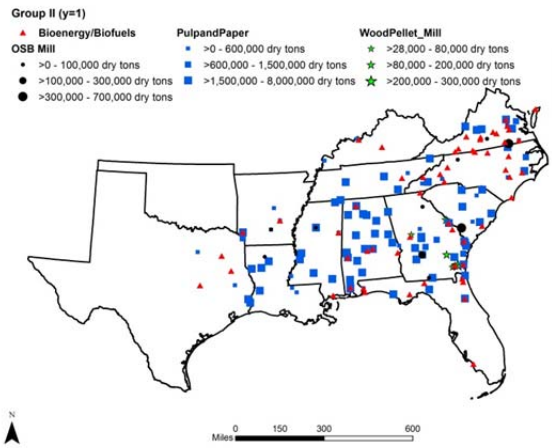
Modeling the Impact of the Emerging Bioeconomy on Transportation Network Flows with Simulation and Bayesian Inference



Final Report

U.S. Department of Transportation

Southeastern SunGrant Center

February, 2014

Principal Investigators:

Timothy M. Young, Ph.D.
Professor
Department of Forestry, Wildlife and Fisheries
Center for Renewable Carbon
The University of Tennessee
2506 Jacob Drive, Knoxville, TN 37996-4570
tmyoung1@uk.edu, 865.946.1119

Lee D. Han, Ph.D.
Professor and Program Coordinator
Civil and Environemntal Engineering
The University of Tennessee
112 Perkins Hall, Knoxville, TN 37996-2010
lhan@utk.edu, 865.974.7707

Frank M. Guess, Ph.D.
Professor
Department of Statistics, Operations, and Management Science
The University of Tennessee,
335 Stokely Mgmt Cntr, Knoxville, TN 37996-0532
fguess@utk.edu, 865.974.1637

Co- Principal Investigators:

Stephanie R. Hargrove
Transportation Engineering Doctoral Student
Civil and Environemntal Engineering
The University of Tennessee
112 Perkins Hall, Knoxville, TN 37996-2010
shargro1@gmail.com
865.406.3837

Xia Huang
Research Associate
Center for Renewable Carbon
The University of Tennessee
2506 Jacob Drive, Knoxville, TN 37996-4570
xhuang8@uk.edu, 865.946.1129

James H. Perdue
Senior Biological Scientist
Decision Support Research
USDA Forest Service
Southern Research Station
Office at The University of Tennessee
Center for Renewable Carbon
2506 Jacob Drive, Knoxville, TN 37996-4570
jperdue@fs.fed.us, 865.946.1123

Co- Principal Investigators:

Chung-Hao Chen, Ph.D.
Assistant Professor
Department of Electrical and Computer Engineering
Old Dominion University
231B Kaufman Hall
Norfolk, VA 23529
cxchen@odu.edu, 757.683.3475

Acknowledgements

Support for this research was provided in part by a grant from the Southeastern Sun Grant Center with funds provided by the U.S. Department of Transportation Research and Innovative Technology Administration. Support for this research was also provided in part by The University of Tennessee, Center for Renewable Carbon, Department of Civil Engineering, Department of Forestry Wildlife and Fisheries, and U.S. Forest Service, Southern Research Station.

Contents

List of Tables vi

List of Figures vii

Abstract 1

Executive Summary 2

1. Introduction 5

1.1 Objectives and Scope 5

2. Methodology 6

2.1 Transportation Network Simulation Flow Model 6

2.1.1 Overview 6

2.2 Logistic Regression Model for Siting Biorefineries 12

2.2.1 Dataset 12

2.2.2 Logistic Regression 16

2.2.3 Study Group 18

2.2.4 Response and Explanatory Variables 20

2.2.5 Modeling Scoring and Interpretation 21

3. Results and Discussion 22

3.1 Group I 22

3.2 Group II 27

3.3 Trucking Costs and Delay Times 32

4. Conclusions 34

5. References 35

Appendix A. 38

Appendix B. 65

List of Tables

Table 1. Level of service criteria for arterials based on volume-to-capacity ratios. …….. 7

Table 2. V/C ratio, arrival time, and LOS for existing traffic. ……. 7

Table 3. Explanatory variables organized by ZCTA. …….. 20

Table 4. Significant variables for Group I mills. ……. 22

Table 5. Summary of classification table for training dataset for Group I mills. …….. 22

Table 6. Summary of classification table for training dataset for Group I mills. …….. 22

Table 7. Analysis of variable estimates from Bayesian inference with a uniform prior for Group I mills. …….. 24

Table 8. Significant variables for Group II mills. …… 27

Table 9. Summary of classification table for training dataset for Group II mills. ……. 27

Table 10. Summary of classification table for validation dataset for Group II mills. …. 27

Table 11. Analysis of significant variables estimates from Bayesian inference with a uniform prior for Group II mills. ……. 29

Table 12. Descriptive statistics of trucking costs ($/dry ton) by state as estimated from the BioSAT model. ……. 31

Table 13. Statistical interval of trucking costs ($/dry ton) by state adjusted for average delay time per state. ……. 32

List of Figures

Figure 1. City of Vonore TN network model. ……..…….. 10

Figure 2. Simulation of a signal controlled intersection. …… 10

Figure 3. Simulation of two STOP controlled intersections. ……. 11

Figure 4. Illustration of forest biomass allocation at the level of 5-digit ZCTA. …… 14

Figure 5. Illustration of Group I woody biomass-using mills. ……. 18

Figure 6. Illustration of Group II woody biomass-using mills. …… 19

Figure 7. ROC charts for model assessment of Group I woody biomass-using mills. …. 23

Figure 8. Lift Charts for model assessment of Group I woody biomass-using mills. …… 23

Figure 9. Estimated probability locations for Group I. ……. 25

Figure 10. ROC charts for model assessment of Group II woody biomass-using mills. ... 28

Figure 11. Lift Charts for model assessment of Group II woody biomass-using mills. … 28

Figure 12. Estimated probability locations for Group II. ……. 30

Appendix A

Figure A1. Estimated probability locations for Group I in Alabama. 39

Figure A2. Estimated probability locations for Group II in Alabama. 40

Figure A3. Estimated probability locations for Group I in Arkansas. 41

Figure A4. Estimated probability locations for Group II in Arkansas. 42

Figure A5. Estimated probability locations for Group I in Florida. 43

Figure A6. Estimated probability locations for Group II in Florida. 44

Figure A7. Estimated probability locations for Group I in Georgia. 45

Figure A8. Estimated probability locations for Group II in Georgia. 46

Figure A9. Estimated probability locations for Group I in Kentucky. 47

Figure A10. Estimated probability locations for Group II in Kentucky. 48

Figure A11. Estimated probability locations for Group I in Louisiana. 49

Figure A12. Estimated probability locations for Group II in Louisiana. 50

Figure A13. Estimated probability locations for Group I in Mississippi. 51

Figure A14. Estimated probability locations for Group II in Mississippi. 52

Figure A15. Estimated probability locations for Group I in North Carolina. 53

Figure A16. Estimated probability locations for Group II in North Carolina. 54

Figure A17. Estimated probability locations for Group I in Oklahoma. 55

Figure A18. Estimated probability locations for Group II in Oklahoma. 56

Figure A19. Estimated probability locations for Group I in South Carolina. 57

Figure A20. Estimated probability locations for Group II in South Carolina. 58

Figure A21. Estimated probability locations for Group I in Tennessee. 59

Figure A21. Estimated probability locations for Group II in Tennessee. 60

Figure A23. Estimated probability locations for Group I in Texas. 61

Figure A24. Estimated probability locations for Group II in Texas. 62

Figure A25. Estimated probability locations for Group I in Virginia. 63

Figure A26. Estimated probability locations for Group II in Virginia. 64

Appendix B

Figure B1. Sample results of travel simulation with 25MG local trucks at intersections near Vonore, TN.....66

Figure B2. Sample results of travel simulation with 50MG local trucks at intersections near Vonore, TN.....67

Figure B3. Sample results of travel simulation with 75MG local trucks at intersections near Vonore, TN.....68

Figure B4. Sample results of travel simulation with 100MG local trucks at intersections near Vonore, TN.....69

Abstract

This project addresses the U.S. Department of Transportation Southeastern Sun Grant priority "*Impact of the emerging bioeconomy on transportation infrastructure and related concerns*." The impact of the emerging bioeconomy on truck transportation network flow was modeled for the southern United States. The study applied Bayesian logistic regression and GIS spatial analysis to estimate site locations for biorefineries in the presence of uncertainty as related to resource competition, delivered costs at the plant gate, and trucking transportation flow. Procurement zones assumed a one-way haul distance of 80 miles as a function of the available road network and interrelated resource as estimated from the BioSAT model www.biosat.net. "*Median Family Income*", "*Timberland Annual Growth-to-Removal Ratio*", and "*Transportation Delays*" were highly significant in influencing mill location. Transportation delays for trucks greatly impact the cost of trucking biomass resources. In some areas of the southern U.S., it increases trucking costs by as much as 60%. The logistic model using Bayesian inference is a good predictive model for predicting preferred site locations and identifying non-preferred locations. The sensitivity of the Bayesian logistic regression models in the validation data set was 89% and specificity was 92.3% for small capacity mills and 88.1% in sensitivity and 89.2% specificity for larger capacity mills similar to pulp and paper mills. The higher probability locations for smaller biomass mills are in southern Alabama, southern Georgia, southeast Mississippi, southern Virginia, western Louisiana, western Arkansas, and eastern Texas. The higher probability locations for large capacity mills are in southeastern Alabama, southern Georgia, central North Carolina, and the Mississippi Delta regions.

Executive Summary

This project addresses the U.S. Department of Transportation Southeastern Sun Grant priority "*Impact of the emerging bioeconomy on transportation infrastructure and related concerns*." The impact of the emerging bioeconomy on truck transportation network flow was modeled for 13 southern states.[1] Bayesian "*priors*" were developed from the network flow models for these regions to estimate "*posterior distributions*" for probabilistic-based prediction of network flow for the emerging bioeconomy. A macroscopic analysis and micro simulation, based on two performance measures, were used to solve the capacity problem. The two objectives of the model were minimizing all travel times for all delivery movement while maximizing number of deliveries to the biofuel plant. A Synchro/SimTraffic model for the vicinity of Vonore TN was developed to model a 'real world' scenario; road network and signal timing was completed for Vonore TN assuming different scale capacities. The model for the Vonore, TN biofuel plant was used for the Bayesian logistic modeling for the southeastern U.S.

This study also involved organizing large volumes of data collected from various sources for the Bayesian logistic modeling (*e.g.*, U.S. Census Bureau, U.S. Forest Service, U.S. National Land Cover Database, U.S. National Elevation Dataset, U.S. Department of Agriculture National Agricultural Statistic Service, U.S. Environmental Protection Agency, and state mill directories).

The BioSAT model (Perdue *et al.* 2011) was used in this study to estimate the availability of woody biomass for procurement zones within a 128.8 km (80 mile) one

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

way travel distance which was not always concentric, *i.e.*, the shape of such zones rely on the available transportation network and location of biomass supply in the surrounding area. National forests, parks, urban areas, and other restricted areas were not considered in the study. All records were organized at the U.S. Census Bureau 5-digit ZIP Code Tabulation Area (ZCTA) level. There were 10,016 ZCTAs in the 13-state 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.

This study applied Bayesian inference for estimation of the parameters in the logistic regression models. 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 (Hilborn *et al.* 1994). The statistical software package WinBUGS® 3.0.2 ("Windows Bayesian Inference Using Gibbs Sampling") was used for Bayesian inference analysis.

Bioenergy and biofuel plants were defined as facilities that integrate woody biomass conversion processes, and equipment to produce wood pellets for energy, biofuels, biopower, or value-added biochemicals (National Renewable Energy Laboratory 2009). Only sixty 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 in the study area were used as surrogates (*e.g.,* sawmills, OSB mills, and pulp and paper mills). The mills were segregated into two groups: smaller mills such as sawmills (Group I) and larger mills such as pulp and paper mills, OSB mills and wood pellets mills (Group II).

