LOGISTIC REGRESSION MODELS OF FACTORS INFLUENCING THE LOCATION OF BIOENERGY AND BIOFUELS PLANTS

Timothy M. Young,[a,]* Russell L. Zaretzki,[b] James H. Perdue,[c] Frank M. Guess,[d] and Xu Liu [e]

Logistic regression models were developed to identify significant factors that influence the location of existing wood-using bioenergy/biofuels plants and traditional wood-using facilities. Logistic models provided quantitative insight for variables influencing the location of woody biomass-using facilities. Availability of "thinnings to a basal area of $31.7m^2$/ha," "availability of unused mill residues," and "high density of railroad availability" had positive significant influences on the location of all wood-using faciities. "Median family income," "population," "low density of railroad availability," and "harvesting costs for logging residues" had negative significant influences on the location of all wood-using faciities. For larger woody biomass-using mills (e.g., biopower) availability of "thinnings to a basal area of $79.2m^2$/ha," "number of primary and secondary wood-using mills within an 128.8km haul distance," and "amount of total mill residues," had positive significant influences on the location of larger wood-using faciities. "Population" and "harvesting costs for logging residues" have negative significant influences on the location of larger wood-using faciities. Based on the logistic models, 25 locations were predicted for bioenergy or biofuels plants for a 13-state study region in the Southern United States.

*Contact Information: a: Professor, University of Tennessee, Center for Renewable Carbon, 2506 Jacob Drive, Knoxville, TN 37996-4570 USA; b: Associate Professor, University of Tennessee, Department of Statistics, Operations, Mgmt. Science, 335 Stokely Management Center, Knoxville, TN 37996-0532 USA; c: Senior Biological Scientist, USDA Forest Service, Southern Research Station, 2506 Jacob Drive, Knoxville, TN 37996-4570 USA; d: Professor, University of Tennessee, Department of Statistics, Operations, Mgmt. Science, 335 Stokely Mgmt Center, Knoxville, TN 37996-0532 USA; e: former Graduate Research Assistant, University of Tennessee, Department of Statistics, Operations, Mgmt. Science, 335 Stokely Mgmt Center, Knoxville, TN 37996-0532 USA. *corresponding author: tmyoung1@utk.edu*

INTRODUCTION

The 20th century was marked by rapid growth and increased prosperity in the world. By 2020, the world's energy consumption is predicted to be 40 percent higher than it is today (Energy Information Agency 2008). Key sources of oil for western markets are located in complex geopolitical environments that increase economic and social risk. About 59 percent of all oil consumed in the United States is imported, with approximately 20 percent coming from the Persian Gulf (Caputo 2009). 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 (Blanchard and Gali 2007). 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 woody biomass (Hubbe and Buehlman 2010; Kumarappan et al. 2009; Cheng and Zhu 2009; Lucia 2008; Pawlak 2008).

Woody biomass is a renewable resource procured from multiple sources which include land clearings, landscaping, industrial byproducts, and abundant forest resources (Caputo 2009). However, developing the new bioeconomy will involve understanding and establishing many relationships (Altman and Johnson 2008; Dasmohapatra 2009; Gronowska et al. 2009). Assessing the significant factors that influence the site location of existing wood-using bioenergy and biofuels plants in the Southern United States, as well as predicting new locations for such plants was the focus of this study. Given the limited number of existing wood-using bioenergy and biofuels plants in the Southern United States, traditional wood-using facilities were used as a group of surrogates (e.g., pulp and paper mills, wood composite mills, sawmills, etc.). Support for the premise that wood-using mills are a surrogate for woody biomass-using bioenergy and biofuels plants comes from the similarities in feedstock requirements and processing/handling technologies, e.g., wood pellet and co-generation plants. Pätäri (2009) noted that 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 (see Knight 2009; Stewart 2009; Cohen et al. 2010; Conrad et al. 2010).

This study builds upon previous research (Velázquez-Martí and Annevelink 2009; Velázquez-Martí and Fernandez-Gonzalez 2010), but is the first study to use logistic regression models to quantify significant factors influencing site location of woody biomass-using bioenergy and biofuels plants and predict potential locations based on probability. In prior studies, Sperling (1984) established a generalized, non-statistical, analytical framework to identify the critical factors for assessing the quality of biomass locations in specific regions. Young et al. (1991) used Geographic Information System (GIS) spatial analysis to assess the economic availability of woody biomass for potential sites for biorefineries in the Southern United States. Lynd (1996) used a sensitivity analysis to determine the relative importance of various site characteristics to the overall financial performance of a biomass ethanol plant in the Northeastern United States.

The logistic regression model is the most widely used method to relate a binary outcome (i.e. success/failure) to a set of explanatory variables in a regression setting. Logistic regression is frequently used in fields as diverse as natural resources, ecology, epidemiology, plant biology, and public health because of its simple form and interpretability (Agresti 2007). Existing data on favorable and unfavorable locations for woody biomass using facilities were used to train logistic regression models. These models were then used to evaluate the suitability of new locations. Economic factors, transportation-related factors, and the availability of biomass feedstocks were included as predictor variables in the logistic regression models.

