a billion-ton annual supply. Oak Ridge, TN: Oak Ridge National Laboratory.
[4] Hubbe, M.A., and U.A. Buelhmann. 2010. A continuing reverence for wood. *Bioresources* 5(1):1-2.
[5] Kumarappan, S., S. Joshi and H.L. MacLean. 2009. Biomass supply for biofuel production: estimates for the United States and Canada. *Bioresources.* 4(3):1070-1087.
[6] Cheng, S.M. and S.D. Zhu. 2009. Lignocellulosic feedstock biorefinery – the future of the chemical and energy industry. *Bioresources*. 4(2):456-457.
[7] Lucia, L.A. 2008. Lignocellulosic biomass: a potential feedstock to replace petroleum. *Bioresources.* 3(4):981-982.
[8] Pawlak, J.J. 2008. A sustainable economy. *Bioresources.* 3(1):1-2.
[9] U.S. Department of Energy. 2011. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts industry. Oak Ridge Oak Ridge National Laboratory. Oak Ridge, TN.
[10] Caputo, J. 2009. Sustainable forest biomass: promoting renewable energy and forest stewardship. Environmental and Energy Study Institute Policy Paper, Washington, D.C. http://www.eesi.org/
[11] Altman, I. and T. Johnson. 2008. The choice of organizational form as a non-technical barrier to agro-bioenergy industry development. *Biomass Bioenerg.* 32(1):28-34.
[12] Dasmohapatra, S. 2009. Future market drivers for the forest products industry. *Bioresources.* 4(4):1263-1266.
[13] Gronowska, M., S. Joshi and H.L. MacLean. 2009. A review of U.S. and Canadian biomass supply studies. *Bioresources.* 4(1):341-369.
[14] U.S. Department of Energy. 2016. 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy, Volume 1: Economic Availability of Feedstocks. M. H. Langholtz, B. J. Stokes, and L. M. Eaton (Leads), ORNL/TM-2016/160. Oak Ridge National Laboratory, Oak Ridge, TN. 448p.
[15] Zalesny, R.S., J.A. Stanturf, J.A., E.S. Gardiner, J.H. Perdue, T.M. Young, D.R. Coyle, W.L. Headlee, G.S. Banuelos, A. Hass. 2016. Ecosystem services of woody crop production systems. *BioEnergy Res*. 9(2):465-491.
[16] Young, T.M., R.L. Zaretski, J.H. Perdue, F.M. Guess and X. Liu. 2011. Logistic regression models of factors influencing the location of bioenergy and biofuels plants. *Bioresources.* 6(1):317-328.
[17] Sperling, D. 1984. An analytical framework for siting and sizing biomass fuel plants. *Energy*.9(11-12):1033-1040.
[18] Young, T.M., D.M. Ostermeier, J.D. Thomas and R.T. Brooks. 1991. The economic availability of woody biomass for the Southeastern United States. *Bioresource Technol.* 37(1):7-15.
[19] Lynd, L.R. 1996. Siting an ethanol plant in the Northeast Burlington VA. C.T., Donovan Associates.
[20] Fried, J., G. Christensen, D. Weyermann and G. Pinjuv. 2003. Modeling opportunities and feasibility of siting wood-fired electrical generating facilities to facilitate landscape-scale fuel treatment with FIA BioSum. *In* Proc. Systems Analysis in Forest Resources, 207-216.
[21] Freppaz, D., R. Minciardi, M. Robba, M. and A. Taramasso. 2004. Optimizing forest biomass exploitation for energy supply at a regional level. *Biomass Bioenerg.* 26:15-25.
[22] Moon, E., B. Saveyn, S. Proost and M. Hermy. 2008. Optimal location of new forests in a suburban region. *J For Econ.* 14(1):5-27.
[23] Panichelli, L. and E. Gnansounou. 2008. GIS-based approach for defining bioenergy facilities location: A case study in Northern Spain based on marginal delivery costs and resources competition between facilities. *Biomass Bioenerg.* 32:289-300.
[24] Patari, S. 2010. Industry-and company-level factors influencing the development of the forest energy business-insights from a Delphi study. *Technological Forecasting & Social Change*. (77):94-109.
[25] Knight, D.K. 2009. Wood bioenergy. Hatton-Brown Publishers, Inc., Montgomery, AL.
[26] Stewart, P. 2009. Energy and the wood fiber supply chain. Forests & People Magazine. (Q1). http://www.forest2market.com/uploads/Legacy/energy-and-the-wood-fiber-supply-chain_1.pdf
[27] Conrad, J.L. IV, M.C. Bolding, W.M. Aust and R.L. Smith. 2010. Wood-to-energy expansion, forest ownership changes, and mill closure: Consequences for U.S. South's wood

supply chain. *Forest Policy Econ.* 12(6):399-406.