"Median Family Income", "Timberland Annual Growth-to-Removal Ratio", and "Transportation Delays" were highly significant in influencing mill location (p-values <0.0001 for both mill groups). Other statistically significant variables to a lesser extent for both groups were "Urban Land Area Ratio" and "Water Area Ratio." A higher family income, and larger urban area had negative coefficients which suggested that urban areas were not suitable for siting mills.

Four ordinal levels for ranking were used to estimate the probability from the logistic model. The higher probability locations for Group I mills were clustered in the southern Alabama, southern Georgia, southeast Mississippi, southern Virginia, west Louisiana, west Arkansas, and east Texas regions. The higher probability locations for Group II mills were clustered in the southeast Alabama, south Georgia, central North Carolina, and the Mississippi Delta regions.

1. Introduction

As noted in BRDI (2008) the national fuel transportation and storage infrastructure must accommodate the current and future growth of domestic biofuels production and transportation. However, investment in the transportation system is not keeping pace with the expected growth in freight transportation in the coming decades. This is leading to rapid increases in congestion and delays along freight corridors (Council of State Governors 2010). Highway bottlenecks impacting freight delay large numbers of truck shipments and is problematic to the emerging bioeconomy. "*Truck-idling*" consumes nearly one billion gallons of diesel fuel annually and produces 11 million tons of CO_2 (Council of State Governors 2010). Trucks are still the only available method for delivering freight for 80% of American companies given the decades of freight rail industry consolidation and the abandonment of rail lines.

This project addresses Southeastern Sun Grant priority "*Impact of the emerging bioeconomy on transportation infrastructure and related concerns*." We modeled the impact of the emerging bioeconomy on truck transportation network flow. Regions with major truck freight were modeled for increased flow and contrasted with regions without comparable flow rates. Bayesian "*priors*" were developed from the network flow models for these regions to estimate "*posterior distributions*" for probabilistic-based prediction (Guess 1996).

1.1 Objectives and Scope

There were five study objectives: (1) Identify regions in the southern U.S. that have high potential for emerging bioeconomies in the presence of high transportation

flow; (2) Assess the current transportation flow for these regions and model the impact of increased truck transportation flow for select types of bioenergy feedstocks and end products, if applicable, *e.g., biofuels, etc.*; (3) Estimate the transportation costs in the select regions and compare such costs with other potential bioeconomy regions that do not have transportation flow bottlenecks; (4) Develop Bayesian "*prior*" and "*posterior*" distributions for transportation flow times for the above regions and biomass types; and (5) Incorporate the aforementioned objectives in optimal siting locations for biorefineries using Bayesian spatial logistic regression models developed by Huang *et al.* (2011), Young *et al*. (2011). The scope of the proposed study was the 13 southern states.[2]

2. Methodology

2.1 Transportation Network Simulation Flow Model

2.1.1 Overview

The rural versus urban location of biofuel plants was examined to determine at what volume-to-capacity ratio a road network should be omitted from the location selection. The computation experiments were implemented using simulation software ARENA 13.5 and OptQuest. A macroscopic analysis and micro simulation, based on two performance measures, was used to solve the capacity problem. The two objectives of the model were minimizing all travel times for all delivery movement while maximizing number of deliveries to the biofuel plant.

The time horizon was set to a three hour peak period and the plant's natural processing capacity is in terms of the number of deliveries processed within a given time

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

interval that is set to be five minutes. The various travel times, vehicle speeds, and traffic signal timing were all estimated statistically based on general transportation data collected. Various assumptions were made to control the size of the problem, namely:

1. Biofuel plant requires a delivery of renewable materials every five minutes to continue operation;
2. One traffic signal was simulated with a cycle length of 100 seconds (60 seconds for the major road and 40 seconds for the minor road);
3. The road network consisted of high-quality arterials and the maximum capacity of all arterials was 1500 vehicles per lane per hour.

To measure the effect of the existing traffic volume the volume to capacity (v/c) ratio was the constraint. The v/c ratio was calculated by dividing the existing traffic volume by the maximum capacity of the road (in this case 1500 vehicle per lane per hour). In Table 1, a description of driving characteristics and Levels of Service (LOS) are listed with correlating v/c ratios. The existing traffic volumes with the resulting arrival times (sec/veh), and v/c ratios are shown in Table 2. General traffic volumes experience two peaks, AM and PM, during a weekday. These peaks range in lengths of one to three hours depending on location and capacity constraints. Each simulation is designed to reproduce a three hour long peak hour.

Using the Arena Process Analyzer, each arrival rate from Table 2 was replicated 60 times as a scenario to determine the average truck travel time and number of deliveries to the biofuel plant. The ranges in arrival times appeared to have a minimal effect on the number of deliveries to the biofuel plant. The maximum average of deliveries was 32.78, while the overall average was 32.62. The effect on travel time appeared to have the same

Table 1. Level of service criteria for arterials based on volume-to-capacity ratios.

Level of Service	Description	V/C[b]
A	Free-flow conditions with unimpeded maneuverability. Stopped delay at signalized intersection is minimal.	0.00 to 0.60
B	Reasonably unimpeded operations with slightly restricted maneuverability. Stopped delays are not bothersome.	0.61 to 0.70
C	Stable operations with somewhat more restrictions in making mid-block lane changes than LOS B. Motorists will experience appreciable tension while driving.	0.71 to 0.80
D	Approaching unstable operations where small increases in volume produce substantial increases in delay and decreases in speed.	0.81 to 0.90
E	Operations with significant intersection approach delays and low average speeds.	0.91 to 1.00
F	Operations with extremely low speeds caused by intersection congestion, high delay, and adverse signal progression.	Greater Than 1.00

[a] For arterials that are multilane divided or undivided with some parking, a signalized intersection density of four to eight per mile, and moderate roadside development.
[b] Volume-to-capacity ratio.

≥ greater than or equal to.
< less than.

Source: Transportation Research Board, *Highway Capacity Manual, Special Report 209* (Washington, D.C., 1994).

Table 2. V/C ratio, arrival time, and LOS for existing traffic.

EXISTING VOLUME (vplph)	ARTERIAL CAPACITY (vplph)	v/c RATIO	ARRIVAL TIME (sec/veh)	LOS
600	1500	0.40	6.00	A
800	1500	0.53	4.50	
900	1500	0.60	4.00	
1000	1500	0.67	3.60	B
1100	1500	0.73	3.27	C
1200	1500	0.80	3.00	
1300	1500	0.87	2.77	D
1400	1500	0.93	2.57	E
1500	1500	1.00	2.40	
1600	1500	1.07	2.25	F
1800	1500	1.20	2.00	

effect, with a maximum average of 523.018 and an overall average of 518.179 seconds; giving a difference of only 4.839 seconds.

It was important to establish whether there was a limit for the v/c ratio for a road network when plant locations were considered. ARENA's OptQuest was utilized to determine the optimal truck travel time and the number of deliveries and the corresponding v/c ratio. The optimal travel time was 479.33 seconds and was caused by a v/c ratio of 0.64. This was a reduction of 38.84 seconds from the overall average determined using the Process Analyzer. The reason for showing the best 25 optimizations versus showing the single best is to point out the varying v/c ratios used to achieve the optimizations. The third best optimization was actually the result of a 0.99 v/c ratio, which is a difference of 0.35 from our optimal travel time's v/c ratio. The same occurs with the number of deliveries to the biofuel plant, with an optimal number of 34.33 deliveries achieved with a v/c ratio of .61. This was an increase of 1.71 deliveries from the Process Analyzers overall average.

The average v/c ratios for the top 25 OptQuests travel time and number of deliveries were .57 and .63, respectively. With the inability to run an optimization for the v/c ratio, the results indicated that with optimization of travel time and number of deliveries the best LOS to have on a road appears to be scenarios A and B (Table 2). Though with the Process Analyzer the most successful v/c ratio was 0.93. The location selection of a biofuel plant was not as simple as determining the ideal v/c ratio for the surrounding road network. A detailed road network simulation was required to model the true effects of freight bottlenecking. Though our model was detailed, the roadway

network was simple with only three routes and two intersections based on the Vonore, TN biofuels plant.

The team had multiple meetings with Dr. Samuel Jackson of Genera Energy, LLC and at the Vonore, TN biofuel plant and acquired operational information and traffic volumes for existing research and projected commercial facilities. Farm-routing information was also acquired and imported into Google Earth for projecting a 'real-world' traffic assignment of Switchgrass bio-products. The team followed several hay trucks in the Vonore area to study their routes and potential height, weight, and width concerns that need to be considered in the traffic modeling effort. The team compared traffic modeling tools including TransCAD, Synchro/SimTraffic, and MATSim and decided to use Synchro for the modeling task. The team met with US DOT/RITA official and presented progress. The team developed a Synchro/SimTraffic model for the vicinity of Vonore. Road network and signal timing was completed.

The team built the transportation simulation framework with a schematic roadway network over aerial photos with proper traffic control devices. With *SimTraffic*, we simulated the traffic operations in the vicinity of the Vonore biofuel plant. Because the study site sits between a couple of counties, the traffic volumes (AADT) were not easily extracted. Nevertheless, the baseline data were acquired and used for the simulation with a *K* factor of 10% and a *D* factor of 0.5. The remaining task was to use realistic increases of traffic resultant from a full size biofuel plant for scenario simulations. Three screen captures of the simulation in action are given in Figures 1, 2 and 3.

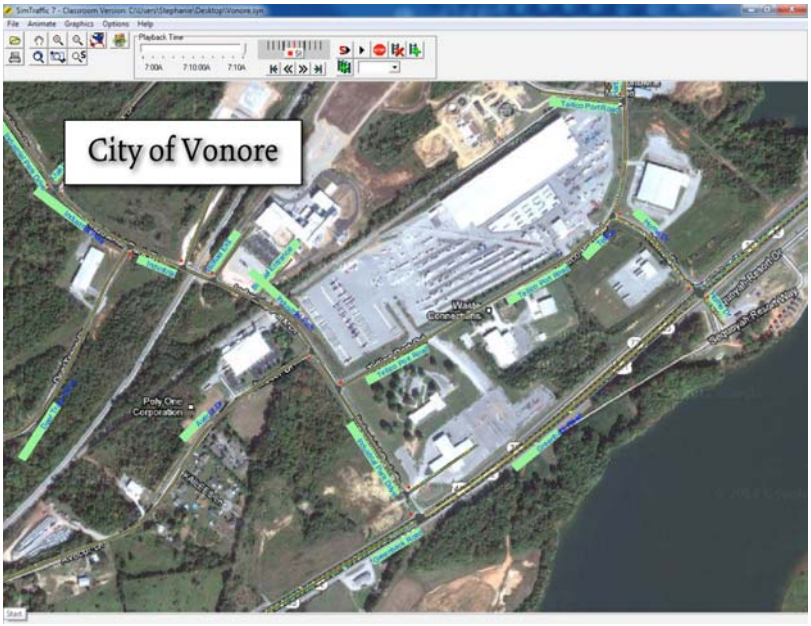


Figure 1. City of Vonore TN network model.

Figure 2. Simulation of a signal controlled intersection.

Figure 3. Simulation of two STOP controlled intersections.

2.2 Logistic Regression Model for Siting Biorefineries

2.2.1 Dataset

This study involved organizing large volumes of data collected from various sources, including the U.S. Census Bureau, U.S. Forest Service, U.S. National Land Cover Database, U.S. National Elevation Dataset, U.S. Department of Agriculture National Agricultural Statistic Service, U.S. Environmental Protection Agency, and state mill directories.

Another resource that was used in the study was the integration of the BioSAT model (Perdue *et al.* 2011). The BioSAT model was used to estimate the availability of woody biomass for procurement zones within a 128.8 km (80 mile) one way travel distance which in most cases was not 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. 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 (*e.g.,* rural paved two-lane roads, privately owned driveway, pedestrian walkway). Navteq data were used for major metropolitan areas (*e.g.,* roads with turn restrictions, physical barriers and gates, one-way streets, restricted access and relative road heights). 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.