METHODS

The strategy followed was to create a probability-based approach to identify suitable site locations. Existing site data were used to fit a logistic model. The resulting fitted models provided probabilities and a combination of predictor variables that were used to identify new locations for woody biomass-using facilities. Twenty-five locations with the highest probabilities are reported for a 13-state study area.[1]

Logistic Regression Models

Logistic regression models the relationship between a two-level categorical response variable (binary response) and explanatory variables which can be continuous or categorical. The outcomes of the response variable y are coded $y=1$ or $y=0$ with respective probabilities p and $1-p$. For a particular subject, the logistic regression equation is,

$$\log (p/[1-p]) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n, \tag{1}$$

where $x_1, x_2, \ldots x_n$ are measurements on a group of predictor variables which relate the probability of an outcome, $y=1$.

An automated procedure, stepwise logistic regression, was implemented for the data set of woody biomass-using facilities to determine which variables played a significant role in determining the quality of a potential site. In addition, a variable transformation procedure was used to preprocess the data set to explore improved predictive ability of the logistic regression.[2] Four methods were used to build the logistic regression models:

> ***Method 1****: Logistic regression without stepwise variable selection and variable transformation;*
> ***Method 2****: Logistic regression with stepwise variable selection only;*
> ***Method 3****: Logistic regression with variable transformation only;*
> ***Method 4****: Logistic regression with both variable selection and variable transformation.*

The full model included in the variable selection procedure contained the full set of 31 explanatory variables along with all two-way interactions.

In each model, the data set was partitioned into two parts: 60 percent of the data were randomly selected to train (i.e. fit) the model while the remaining 40 percent of the data were used to evaluate the quality of the model predictions by constructing a classification table (i.e., validation data set). The classification table cross classifies the actual binary outcomes of y with the models prediction $\hat{y} = 0$ or $\hat{y} = 1$.

---

[1]Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, and Virginia.
[2]A logarithmic variable transformation was included to maximize the linear relationship between the independent variables and the response.

The four methods (above) were compared by the Akaike's Information Criteria (AIC) and Bayesian Information Criterion (BIC), which are defined as,

$$-2ln(L_m) + 2m, \qquad (2)$$

$$-2L_m + mln(n), \qquad (3)$$

where n is the sample size, L_m is the maximized *log-likelihood* of the model, and m is the number of parameters in the model. The AIC and BIC take into account both the statistical goodness-of-fit and the number of parameters in order to avoid modeling the noise in the data also known as overfitting (McQuarrie and Tsai 1998).

Response and Explanatory Variables

This study involved organizing large volumes of data collected from various sources, including the U.S Census Bureau, U.S. Forest Service, railroad companies, and trucking firms (U.S. Census Bureau 2000; University of Wisconsin-Milwaukee Employment and Training Institute 2000; Perlack et al. 2005; U.S. Army Corps of Engineers 2008; U.S. Department of Agriculture Forest Service 2008; IEA Bioenergy Task 39 2009; Renewable Fuels Association 2009; Galik et al. 2009).[3]

Another resource was the BioSAT model data from Young et al. (2008). The BioSAT model estimated marginal cost, average cost, total cost of delivered mill and logging residues, and quantity of available mill and logging residues within 128.8 km haul distance. The BioSAT model estimates the cost and availability of woody biomass for procurement zones which may not be concentric, i.e., the shape of such zones rely on the available transportation network and biomass supply. National forests, parks, urban areas, and other restricted areas were not considered in BioSAT when estimating availability. Travel times and distances were estimated from the Microsoft[©] MapPoint[®] 2006 (http://www.microsoft.com/MapPoint/en-us/default.aspx). Road networks in MapPoint[®] were a combination of the Geographic Data Technology, Inc. (GDT) and Navteq data. GDT data were used for rural areas and small to medium size cities. Navteq data were used for major metropolitan areas. In the BioSAT model, estimates of all-live total biomass, as well as average annual growth, removals, and mortality were obtained from the Forest Inventory and Analysis Database (FIADB) version 3.0 (http://srsfia2.fs.fed.us/). The Subregional Timber Supply (SRTS) model was used to estimate current logging residues and project future logging residues (Abt 2008). The Fuel Reduction Cost Simulator (FRCS), as modified for the Billion Ton Study, was used to estimate the costs of harvesting logging residues (Fight et al. 2006; Dykstra 2008).

---

[3]Railroad location data were collected from: Burlington Northern Santa Fe Railway, CSX Corporation, Inc., CSX Corporation, Inc., Norfolk Southern System, and Union Pacific. Trucking cost rate data were collected from Pemberton Truck Lines (TN), Skyline Transportation, Inc. (TN), Mason Dixon (TN), Mason Dixon (AL), Patterson Chip Company (KY), Tennessee Department of Agriculture (Nashville, TN).

All records were organized at the U.S. Census Bureau 5-digit ZIP Code Tabulation Area (ZCTA) level (U.S. Census Bureau 2000). There were 9,221 ZCTAs in the 13-state study region which corresponded to 9,221 potential site locations for woody biomass-using facilities. The average area for 5-digit ZCTAs in the 13-state study region was 237.48 km^2. Plant locations were selected based on the outcomes of the logistic models; most variables were related to biomass supply and cost. These supply and cost outcomes were also noted in a UK study by Adams et al. (2010). The studies by Velázquez-Martí and Annevelink (2009) and Velázquez-Martí and Fernandez-Gonzalez (2010) focused on point plant location siting based on the consumption areas. However, the aforementioned studies could have also focused on biomass source areas which were the focus of this study.