[28] Thorp B., H. Seamans and M. Akhtar. 2013. Cellulosic bioproducts will be important to pulp and paper. *Paper 360° TAPPI*. 8(6):18-19.

[29] U.S. Census Bureau. 2010a. 2010 Census ZIP code tabulation areas [Data file]. http://www.census.gov/geo/ZCTA/zcta.html

[30] U.S. Census Bureau. 2010b. 2010 Urban and rural classification [Data file]. http://www.census.gov/geo/reference/urban-rural.html

[31] U.S. Forest Service. 2009. Lands in public preserves [Data file]. Personal contact on 08/2011.

[32] U.S. National Land Cover Database. 2006. Multi-resolution land characteristics consortium [Data file]. http://www.mrlc.gov/nlcd2006.php.

[33] U.S. National Elevation Dataset. 2010. National elevation dataset 1 arc second [Data file]. http://seamless.usgs.gov/ned1.php.

[34] U.S. Department of Agriculture National Agricultural Statistics Service. 2008. Census of agriculture: farm net income [Data file]. http://www.nass.usda.gov

[35] U.S. Environmental Protection Agency. 2011. Ecoregions of the United States [Data file]. http://www.epa.gov/bioiweb1/html/usecoregions.html

[36] Perdue, J.H., T.M. Young and T.G. Rials. 2011. The Biomass Site Assessment Model – BioSAT. Final Report for U.S. Forest Service, Southern Research Station submitted by The University of Tennessee, Knoxville. 282p. www.biosat.net

[37] Kutner, M.H., C.J. Nachtsheim and J. Neter. 2004. Applied linear regression models. McGraw-Hill Irwin. New York, NY.

[38] Gustafsson, M.G., M.B. Wallman, H. Göransson, M. Fryknäs, C.R. Andersson and A. Isaksson. 2010. Improving Bayesian credibility intervals for classifier error rates using maximum entropy empirical priors. *Artif Intell Med.* 49(2):93-104.

[39] National Renewable Energy Laboratory. 2009. What is a biorefinery? http://www.nrel.gov/biomass/biorefinery.html

[40] Gary K. and L. Zeng. 2001. Logistic regression in rare events data. *Political Analysis*. 9:137-163.

[41] Cohen, J., V. Janssen and P. Stuart. 2010. Critical analysis of emerging forest biorefinery (FBR) technologies for ethanol production. *Pulp & Paper Canada.* (1):24-30.

[42] Wear, D., R. Liu, R., J. Foreman and R. Sheffield. 1999. The effects of population growth on timber management and inventories in Virginia. *Forest Ecology and Management*. 118:107-115.

[43] White, E.M. and R. Mazza. 2008. A closer look at forests on the edge: future development on private forests in three states. Gen. Tech. Rep. PNW-GTR-758. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

[45] Huang, X., J.H. Perdue, and T.M. Young. 2012. A spatial index for identifying opportunity zones for woody cellulosic conversion facilities. *Int J of For Res.* 2012:1-11.

[46] Carney, P.A. E.A. Sickles, B.S. Monsees, L.W. Bassett, R.J. Brenner, S.A. Feig, R.A. Smith, R.D. Rosenberg, T.A. Bogart, S.B. Browning, J.W. Kelly, K.A. Tran, K. and D.L. Miglioretti. 2010. Identifying minimally acceptable interpretive performance criteria for screening mammography. *Radiology.* 255(2):354-361.

[47] Alig, R.J., J.D. Kline, J.D. and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. *Landscape Urban Plan*. 69(2/3):219–234.

[48] Cho, S.H., J. Wuand and R. Alig. 2005. Land development under regulation: comparison between the east and west sides of the Cascade Range in Oregon, Washington, and California. *Rev Urban Regional Stud.* 17(1):1–17.

[49] Kimsey, M., D. Page-Dumroese and M. Coleman. 2011. Assessing bioenergy harvest risks: geospatial explicit tools for maintaining soil productivity in Western US forests. *Forests.* 2:797-813.

[50] Luppord, W. and J. Baumgras. 1999. The interaction between forest industry and the forest resource in West Virginia. *In Proc.* Improving forest productivity for time – a key to sustainability, 159-164.