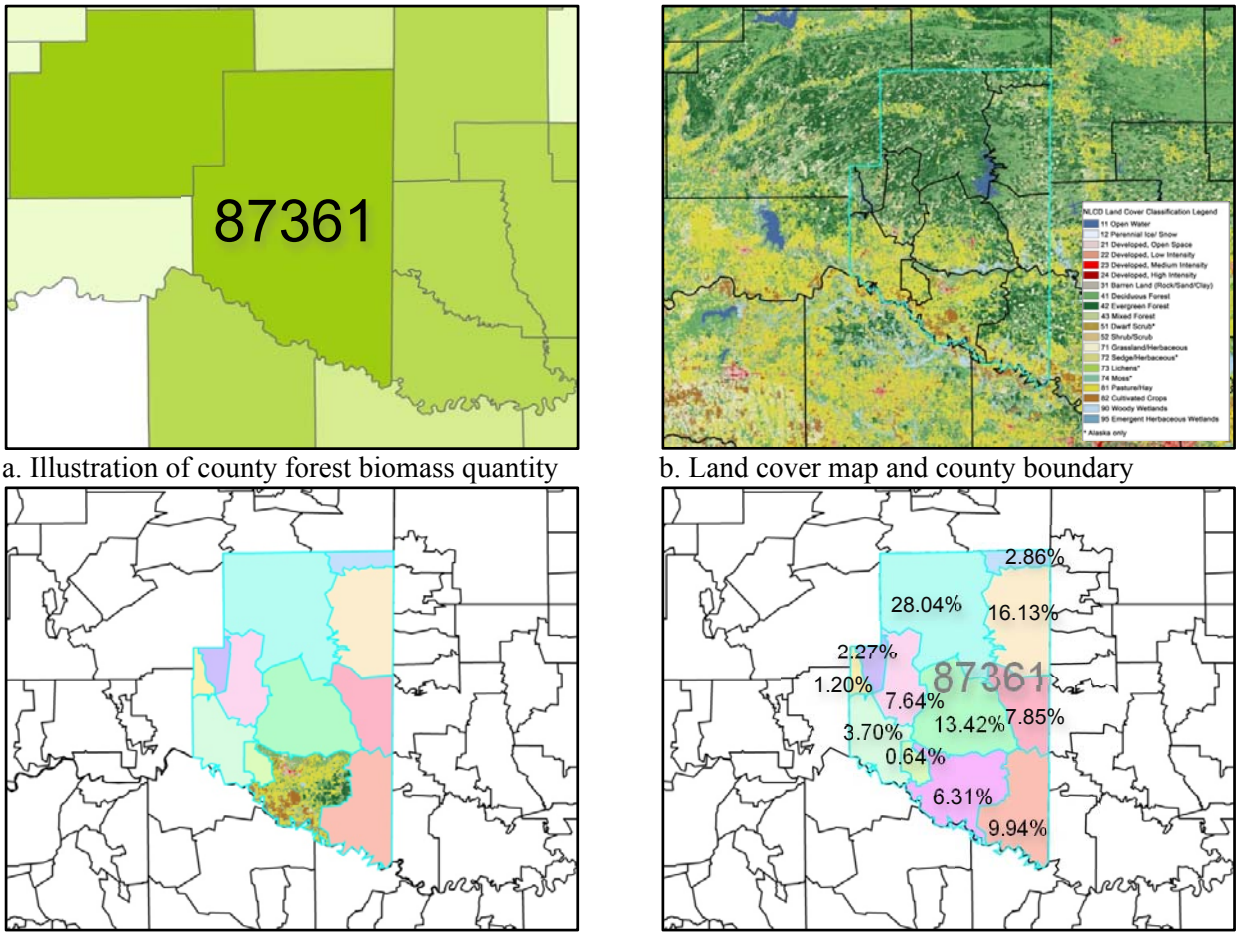
All records were organized at the U.S. Census Bureau 5-digit ZIP Code Tabulation Area (ZCTA) level. There were 10,016 ZCTAs in the 13-state[3] study region which corresponded to 10,016 potential analytical polygons or potential sites for woody biomass-using facilities. The average area size for 5-digit ZCTAs in the 13-state study regions was 209.84 km^2.

2.2.1.1 Biomass Estimation using GIS -- Forest biomass annual growth and removal quantity data were collected at the county level from Forest Inventory and Analysis Database (FIADB) version 3.0 (Figure 4a), and reallocation was done for each of the 10,016, 5-digit ZCTAs using a geographic information system (GIS) technology. National land cover data and digital raster map 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 4b). Forest biomass annual growth and removal

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

quantities were proportionally allocated to each 5-digit ZCTA using the county boundary, 5-digit ZCTA, and the land cover image data with GIS spatial overlay techniques.

Due to the mismatch of county boundary and 5-digit ZCTA (*i.e.,* some 5-digit ZCTAs cross county), each forest biomass county was split into multiple area parts via the 5-digit ZCTA area shape 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 were estimated (Figure 4c). By summing up the pixels of deciduous forest, evergreen forest and mixed forest, which together represented forestland, in the unit of county, a forestland pixel ratio for each area part to its belonging county was calculated and the forest biomass quantity in every area part was derived for this pixel ratio (Figure 4d). A summed quantity value for all area parts belonging to the same 5-digit ZCTA were then calculated as the forest biomass quantity in this 5-digit ZCTA.



a. Illustration of county forest biomass quantity

b. Land cover map and county boundary

c. Land cover for 5-digit ZCTA boundary

d. Forest biomass allocation by 5-digit ZCTA

Figure 4. Illustration of forest biomass allocation at the level of 5-digit ZCTA.

2.2.2 Logistic Regression

The logistic regression methodology was inspired by Young *et al.* (2011). In Young *et al.* (2011) logistic regression was used to identify site locations for biorefineries 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 (see equations [1] and [2]). 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 (Hilborn *et al.* 1994). In practice, Bayesian inference expresses the uncertainties of parameters in terms of probability distributions called prior probability distributions and derives posterior estimate probability distributions from the priors. For example, by expressing the uncertainties in parameter vector β for a given model M as the posterior probability distribution $p(\beta|M, D)$, where D is the observed data, we have,

$$p(y = 1|x, M, D) = \int_{\beta}^{a} p(y = 1, \beta\,|x, M, D)d\beta = \int_{\beta}^{a} p(y = 1|x, \beta, M)p(\beta|M, D)d\,\beta, \qquad [1]$$

where, $$p(y = 1\,|x, \beta, M) = \frac{1}{1+\exp(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}\ . \qquad [2]$$

The key of Bayesian inference is to choose the parametric family for the prior probability distributions. Two categories were used: non-informative prior distributions and informative prior distributions. A non-informative prior distribution expresses vague or general information about a parameter. A common non-informative prior distribution is the uniform distribution, which assigns equal probabilities to all possible values, and

always yields similar results as classical statistics. Thus, Bayesian and classical statistics are not exclusive, yet overlapped to some extent. In fact, classical approaches are approximate Bayesian using certain priors. An informative prior distribution reflects specific and definite information about a parameter. If both prior and posterior distributions belong to the same family, the prior distribution is called a conjugate prior distribution, which is a case in informative prior distributions. In this study two prior distributions were selected from non-informative and informative prior distributions and were constructed on parameter β:

Prior 1: Uniform prior distribution $p(\beta) \propto \text{constant}$, [3]

Prior 2: Gaussian prior distribution $(\beta \mid \mu, \sigma^2) \propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\beta-\mu)^2}{2\sigma^2})$. [4]

The statistical software package WinBUGS® 3.0.2 ("Windows Bayesian Inference Using Gibbs Sampling") was used for Bayesian inference analysis. It provided a convenient environment to conduct a Markov Chain Monte Carlo simulation (MCMC) of a random walk in the space of parameters β , which converges to a stationary distribution approximating the joint distribution. In each analysis, one independent chain was run for 10,000 iterations. Convergence was assessed by visual inspection of the chain and by the Gelman *et al.* (2000) shrink factor. In addition to providing a posterior estimate of the mean $\hat{\beta}$, the stationary distribution also provided a posterior standard deviation, the median and a 95% credibility interval (CI)[4] of $\hat{\beta}$ via the estimated

[4]A powerful tool that provides a measure of the uncertainty of classification performance based on the current available information is the Bayesian CI. To obtain such an interval, a binary classifier is designed using N_d samples per class, and tested using N_t class-specific holdout test

covariance for β (Jaynes 1976). The data 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 by constructing classification tables.

2.2.3 Study Group

Bioenergy and biofuel plants were defined as facilities that integrate woody biomass conversion processes, and equipment to produce wood pellets for energy, biofuels, biopower, or value-added biochemicals (National Renewable Energy Laboratory 2009). 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 (Patari 2010, Knight 2009, Stewart 2009, Cohen *et al.* 2010, Moon *et al.* 2008). The groups were:

Group I: Sawmills;

Group II: Pulp and paper mills, OSB mills and wood pellets mills.

Groups I locations are illustrated in Figure 5. Group II mills were illustrated in Figure 6.

---

examples that results in k_t errors. Bayes' theorem is used to transform the test result into the posterior (*a posteriori*) pdf. Based on the posterior pdf, an α% Bayesian CI is defined as any interval of the posterior that covers α% of its probability mass (Gustafsson *et al.* 2010).

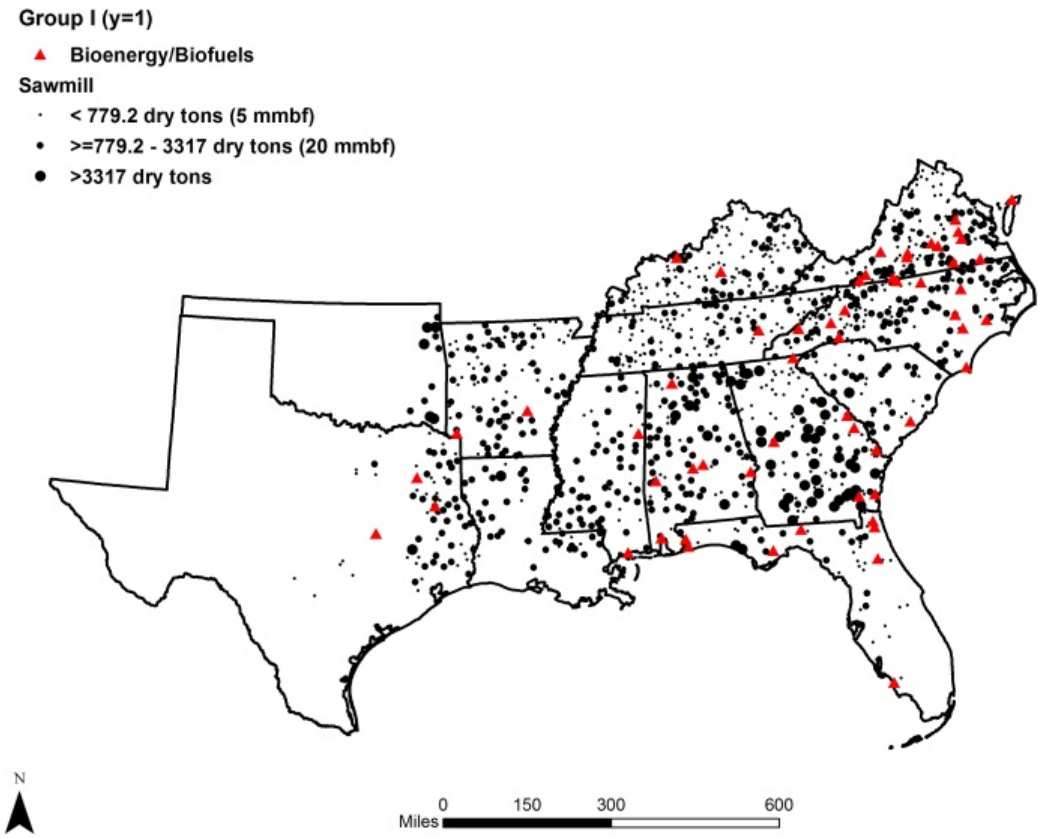


Figure 5. Illustration of Group I woody biomass-using mills.