Group I and II Subsets

In this study, bioenergy and biofuel plants were defined as facilities that use all possible wood residues in an integrated biomass conversion process to produce wood pellets for energy, biofuels, biopower, or biochemicals (National Renewable Energy Laboratory 2009). Only 29 such facilities existed 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 were used as surrogates (e.g., primary wood mills, secondary wood mills, and pulp and paper mills). The assumption was that similar factors may affect the attractiveness and suitability of a site given the commonality in feedstocks (recall Pätäri 2009; Knight 2009; Stewart 2009; Cohen et al. 2010; Conrad et al. 2010). Definitions for locations were:

> **Group I**: *All wood-using mills and wood-using bioenergy or biofuels plants;*

> **Group II**: *Pulp and paper mills and wood-using bioenergy or biofuels plants.*

Group I locations are illustrated in Fig. 1.[4] Group II mills are illustrated in Fig. 2.

Response and explanatory variables

Two separate response variables were considered for modeling and ranking of potential sites. For Group I the response variable, $y_{i1}=1$, if the ith ZCTA has a woody biomass-using facility and $y_{i2}=1$ was defined similarly for a Group II mill. Thirty-one explanatory variables were examined in the logistic models (Table 1).

---

[4]As defined by Perlack et al. (2005), primary wood processing mills include sawmills, medium density fiberboard, oriented strand board, particleboard, plywood, veneer post, pole, piling, dealer, yard, energy, and wood chips. Secondary mills utilize the products of primary mills (e.g., millwork, containers and pallets, buildings, furniture, flooring, paper and paper products. Secondary wood processing mills also include mills processing the above products but also produce planed wood products, remanufactured wood products, trusses, moldings, kiln dried products, treated wood products, plants, decking, and siding.
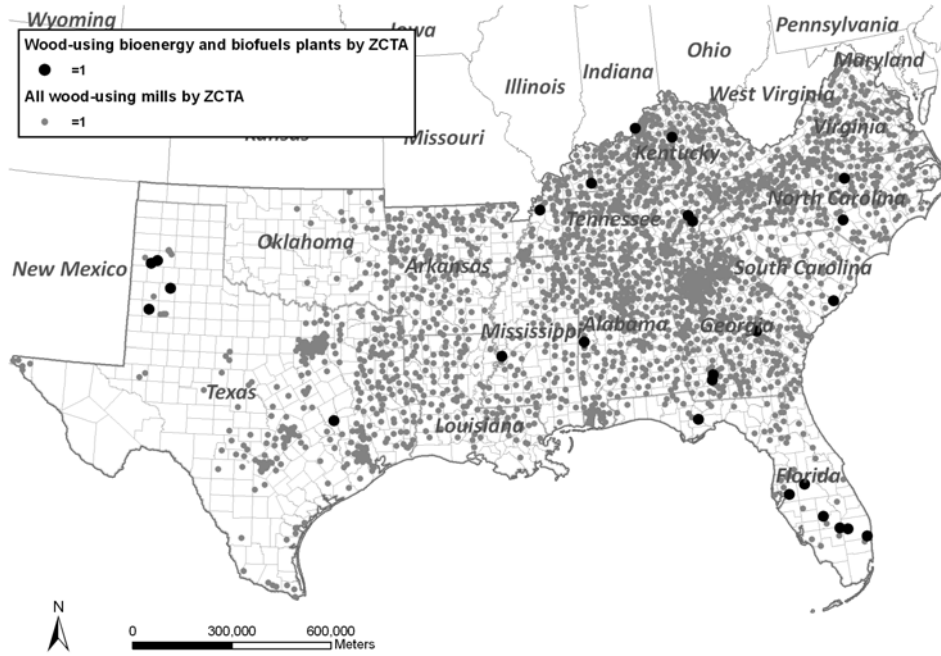
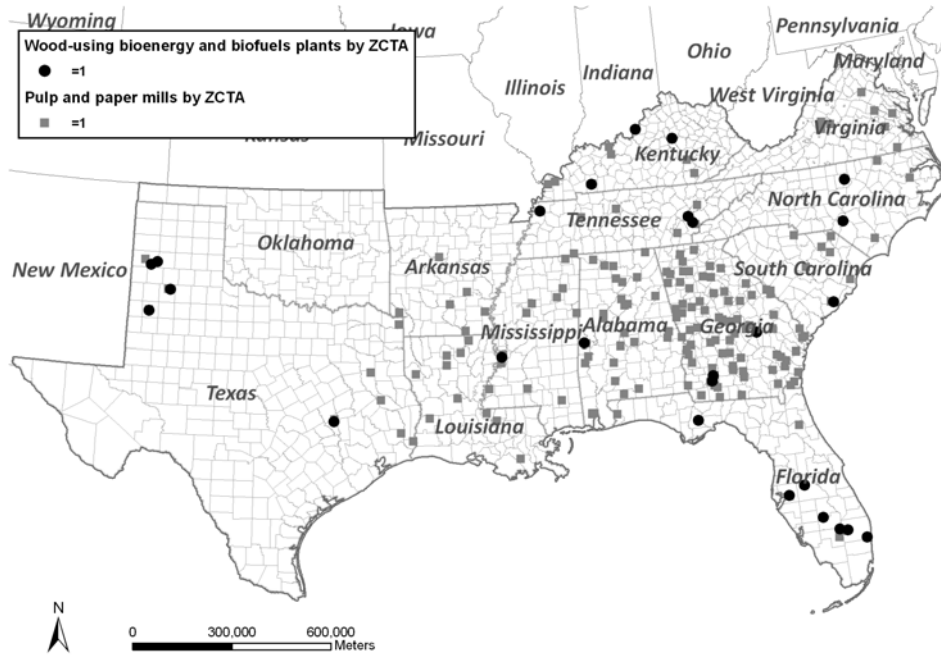**Fig. 1.** Illustration of Group I woody biomass plants