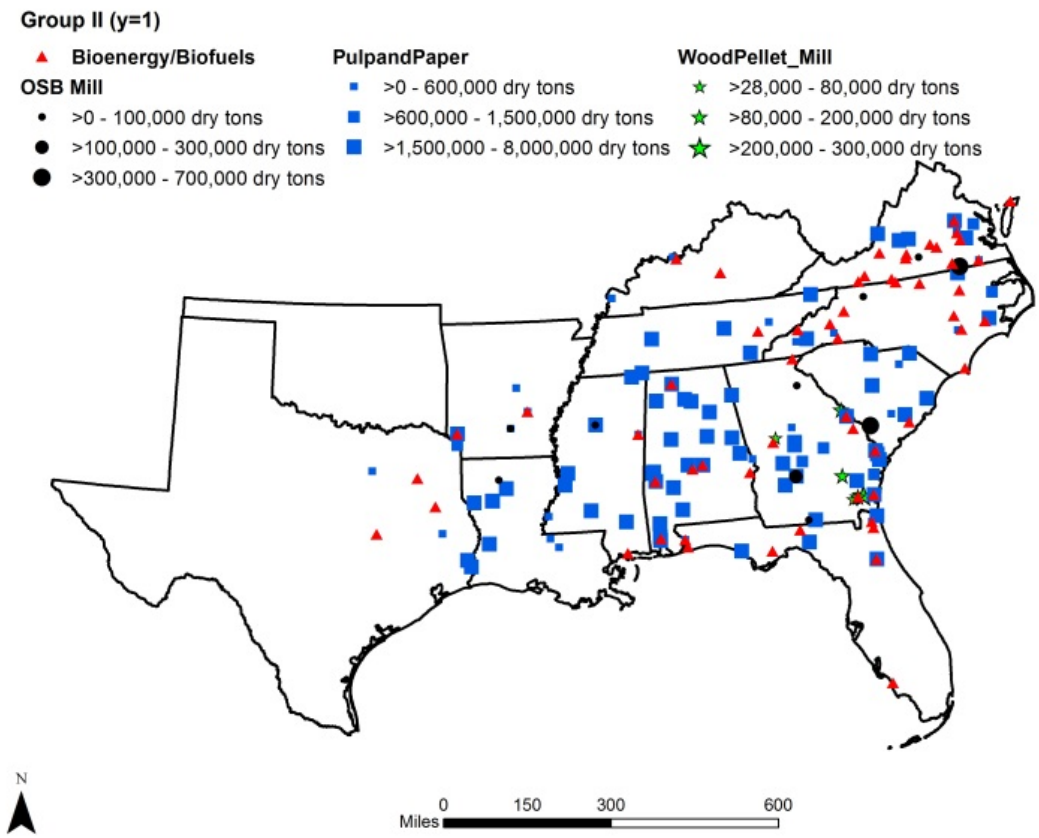


Figure 6. Illustration of Group II woody biomass-using mills.

2.2.4 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 one if *ith* ZCTA had at least one woody biomass-using facility, and y_{i2} equals one was defined similarly for Group II mills. Thirteen explanatory variables for data available in the public domain were examined in the Bayesian logistic models (Table 3).

Table 3. 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			U.S. National Land Cover Database (2006)
Water Area Ratio	5-digit ZCTA	percent	
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
Transportation Delays	5-digit ZCTA	seconds	Average traffic total delays within in a 10-mile distance using the transportation network simulation flow model

* No timberland growth value available in the west Oklahoma and Texas

** As an independent explanatory variable only in Group II subset.

2.2.5 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. The probability was used in the validation data set to compare ZCTAs with actual mill locations.

3. Results and Discussion

3.1 Group I

Five out of the possible fourteen predictor variables were statistically significant (p-value < 0.05) from the stepwise logistic regression (Table 4). To compare the MLE and Bayesian inference estimation methods for parameter coefficients, the classification tables as well as ROC and lift charts for training and validation data sets were displayed in Tables 5 and 6, and Figures 7 and 8. The classification tables confirmed that the logistic regression with Bayesian Inference assuming a uniform prior had good predictive power for the siting locations of the Group I woody biomass-using facilities (Tables 5 and 6). Graphically, ROC and lift charts confirmed that Bayesian inference assuming a uniform prior had a better overall prediction performance since the ROC curve was closer to the left-corner optimal point and the lift value was higher across the training and validation data sets (Figures 7 and 8). The sensitivity of this model using Bayesian inference assuming a uniform prior in the validation data set was 89% (*e.g.*, the model predict a mill location correctly 89% of the time in validation), and specificity (*e.g.*, predicting no mill location where no mill by 2010) was 92.3% (Tables 5 and 6). The sensitivity rates in training and validation data sets were acceptable, which was higher than 75% based on the stringent criteria of medical radiology screening (Carney *et al.* 2010). It indicated a good performance of the model that had a significant power for correctly predicting a mill location. However, the specificity rates of the model in the training and validation data sets were lower than the points[5] used by the medical industry

[5] 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%, recall false positive rate (Type I error) less than 5% or greater than 12%.

but close to acceptable. This logistic model for Group I appeared better at predicting preferred site locations than at predicting where not to location a plant.

Table 4. Significant variables for Group I mills.

Significant variables	**p-value**
Median Family Income	<0.0001
Urban Land Area Ratio	<0.0012
Water Area Ratio	<0.0010
Timberland Annual Growth-to-Removal Ratio	<0.0001
Transportation Delays	<0.0001

Table 5. 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)		3136	202	90	820	93.9%	90.1%
Bayesian Inference	*Uniform*	3140	198	86	824	94.1%	90.5%
	Gaussian	3136	200	88	822	93.9%	90.3%

Table 6. 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)		769	65	27	200	92.2%	88.1%
Bayesian Inference	*Uniform*	770	64	25	202	92.3%	89.0%
	Gaussian	769	64	27	200	92.2%	88.1%

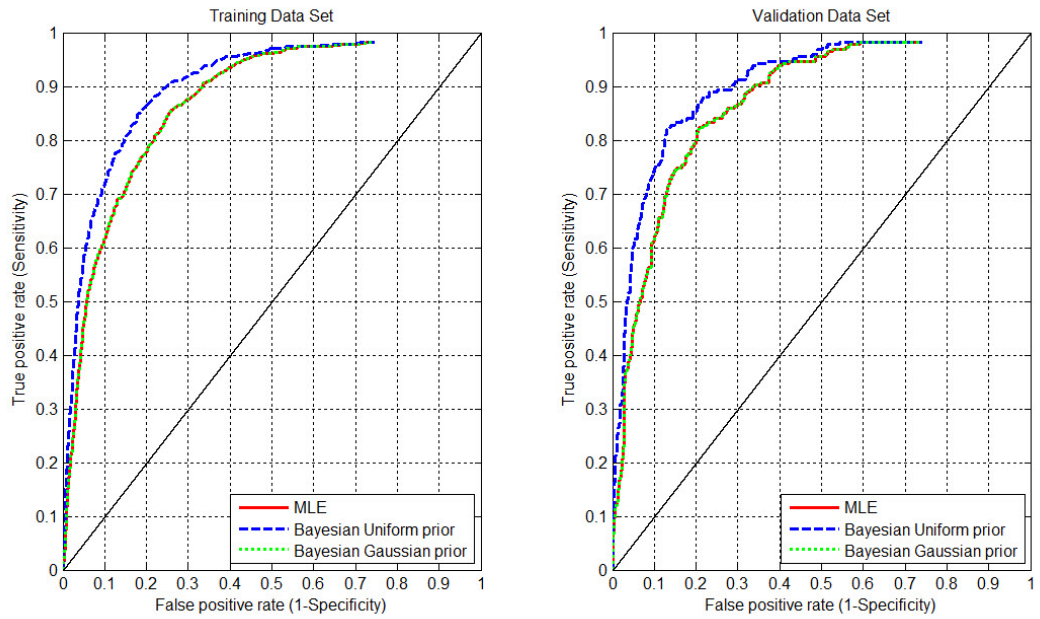


Figure 7. ROC charts for model assessment of Group I woody biomass-using mills.

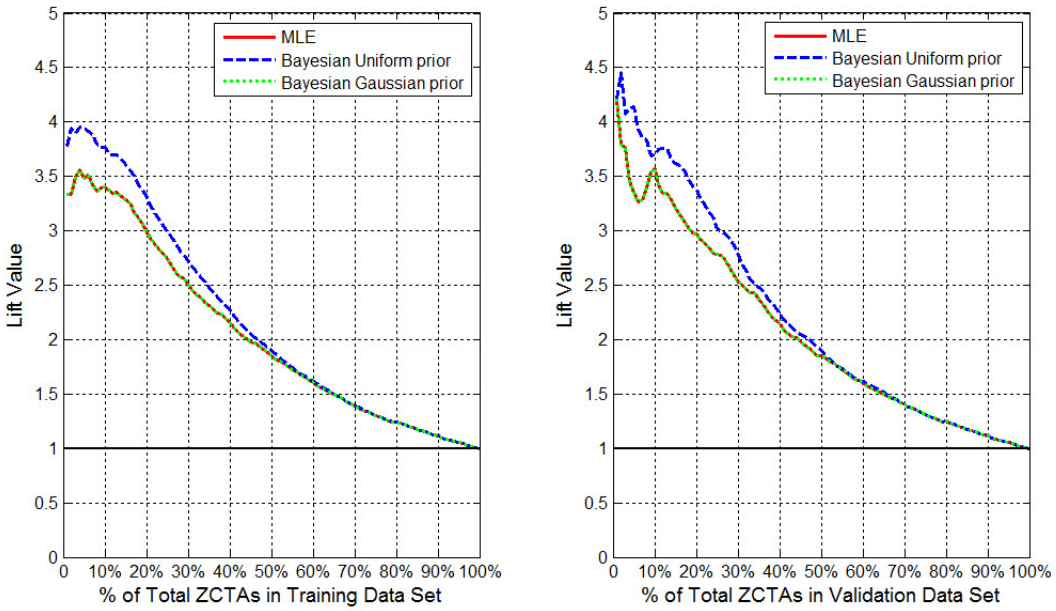


Figure 8. Lift Charts for model assessment of Group I woody biomass-using mills.

"Median Family Income," "Timberland Annual Growth-to-Removal Ratio," and "Transportation Delays" were highly significant in influencing mill location (p-values <0.000, Table 7). Other statistically significant variables were "Urban Land Area Ratio," and "Water Area Ratio." A higher family income and larger urban area had negative coefficients (Table 7), which suggested that urban developed areas were not suitable for siting mills. This result is in agreement with other studies that mill locations were nearer the rural biomass supply. "Timberland Annual Growth-to-Removal Ratio" and "Water Area Ratio" had positive coefficients. This indicates that landscape with abundant forestland and water areas are preferred. "Transportation Delays" had a positive coefficient, which shows the mill location may have significant impacts on the local transportation networks. This result suggests the importance of landscape suitability and woody biomass availability on mill location as well as mill location influence on the adjoining transportation system.

Table 7. Analysis of variable estimates from Bayesian inference with a uniform prior for Group I mills.

Significant Variables	**Mean Estimates**
Median Family Income	-0.3080
Urban Land Area Ratio	-1.3204
Water Area Ratio	0.7580
Timberland Annual Growth-to-Removal Ratio	3.7814
Transportation Delays	6.3087

Four ordinal levels for ranking the estimated probability from the logistic model (Bayesian Inference with a uniform prior) in the study region are given in Figure 9. The higher probability locations for Group I mills were clustered in the southern Alabama, southern Georgia, southeast Mississippi, southern Virginia, western Louisiana, western Arkansas, and eastern Texas regions.

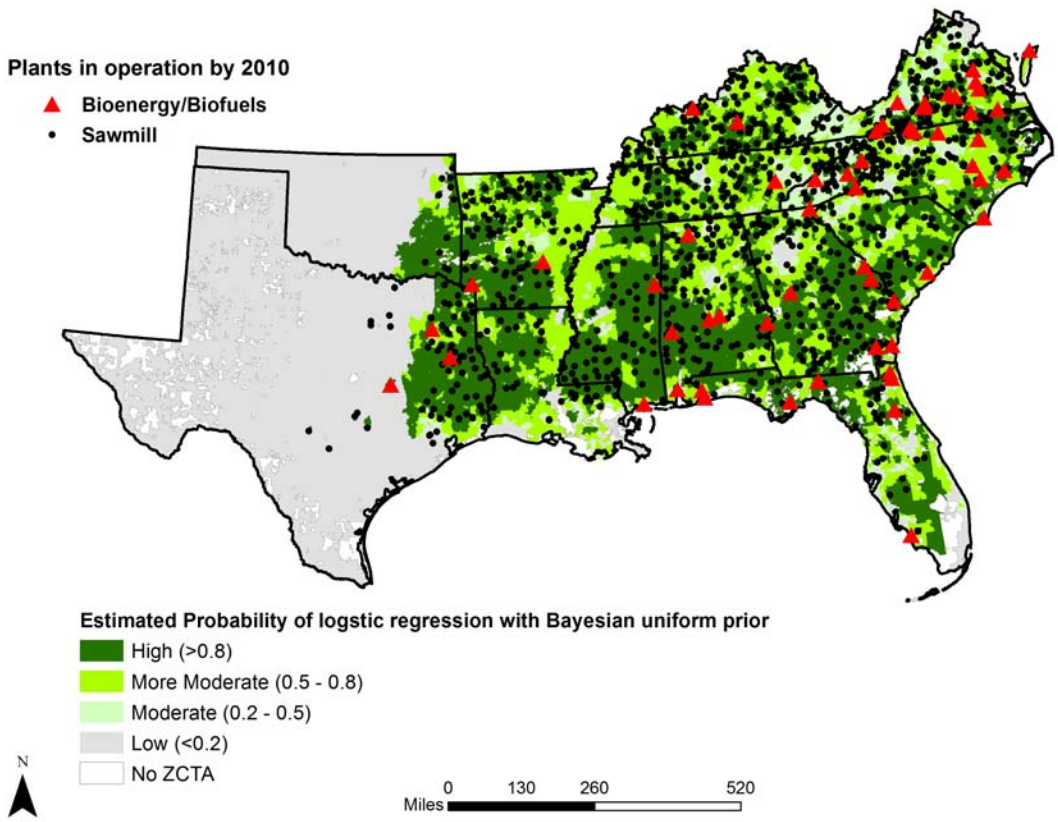


Figure 9. Estimated probability locations for Group I.

3.2 Group II

Five out of fifteen predictor variables were statistically significant (p-values <0.05) from the stepwise logistic regression (Table 8). The classification tables as well as ROC and Lift Charts for training and validation data sets were constructed to compare the MLE and Bayesian inference estimation methods for parameters (Tables 9 and 10, Figures 10 and 11). The classification tables confirmed that the logistic regression with Bayesian inference assuming a uniform prior had good predictive power for the siting locations of the Group II woody biomass-using facilities (Tables 9 and 10). ROC and Lift Charts graphically suggested the predictive performance of three methods were close but Bayesian inference with a uniform prior had a better overall prediction of siting Group II mills for validation data set. The classification table for the model using Bayesian Inference with a uniform prior had a sensitivity of 88.1% and specificity of 89.2% for Group II training data (Table 8). The sensitivity of this model in the validation data set was 90.5% and specificity was 87.7% (Table 9). The sensitivity rates of the model for the training and validation data sets were acceptable. The specificity rates of the model in the training and validation data sets were very close to acceptable. The logistic model using Bayesian Inference with a uniform prior for Group II was a good predictive model for predicting preferred site locations and identifying non-preferred locations.

Table 8. Significant variables for Group II mills.

Significant variables	**p-value**
Median Family Income	<0.0001
Urban Land Area Ratio	0.0001
Water Area Ratio	<0.03
Number of Primary Wood Processing Mills in Each ZCTA	<0.0001
Transportation Delays	<0.0001

Table 9. 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)		519	66	12	72	88.7%	85.7%
Bayesian Inference	*Uniform*	522	63	10	74	89.2%	88.1%
	Gaussian	519	66	12	72	88.7%	85.7%

Table 10. 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)		125	21	3	18	85.6%	85.7%
Bayesian Inference	*Uniform*	128	18	2	19	87.7%	90.5%
	Gaussian	125	21	3	18	85.6%	85.7%

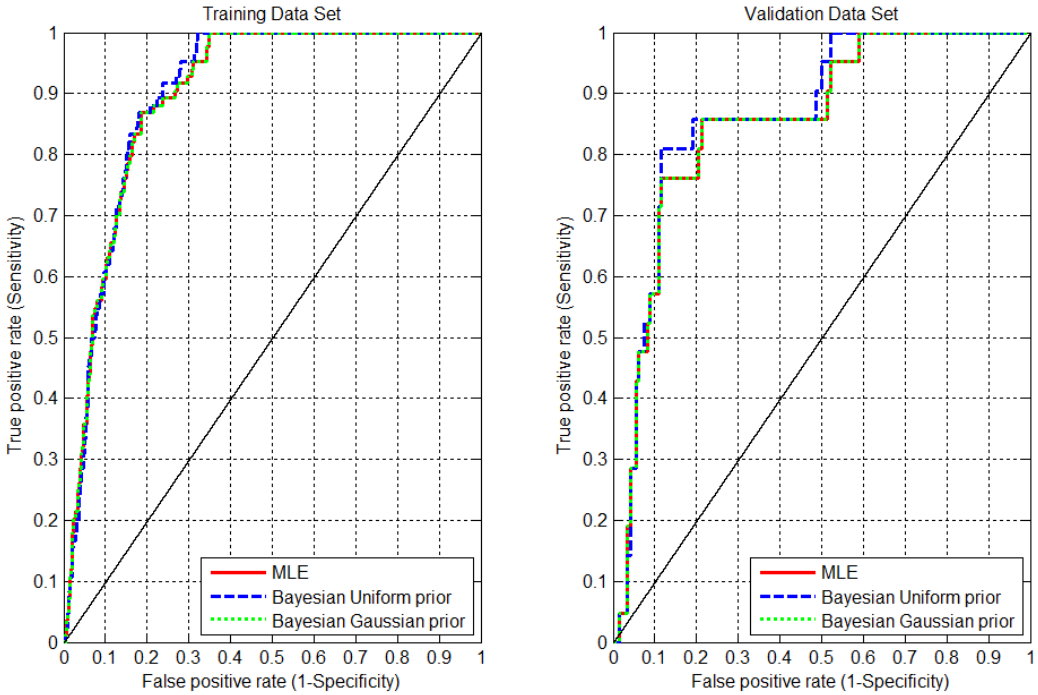


Figure 10. ROC charts for model assessment of Group II woody biomass-using mills.

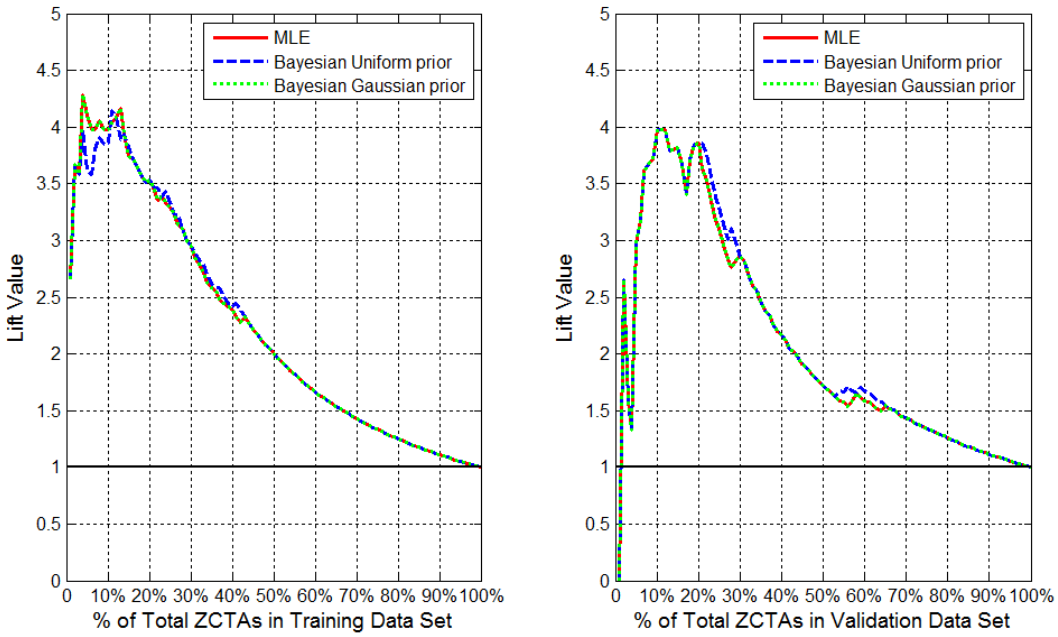


Figure 11. Lift Charts for model assessment of Group II woody biomass-using mills.

"Median Family Income," "Urban Land Area Ratio," "Number of Primary Wood Processing Mills in Each ZCTA," and "Transportation Delays" were highly significant influencing mill location, p-values < 0.0001 (Table 10). Another statistically significant variable was "Water Area Ratio." "Median Family Income" had negative influences on mill location for Group II. "Urban Land Area Ratio," "Water Area Ratio," "Number of Primary Wood Processing Mills in Each ZCTA," and "Transportation Delays" had positive influences on mill location. This may indirectly reflect that this group is located in an area that has better transportation road network which may be somewhat closer to the urban area relative to Group I because of the synergistic relationship that exists between primary wood processing mills and larger-size Group II facilities which may depend on the residue feedstock from primary wood processing mills. It also confirms the local transportation system impacts from Group II facilities.

Four ordinal levels for ranking the estimated probability from the logistic model (Bayesian Inference with a uniform prior) in the study region are given in Figure 12. The higher probability locations for Group II mills are clustered in the southeast Alabama, southern Georgia, central North Carolina, and Mississippi Delta regions.

Table 11. Analysis of significant variables estimates from Bayesian inference with a uniform prior for Group II mills.

Significant Variables	**Mean Estimates**
Median Family Income	-3.388
Urban Land Area Ratio	2.343
Water Area Ratio	1.344
Number of Primary Wood Processing Mills in Each ZCTA	1.814
Transportation Delays	2.597

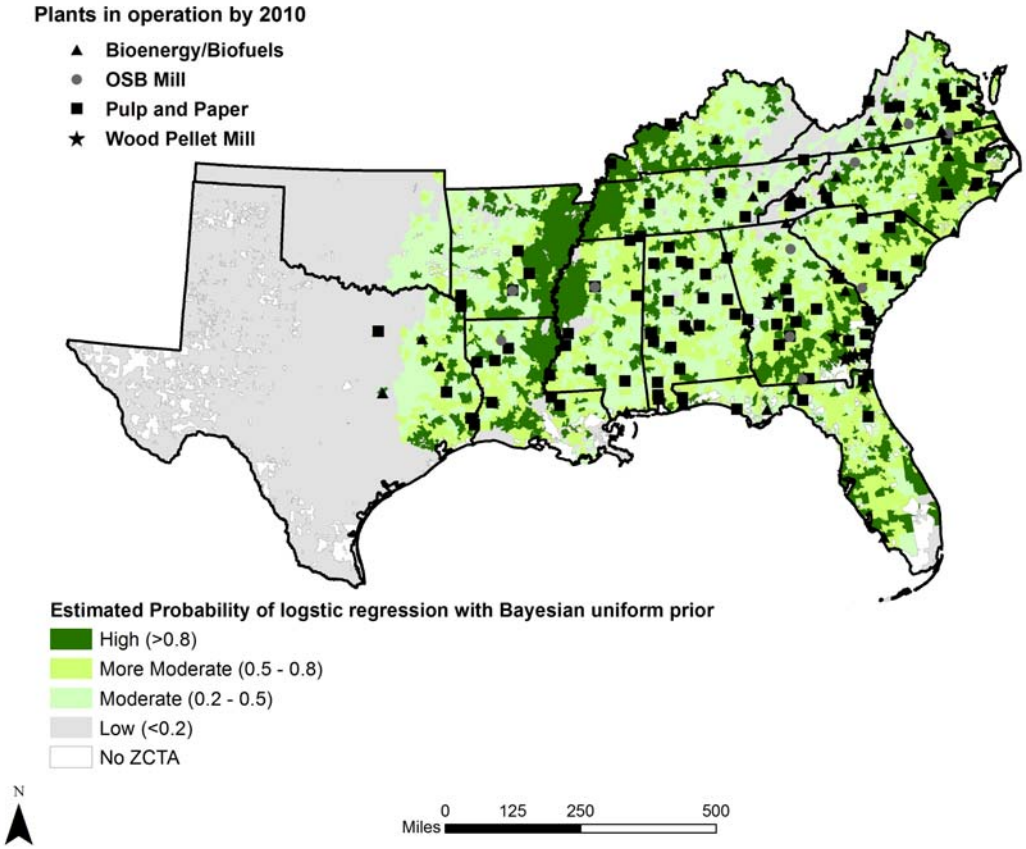


Figure 12. Estimated probability locations for Group II.