Fig. 2. Illustration of Group II woody biomass plants

Table 1. Explanatory Variables for Two Study Groups Organized by ZCTA

Description	Definition
Employment	Employed persons for all industries
Population	Total population
Population_Density	Population density (people/km^2)
Sqmiwater	Water area (km^2)
Median_Family_Income	Median of family incomes in 1999 $
Income_index	Median family income per employed person ($/person)
LOG_RES_HW	Logging residues of hardwoods (dry tonnes)
LOG_RES_SW	Logging residues of softwoods (dry tonnes)
LOG_RES_TOT	Logging residues of both (dry tonnes)
OTHR_REM_HW	Other removal of hardwoods (dry tonnes)[5]
OTHR_REM_SW	Other removal of softwoods (dry tonnes)
OTHR_REM_TOT	Other removal of both (dry tonnes)
THIN_40	Thinnings to a basal area 9.2 (m^2/ha)
THIN_80	Thinnings to a basal area 18.4 (m^2/ha)
THIN_120	Thinnings to a basal area 27.6 (m^2/ha)
THIN_200	Thinnings to a basal area 445.9 (m^2/ha)
TOTAL_MILL_RES	Total mill residues supplied for each ZCTA (dry tonnes)[6]
UNUSED_MILL_RES	Total unused mill residues for each ZCTA (dry tonnes)
MCost_p5M	Marginal cost of trucking 0.5 million tons within 128-km ($/dry tonne)
MCost_1M	Marginal cost of trucking 1.0 million tons within 128-km ($/dry tonne)
MCost_1p5M	Marginal cost of trucking 1.5 million tons within 128-km ($/dry tonne)
TCost_80	Total trucking cost within an 80- 128-km haul distance ($/dry tonne)
ACost_80	Average trucking cost within an 128-km haul distance ($/dry tonne)
TQty_80	Total quantity of mill residues within an 128-km haul distance
URBAN_WASTE	Municipal solid waste, construction and demolition debris (dry tonnes)
Log_Res_Harvest_Cost	Harvesting cost of logging residues ($/dry tonne)
RailroadAvailability_#	Railroad accessible index (ordinal ranking of 0, 1, 2, 3, 4)[7]
Numberports	Number of water ports in each ZCTA (categorical)
Primary_mill_total	Number of primary wood processing mills in each ZCTA (categorical)
Secondary_mill_total	Number of secondary wood processing mills in each ZCTA (categorical)
Other_Mill_total	Number of other wood processing mills in each ZCTA (categorical)

Model Scoring and Interpretation

Given a specific response variable and set of predictor variables, the fitted logistic regression model provided an estimated probability that a ZCTA will contain a woody biomass-using facility. The estimated probability was used in the validation data set to

---

[5] Other removal is the unutilized wood volume from cut or otherwise dead growing stock, from cultural operations, e.g., precommercial thinning, or timberland clearing. This doesn't include volume removed from inventory through reclassification of timberland to productive reserved forest land (Perlack et al. 2005)

[6] Residues generated from primary mills, secondary mills and pulp and paper mills, which include bark, coarse residues (chunks and slabs), fine residues (shavings and sawdust), sawdust, sander dust, wood chips and shavings, board and cut-offs, miscellaneous scrap wood and black liquor ("solution of lignin-residue and the pulping chemicals used to extract lignin during the manufacture of paper" (Perlack et al. 2005)

[7] Railroad accessible index ranked by four railroad companies as 0, 1, 2, 3, and 4. "0" means ZCTA has no railroad; "1" means one out of four railroad companies in this ZCTA has railroad access.

compare ZCTAs with actual mill locations. This probability was used to rank potential ZCTAs for new woody biomass mill location.

RESULTS AND DISCUSSION

Logistic Regression Models - Group I

The AIC, BIC values and misclassification rates of the four models are given in Table 2. The logistic model using Method 3 had the lowest AIC score of 1923.35, lowest BIC score of 1977.13, and the lowest misclassification rate of 12.5 percent.