3.3 Trucking Costs and Delay Times

Delay times for each state were estimated from the study using the Synchro and SimTraffic models. Average trucking costs ($/dry ton) were estimated using the trucking cost model of BioSAT, www.biosat.net. Descriptive statistics of the trucking costs from the BioSAT model are given in Table 12. Average trucking costs varied from $15.52/dry-ton to $17.34/dry-ton across the 13 southern states study region. The coefficient of variation (CV) for these trucking costs varied from 5.09% in South Carolina to 30.59% in Florida and Tennessee.

Table 12. Descriptive statistics of trucking costs ($/dry ton) by state as estimated from the BioSAT model.

State	Average ($\bar{x}$)	Median ($\tilde{\mu}$)	Standard Deviation (s)	Coefficient of Variation (CV)
Alabama	15.74	16.00	1.25	7.92%
Arkansas	16.27	17.05	4.57	28.10%
Florida	15.88	15.25	4.86	30.59%
Georgia	16.00	16.40	1.54	9.65%
Kentucky	16.35	16.38	0.88	5.37%
Louisiana	16.12	16.31	1.39	8.59%
Mississippi	16.11	16.24	0.96	5.98%
North Carolina	16.32	16.33	1.10	6.74%
Oklahoma	17.34	17.03	1.77	10.19%
South Carolina	15.91	15.90	0.81	5.09%
Tennessee	15.52	16.36	4.75	30.59%
Texas	16.34	17.13	3.62	22.13%
Virginia	15.83	15.98	1.15	7.27%

The costs associated with transportation delay times for trucks were derived from Gillett (2011). Average delay minutes per state varied from 1.12 minutes in Oklahoma to 17.93 minutes in Tennessee (Table 13). Delay times had significant influences on trucking costs by state. Delay times have the potential to increase trucking costs by as much as 61% in certain states within the study region. Recall the statistical significance

of transportation delays in the Bayesian logistic regression models determining site location.

Table 13. Statistical interval of trucking costs ($/dry ton) by state adjusted for average delay time per state.

State	Average Cost ($/ton)	Average Delay (minutes)	Delay Standard Deviation (minutes)	Statistical Interval* cost with delay times ($/ton)
Alabama	15.74	15.11	36.86	[12.26 , 17.96]
Arkansas	16.27	16.80	43.75	[13.27 , 20.32]
Florida	15.88	13.92	17.70	[12.81 , 15.03]
Georgia	16.00	17.68	42.92	[14.59 , 20.78]
Kentucky	16.35	16.62	47.70	[13.25 , 19.99]
Louisiana	16.12	5.85	21.56	[10.98 , 21.26]
Mississippi	16.11	12.55	34.03	[11.96 , 20.26]
North Carolina	16.32	15.13	28.08	[13.49 , 19.15]
Oklahoma	17.34	1.12	8.96	[6.64, 28.05]
South Carolina	15.91	8.91	29.31	[10.93 , 20.89]
Tennessee	15.52	17.93	42.46	[12.66 , 18.38]
Texas	16.34	2.84	15.45	[12.38 , 20.30]
Virginia	15.83	12.58	35.18	[12.94 , 18.73]

**Statistical interval was the 95% confidence interval assuming the t-distribution and unknown variance.*

4. Conclusions

The analysis of mill locations for the emerging bioeconomy using Bayesian logistic regression in the presence of trucking delays had several important outcomes.

- Transportation delays for trucks are statistically significant in influencing site location of bioenergy plants.
- Transportation delays for trucks can greatly impact the cost of trucking biomass.
- The logistic model using Bayesian inference is a good predictive model for predicting preferred site locations and identifying non-preferred locations.
- The higher probability locations for larger biomass using mills are clustered in southeast Alabama, southern Georgia, central North Carolina, and the Mississippi Delta regions.
- The higher probability locations for smaller biomass mills are clustered in southern Alabama, southern Georgia, southeast Mississippi, southern Virginia, west Louisiana, west Arkansas, and the east Texas regions.

5. References

Berwick, M. and Farooq, M. 2003. Trucking cost model for transportation managers. Upper Great Plains Transportation Institute, North Dakota State University. Fargo, ND.

Biomass Research and Development Board. 2008. National biofuels action plan. U.S. Department of Agriculture (USDA) and U.S. Department of Energy (DOE). 24p.

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., Bowning, 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 225(2), 354-361.

Chin, S., Hwang H., Peterson B.E., Han L.D. and Chin C. 2009. "Routing hazardous materials around the District of Columbia area," Journal of Transportation Safety and Security 1(4),296-313.

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

Council of State Governors. 2010. Green freight transportation. http://knowledgecenter.csg.org/drupal/system/files/CR_GreenFreight.pdf

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. 2000. Bayesian Data Analysis. Chapman & Hall, London.

Gillett, J.C. 2011. Monetizing truck freight and the cost of delay for major truck routes in Georgia. M.S. Thesis. Georgia Institute of Technology. Atlanta, GA. 59 p.

Guess, F. 1996. "Bayesian applications," STATS 15, 15-17. Invited Paper.

Han, L.D., Chin, S. and Hwang, H. 2003. Estimating adverse weather impacts on major US highway network. TRB Annual Meeting Compendium of Papers CD ROM, No.03-2155, January 2003, 17p.

Han, L.D., Yuan, F. and Urbanik II, T. 2007. "What is an effective evacuation operation," ASCE Journal of Urban Planning and Development 133(1), 3-8.

Hilborn, R., Pikitch, E., and McAllister, M. 1994. "A Bayesian estimation and decision analysis for an age-structured model using biomass survey data," Fisheries Research 19, 17-30.

Hwang, H., Chin S., Han L.D. and Yuan F. 2006. Synthesizing truck origin-destination table for local transportation analysis zones: a Nashville metropolitan area case study. *Transportation Research Circular*, E-C088, pp.145-148.

Jaakkola, T.S. and Jordan, M.I. 1996. A variational approach to Bayesian logistic regression models and their extensions. Proceedings of the sixth international workshop on artificial intelligence and statistics.

Jaynes, E.T. 1976. "Confidence interval vs. Bayesian intervals," In; W.L. Harper and C.A. Hooker, Editors, Foundations of probability theory, statistical inference, and statistical theories of science, vol. II, D. Reidel, Dordrecht, Netherlands. 175-257.

Knight, D.K. 2009. "Wood bioenergy," Hatton-Brown Publishers, Inc., Montgomery, AL.

Lambert, B. 2008. Thoughts on the importance of freight to the state of Alabama. Institute for Trade and Transportation Studies. http://www.alabamat2.org/files/pdf/2010-trans conf/Freight%20in%20the%20Southeast%20Bruce%20Lambert%201.pdf

Moon, E., Saveyn, B., Proost, S., and Hermy, M. 2008. "Optimal location of new forests in a suburban region," Journal of Forest Economics 14(1), 5-27.

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

Oliveira-Neto, F.M., Han L.D. and Jeong, M.K. 2009. Tracking large trucks in real time with license plate recognition and text-mining techniques. Transportation Research Record, No.2121, TRB of the National Academies, Washington, D.C., pp.121-127. DOI: 10.3141/2121-13

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.

Perdue, J.H., Young, T.M., and Rials, T.G. 2011. "*The Biomass Site Assessment Model – BioSAT*," Final Report for U.S. Forest Service, Southern Research Station submitted by the University of Tennessee, Knoxville, 282.

Si, J., Urbanik II, T. and Han, L.D. 2007. Effectiveness of alternative detector configurations for option zone protection on high-speed approaches to traffic signal. Transportation Research Record. Vol. 2035, TRB of the National Academies, Washington, D.C. pp.107-113. DOI:10.3141/2035-12

Stewart, P., 2009. "Energy and the wood fiber supply chain," *Forests & People Magazine* (Q1). Retrieved from (http://www.forest2market.com/uploads/Legacy/energy-and-the-wood-fiber-supply-chain_1.pdf).

U.S. Census Bureau, 2010(a). "2010 Census ZIP code tabulation areas" [Data file], Retrieved from (http://www.census.gov/geo/reference/zctas.html).

U.S. Census Bureau, 2010(b). "2010 Urban and rural classification" [Data file], Retrieved from (http://www.census.gov/geo/reference/urban-rural.html).

U.S. Department of Agriculture National Agricultural Statistics Service. 2008. "Census of agriculture: farm net income" [Data file], Retrieved from (http://www.nass.usda.gov/#top)

U.S. Environmental Protection Agency. 2011. "Ecoregions of the United States" [Data file], Retrieved from(http://www.epa.gov/bioiweb1/html/usecoregions.html).

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

U.S. National Elevation Dataset. 2010. "National elevation dataset 1 arc second"[Data file], Retrieved from(http://seamless.usgs.gov/ned1.php).

U.S. National Land Cover Database. 2006. "Multi-resolution land characteristics consortium" [Data file], Retrieved from (http://www.mrlc.gov/nlcd2006.php).

Young, T.M., Perdue J.H., Hartsell A.J., Hodges, D.G. and Rials, T.G. 2008. Optimal siting for biorefineries using the Biomass Siting Analysis Tool (BioSAT). 2008 Forest Inventory and Analysis Symposium. Oct. 21-23. Park City, UT. USDA Forest Service Proceedings – RMRS-P-56. 42:1-22.

Young, T.M., Zaretski, R.L., Perdue, J.H., Guess, F.M., and Liu, X. 2011. "Logistic regression models of factors influencing the location of bioenergy and biofuels plants," Bioresources 6(1), 317-328.

Appendix A.