Table 2. Group I Model Results by AIC, BIC Criteria and Misclassification Rate

Model Method	AIC Value	BIC Value	Misclassification Rate[8]
Model 1: Logistic regression without variable selection and variable transformation	2001.71	2055.50	0.1305
Model 2: Logistic regression with stepwise variable selection only	1992.96	2034.80	0.1374
Model 3: Logistic regression with variable transformation only	**1923.35**	**1977.13**	**0.1250**
Model 4: Logistic regression with both variable selection and variable transformation	1927.71	1975.52	0.1288

The classification table confirms that logistic model using Method 3 had good predictive power for the siting locations of the Group I woody biomass-using facilities (Table 3). The sensitivity of this model in the validation data set was $P(\hat{y}=1 \mid y=1) =$ 86.7% (e.g., the model predicted a mill location correctly 86.7% of the time in validation), and specificity (e.g., predicting no mill location where there is currently no mill) was $P(\hat{y}=0 \mid y=0) = 87.0\%$ (Table 4). The sensitivity rates of the model in the training and validation data sets were acceptable, if based on the stringent criteria of medical radiology screening (Carney et al. 2010).[9] The specificity rates of the model in the training data set were acceptable and very close to acceptable in the validation data set.

Table 3. Classification Table for Training Data Set for Group I

Actual value	Predictive Value		
	0	1	Total
0	**1251 (89.4%)**	148	1399
1	216	**1296 (85.7%)**	1512
Total	1467	1444	2911

---

[8] The misclassification is an error rate, e.g., for Table 3(148+216)/2911=0.1250

[9] In medical research (see Carney et al. 2010) final cut points to identify low performance for screening mammography were: sensitivity less than 75%, specificity less than 88% or greater than 95%,

---

Table 4. Classification Table for Validation Data Set for Group I

Actual value	Predictive Value		
	0	1	Total
0	**808 (87.0%)**	122	929
1	134	**876 (86.7%)**	1011
Total	942	998	1940

"Railroad availability" (RailroadAvailability_2), "thinnings to a basal area of 31.7m^2/ha," (Thin_80), and "logging residue harvest costs" (Log_Res_Harvest_Cost) were highly significant in influencing mill location, p-values < 0.0001 (Table 5). Other statistically significant variables were "thinnings to a basal area of 47.5m^2/ha" (Thin_120), "median family income" (Median_Family_Income), "unused mill residues" (UNUSED_MILL_RES), "population" (Total population), and "railroad availability" (RailroadAvailability_1). A higher density of railroad availability (RailroadAvailability_2) had a positive coefficient relative to a lower density of railroad availability (RailroadAvailability_1) which had a negative coefficient. This may suggest that a higher density of railroads has an economic advantage over one railroad provider and its oligopoly market influence, see Ivaldi and McCullough (2001) on railroad integration and oligopoly advantage. "Thinnings to a basal area of 31.7m^2/ha" (Thin_80) had a positive coefficient, while "thinnings to a basal area of 47.5m^2/ha" (Thin_120) had a negative coefficient. This may indicate that availability of higher volumes of woody biomass from thinning is preferred over lower volumes of thinning. This result is in agreement with other studies as related to the importance of biomass supply (Adams et al. 2010; Stewart 2009; Conrad et al. 2010).

Table 5. Analysis of Maximum Likelihood Estimates of Parameters for Group I

Parameter	Estimate	Wald Chi_Square	Pr>ChiSq
Intercept	-0.146	0.84	0.3607
RailroadAvailability_2	0.579	47.0	<0.0001
Thin_80	11.287	39.25	<0.0001
Log_Res_Harvest_Cost	-0.218	132.49	<0.0001
Thin_120	-6.092	12.64	0.0004
Median_Family_Income	-0.163	8.61	0.0033
UNUSED_MILL_RES	7.746	7.27	0.0070
Population	-0.325	6.21	0.0127
RailroadAvailability_1	-0.218	4.09	0.0430

Unused mill residues had a positive influence on mill location, while costs of logging residue harvesting (recall Adams et al. 2010), median family income, and population have negative influences mill locations, i.e., it is not surprising that population (and higher incomes associated with urban areas) had a negative influence on mill locations which would be nearer the rural biomass supply.