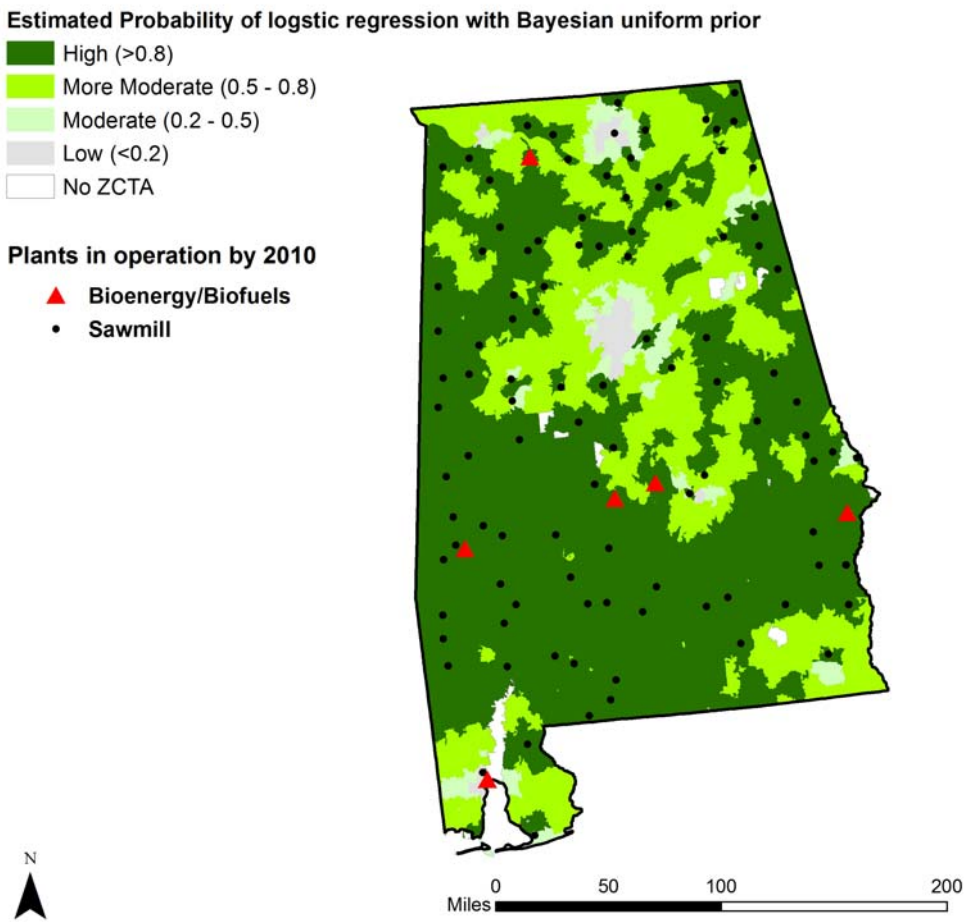


Figure A1. Estimated probability locations for Group I in Alabama.

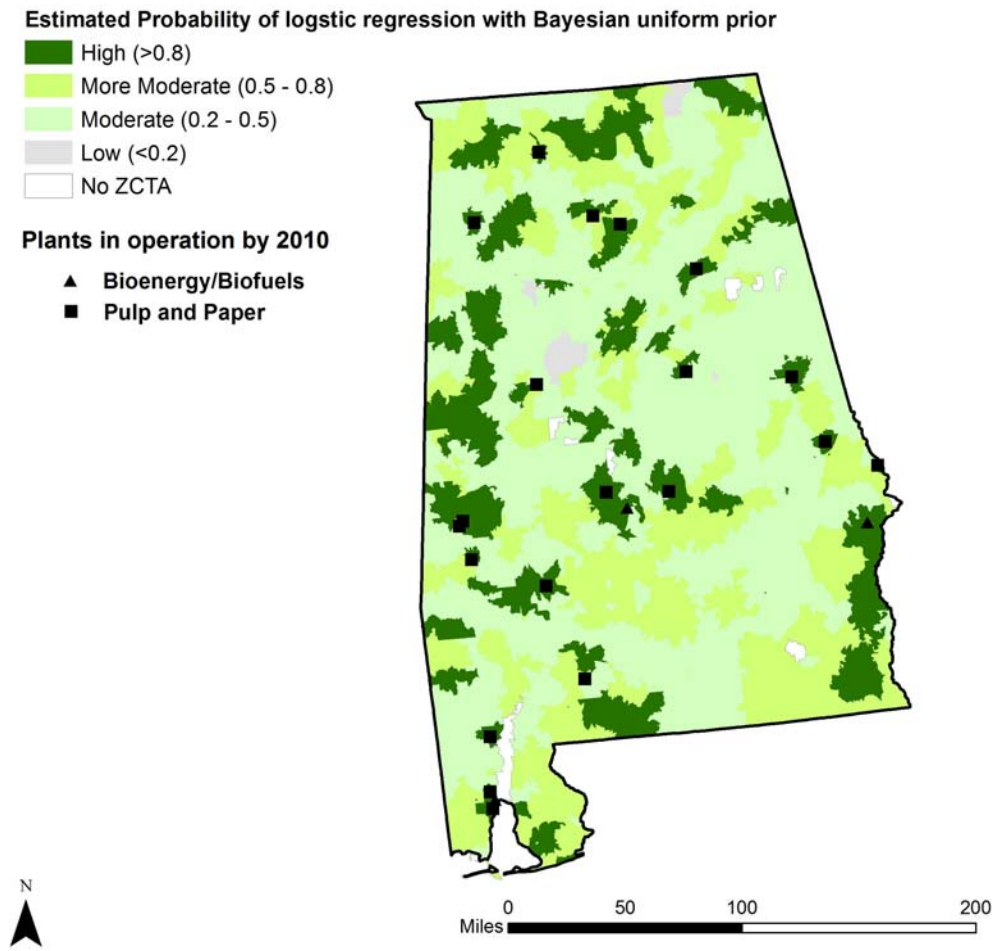


Figure A2. Estimated probability locations for Group II in Alabama.

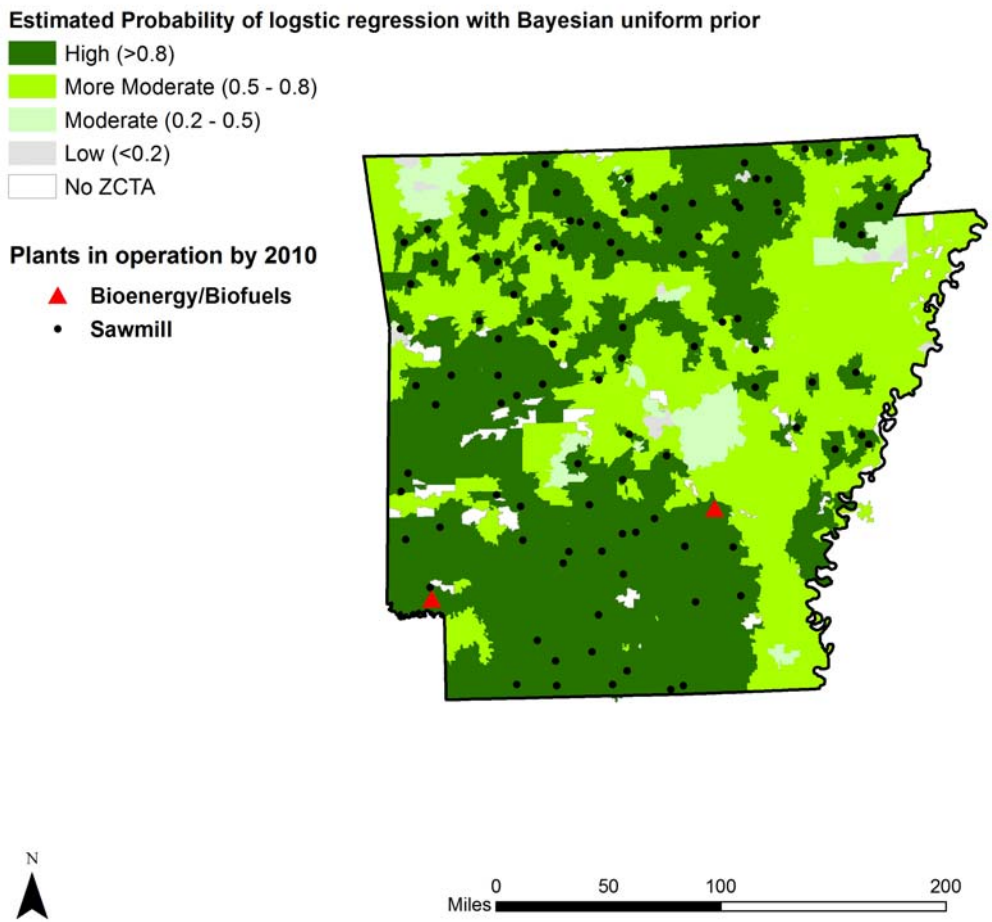


Figure A3. Estimated probability locations for Group I in Arkansas.

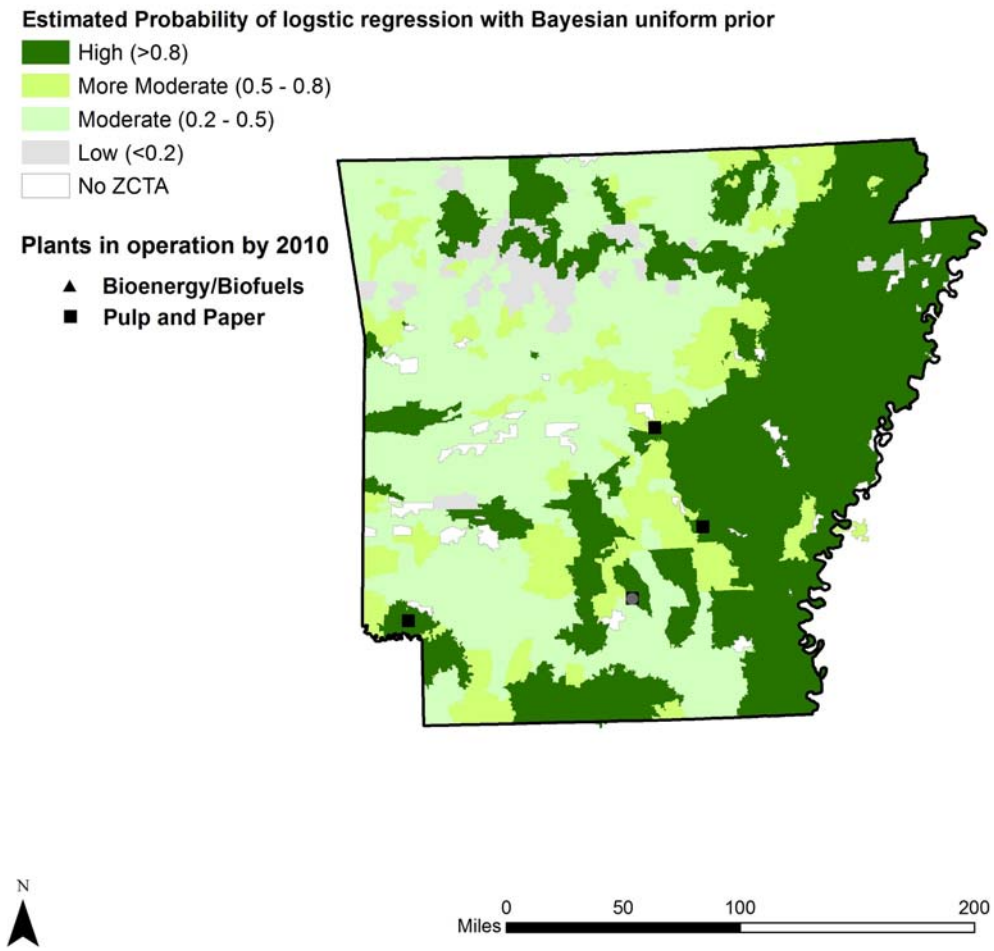


Figure A4. Estimated probability locations for Group II in Arkansas.