The 25 mill locations (ZCTAs) with the highest probability from the logistic model (Method 3) are given in Fig. 3. There were ten possible locations in Mississippi, eight in Tennessee, three in Virginia, three in Louisiana, and one in Georgia.
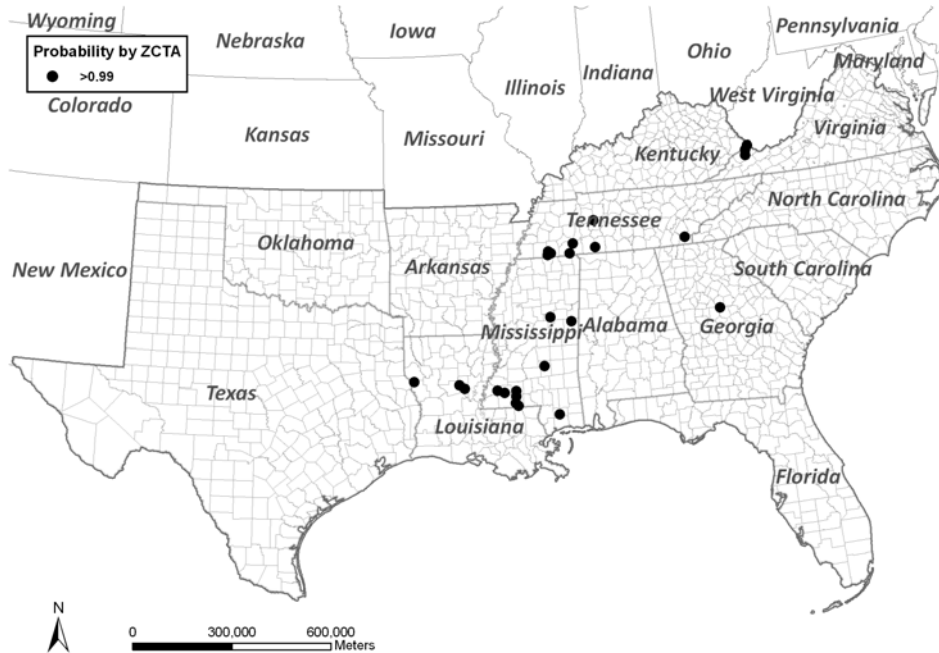
Fig. 3. Highest probability locations for Group I

Logistic Regression Models - Group II

The AIC values, BIC values, and misclassification rates for the logistic models are given in Table 6. The logistic model using Method 3 had the lowest AIC score of 322.19, lowest BIC score of 360.55, and the lowest misclassification rate of 2.5 percent.

Table 6. Group II Model Results by AIC, BIC Criteria and Misclassification Rate

Model Method	AIC Value	BIC Value	Misclassification Rate
Model 1: *Logistic regression without variable selection and variable transformation*	324.32	362.68	0.0248
Model 2: *Logistic regression with stepwise variable selection only*	355.89	383.29	0.0299
Model 3: Logistic regression with variable transformation only	**322.19**	**360.55**	**0.0254**
Model 4: *Logistic regression with both variable selection and variable transformation*	358.69	386.10	0.0299

The classification table for Group II training data had a sensitivity of $P(\hat{y} = 1 | y = 1) = 66.4\%$, specificity $P(\hat{y} = 0 | y = 0) = 99.5\%$ (Table 7). The sensitivity of this model in the validation data set was $P(\hat{y} = 1 | y = 1) = 60.5\%$, and specificity was $P(\hat{y} = 0 | y = 0) = 99.1\%$ (Table 8). The specificity rates of the model in the training and validation data sets were close to acceptable (recall footnote 10). The sensitivity rates of the model in the training and validation data sets were outside the cut points used by the medical industry (recall Carney et al. 2010). The logistic model for Group II appears better at predicting locations where not to locate a plant then at predicting preferred site locations.

Table 7. Classification Table for Training Data Set for Group II

Actual value	Predictive Value		
	0	1	Total
0	**1655 (99.5%)**	8	1663 (93.8%)
1	37	**73 (66.4%)**	110 (6.2%)
Total	1692	81	1773

Table 8. Classification Table for Validation Data Set for Group II

| Actual value	Predictive Value		
	0	1	Total
0	**1091 (99.1%)**	10	1101 (93.8%)
1	32	**49 (60.5%)**	81 (6.2%)
Total	1123	59	1182

"Logging residue harvest costs" (Log_Res_Harvest_Cost), population, and "thinnings to a basal area of 79.2m^2/ha" (Thin_200) were highly significant in influencing mill location, p-values < 0.0001 (Table 9). Other statistically significant variables were "number of secondary mills" (Secondary_Mill_Total), "number of primary mills" (Primary_Mill_Total), and "amount of total mill residues" (TOTAL_MILL_RES). "Logging residue harvest costs" and population had negative influences on mill location for Group II. "Thinnings to a basal area of 79.2m^2/ha" had a positive influence on mill location. This may reflect that this group may accept smaller-sized woody biomass feedstocks relative to Group I. "Number of secondary mills," "number of primary mills," and "amount of total mill residues" all had positive influences on Group II facilities. This may reflect the synergistic relationship that exists between primary and secondary wood-using mills and larger Group II facilities that may depend on their residue feedstocks. The feasibility of residues for wood-based materials and bioenergy has been documented (Garay et al. 2009).

Table 9. Analysis of Maximum Likelihood Estimates of Parameters for Group II

Parameter	Estimate	Wald Chi_Square	Pr>ChiSq
Intercept	-2.0490	22.64	<0.0001
Log_Res_Harvest_Cost	-0.3060	20.67	<0.0001
Population	-2.0097	23.62	<0.0001
Thin_200	4.4776	66.03	<0.0001
Secondary_Mill_Total	0.9600	12.13	0.0005
Primary_Mill_Total	1.2937	10.19	0.0014
TOTAL_MILL_RES	1.3233	7.54	0.0060

The importance of biomass supply (mill residues) also supports previous studies (Adams et al. 2010; Stewart 2009; Conrad et al. 2010). These results are also in agreement with the study by Pätäri (2009) where the complementary resources held by forest and energy companies make collaboration in the bioenergy business favorable. The significance of harvesting costs in the logistic model is also noted by Pätäri (2009), where a synergistic relationship with pulp and paper mill locations has been considered by Cohen et al. (2010).

The 25 potential locations for Group II with the highest probabilities are plotted in Fig. 4. The logistic model identified seven possible locations in Georgia, six in North Carolina, four in Arkansas, three in Mississippi, two in Kentucky, two in Texas, and one in Louisiana.
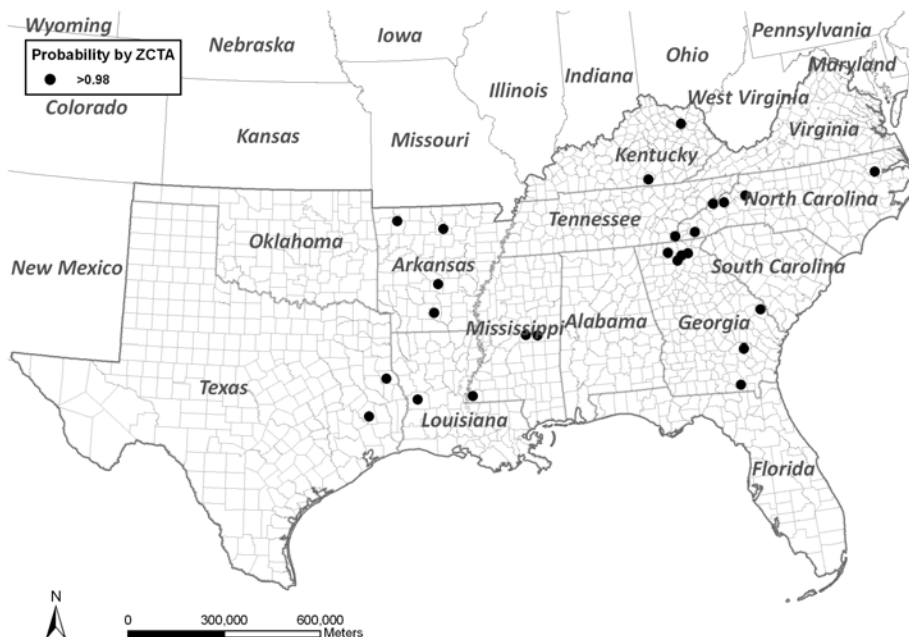
Fig. 4. Highest probability locations for Group II

CONCLUSIONS

1. Logistic regression models developed in this study provided quantitative insight and a probability-based approach to understanding variables that influence the location of woody biomass-using facilities. "Thinnings to a basal area of $31.7m^2/ha$," "availability of unused mill residues," and "a high density of railroad availability" have a positive significant influences on all woody biomass-using facilities. For larger woody biomass-using mills (e.g., pulp and paper or biopower) "availability of thinnings to a basal area of $79.2m^2/ha$," "number of primary and secondary wood-using mills within a 128.7 km haul distance," and "amount of total mill residues" have positive significant influences on site location.
2. The logistic model quantified the synergistic relationship between primary and secondary wood-using mills, and larger, wood-using facilities (e.g., pulp and paper, or biopower) by predicting correctly mill locations 86.7% of the time in validation.

ACKNOWLEDGMENTS

REFERENCES CITED

Abt, R. C. (2008). "Sub-regional timber supply model - data, model, and projection updates," Presentation at SOFAC IV, North Carolina State University, Raleigh, NC, http://www.ces.ncsu.edu/nreos/forest/feop/SOFAC2008meeting-info.html.

Adams, P. W., Hammond, G. P., McNanus, M. C., and Mezzullo, W. G. (2010). "Barriers to and drivers for UK bioenergy development," *Renewable and Sustainable Energy Reviews*, In Press, doi:10.1016/j.rser.2010.09.039

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, John Wiley and Sons, Hoboken, NJ.

Altman, I., and Johnson, T. (2008). "The choice of organizational form as a non-technical barrier to agro-bioenergy industry development," *Biomass Bioenerg* (32), 28-34.

Blanchard, O. J., and Gali, J. (2007). "The macroeconomic effects of oil price shocks: why are the 2000s so different from the 1970s?" Proc. of NBER ME Conference on International Dimensions of Monetary Policy. S'Agaro, Catalonia, Spain. http://www.crei.cat/people/gali/pdf_files/bgoil07wp.pdf

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/.

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

Cheng, S. M., and Zhu, S. D. (2009). "Lignocellulosic feedstock biorefinery – The future of the chemical and energy industry," *Bioresources* 4(2), 456-457.

Cohen, J., Janssen, M., Chambost, V., and Stuart P. 2010. "Critical analysis of emerging forest biorefinery (FBR) technologies for ethanol production," *Pulp and Paper Canada*. 111(3 May/June), T42-T48.

Conrad, J. L. IV, Bolding, M. C., Aust, W. M., and Smith, R. L. (2010). "Wood-to-energy expansion, forest ownership changes, and mill closure: Consequences for U.S. South's wood supply chain," *Forest Policy and Economics* 12(6), 399-406.

Dasmohapatra, S. (2009). "Future market drivers for the forest products industry," *Bioresources* 4(4), 1263-1266.

Dykstra, D. P. (2008). "Subject: estimating biomass collection costs for the 'Billion-Ton Study' update. Memo: "Estimating forest biomass collection costs for the billion-ton study update (BTS2)". Dykstra, April 25, 2008.

Energy Information Administration. (2008). *International Energy Outlook,* Report #:DOE/EIA-0484(2008). http://www.eia.doe.gov/oiaf/ieo/world.html

Fight, R. D., Hartsough, B. R., and Noordijk, P. (2006). *Users Guide for FRCS: Fuel Reduction Cost Simulator Software*, Gen. Tech. Rep. PNW-GTR- 668, Portland, Research Station, Forest Service, U.S. Department of Agriculture. http://www.fs.fed.us/pnw/pubs/pnw_gtr668.pdf .

Galik, C. S., Abt, R. C., and Wu, Y. (2009). "Forest biomass supply in the Southeastern United States: Implications for industrial roundwood and bioenergy production," *J. For.* 107(2), 69-77.

Garay, R. M., McDonald, F., Acevedo, M. L., Calderon, B., and Araya, J. E. (2009). "Particleboard made with crop residues mixed with wood from Pinus radiata," *Bioresources* 4(4), 1396-1408.

Gronowska M., Joshi S., and MacLean, H. L. (2009). "A review of U.S. and Canadian biomass supply studies," *Bioresources* 4(1), 341-369.

IEA Bioenergy Task 39. (2009). "Status of 2nd generation biofuels demonstration facilities," [Data file], Retrieved from <http://biofuels.abc-energy.at/demoplants>.

Hubbe, M. A., and Buelhmann U. A. (2010). "A continuing reverence for wood," *BioResources* 5(1), 1-2.

Ivaldi, M., and McCullough, G. (2001). "Density and integration effects on class I U.S. freight railroads," *J. Regul. Econ.* 19 (2), 161-182.

Knight, D. K. (2009). *Wood Bioenergy*, Hatton-Brown Publishers, Inc., Montgomery, AL.

Kumarappan, S., Joshi, S., and MacLean, H. L. (2009). "Biomass supply for biofuel production: Estimates for the United States and Canada," *Bioresources* 4(3), 1070-1087.

Lucia, L. A. (2008). "Lignocellulosic biomass: A potential feedstock to replace petroleum," *BioResources* 3(4), 981-982.

Lynd, L. R. (1996). "Siting an ethanol plant in the Northeast*,* Burlington," VA: C.T., Donovan Associates.

McQuarrie, A. D. R., and Tsai, C.-L. (1998). *Regression and Time Series Model Selection*, Singapore [u.a.]: World Scientific Publishing.

National Renewable Energy Laboratory. (2009). "What is a biorefinery?" [Biorefinery definition]. Retrieved from <http://www.nrel.gov/biomass/biorefinery.html>.

Pätäri, 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.

Pawlak, J. J. (2008). "A sustainable economy," *BioResources* 3(1), 1-2.

Perlack, R. D., Wright, L. L., Turhollow, A. F., Graham, R. L., Stokes, B. J., and Erbach, D. C. (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.

Renewable Fuels Association. (2009). "Biorefinery locations*,*" [Data file], Retrieved from <http://www.ethanolrfa.org/industry/locations/>.

Sperling, D. (1984). "An analytical framework for siting and sizing biomass fuel plants," *Energy* 9(11-12), 1033-1040.

Stewart, P. (2009). "Energy and the wood fiber supply chain," *Forests & People Magazine* (Q1). http://www.forest2market.com/f2m/us/f2m1/free/publications/Energy-and-Wood-Fiber-Supply-Chain.

University of Wisconsin-Milwaukee Employment and Training Institute. (2000). "Workforce and household/income data," [Data file], Available from <http://www4.uwm.edu/eti/PurchasingPower/ETIshapefiles.htm>.

U.S. Army Corps of Engineers Navigation Data Center. (2008). "U.S. waterway data," [Data file], Retrieved from <http://www.ndc.iwr.usace.army.mil/data/datapwd.htm>.

U.S. Census Bureau. (2000). "Population, land area, and water area data," [Data file], Available from <http://www.census.gov/geo/www/gazetteer/places2k.html>.

U.S. Department of Agriculture Forest Service. (2008). *Woody Biomass Utilization Strategy*, U.S. Department of Agriculture, FS-899, GPO: Washington, D.C.

Young, T. M., Ostermeier, D. M., Thomas, J. D., and Brooks, R. T. (1991). "The economic availability of woody biomass for the Southeastern United States," *Bioresource Techn* 37(1), 7-15.

Young, T. M., Perdue, J. H., Hartsell, A., Abt, R. C., Hodges, D. G., and Rials, T. G. (2008). "A real-time, web-based optimal Biomass Site Assessment Tool (BioSAT): Module 1 - An economic assessment of mill residues for the southern," U.S. Forest Inventory and Analysis (FIA) Symposium 2008, October 21-23, 2008, Park City, UT. Retrieved from http://www.fs.fed.us/rm/pubs/rmrs_p056/rmrs_p056_42_young.pdf>.