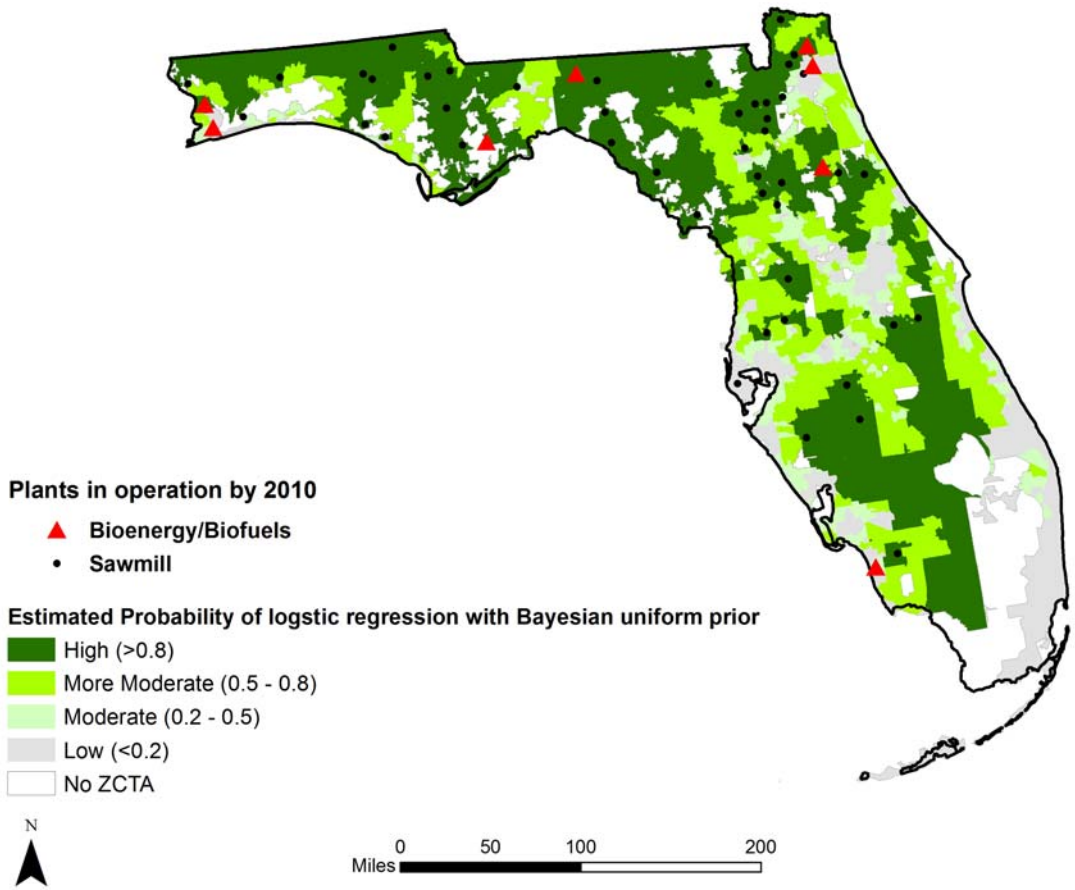


Figure A5. Estimated probability locations for Group I in Florida.

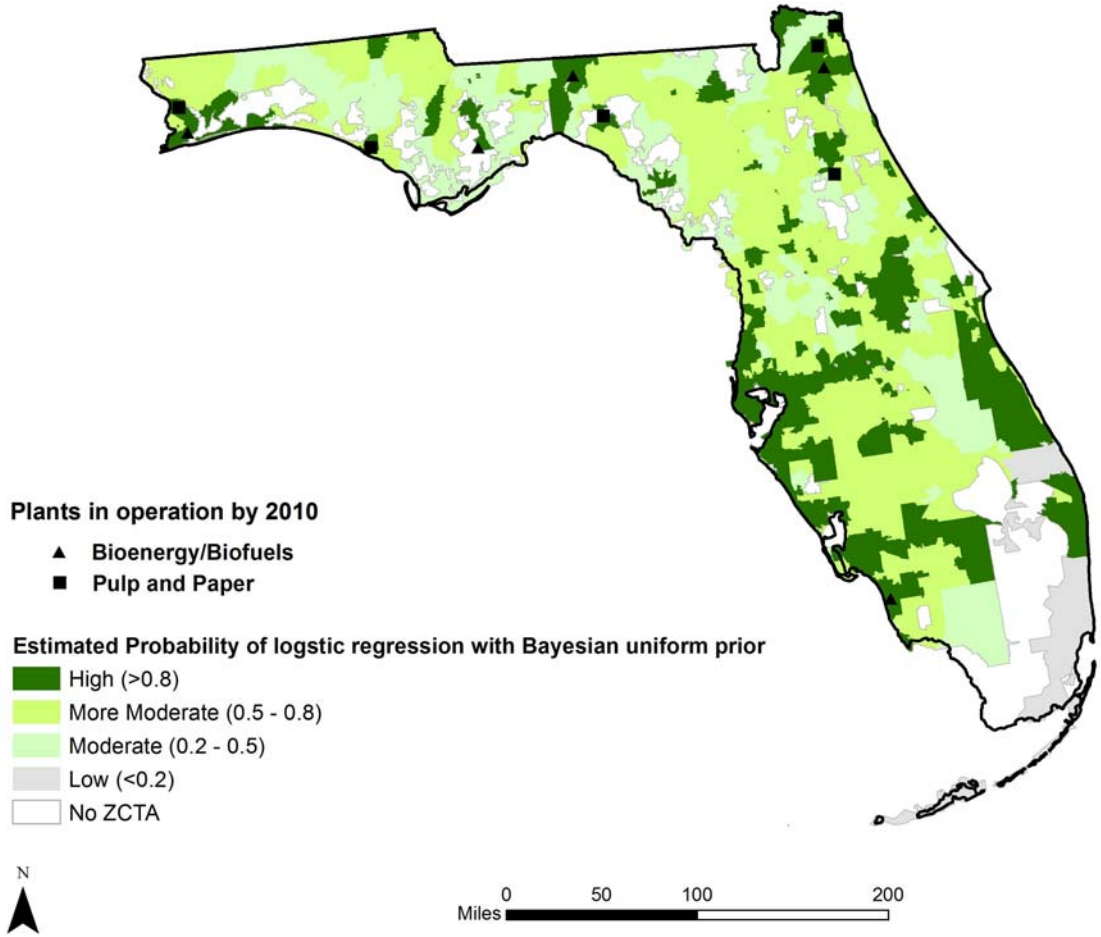


Figure A6. Estimated probability locations for Group II in Florida.

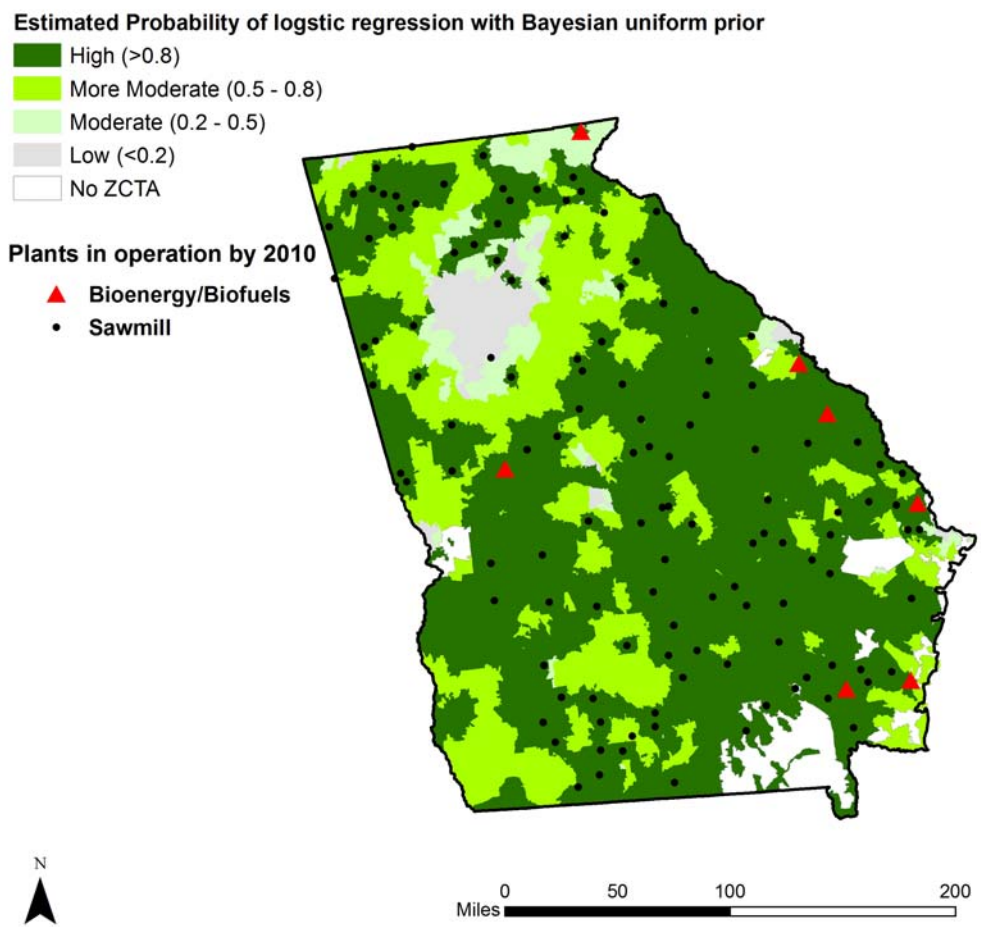


Figure A7. Estimated probability locations for Group I in Georgia.

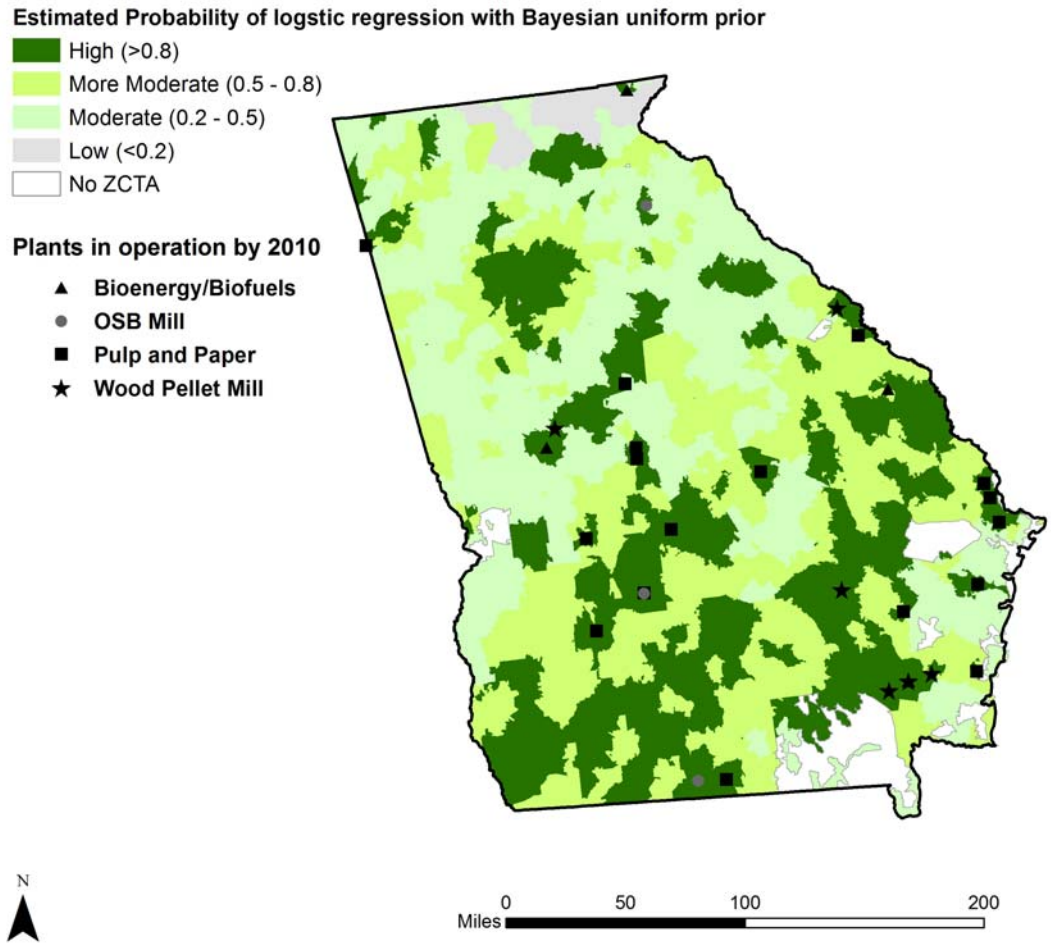


Figure A8. Estimated probability locations for Group II in Georgia.

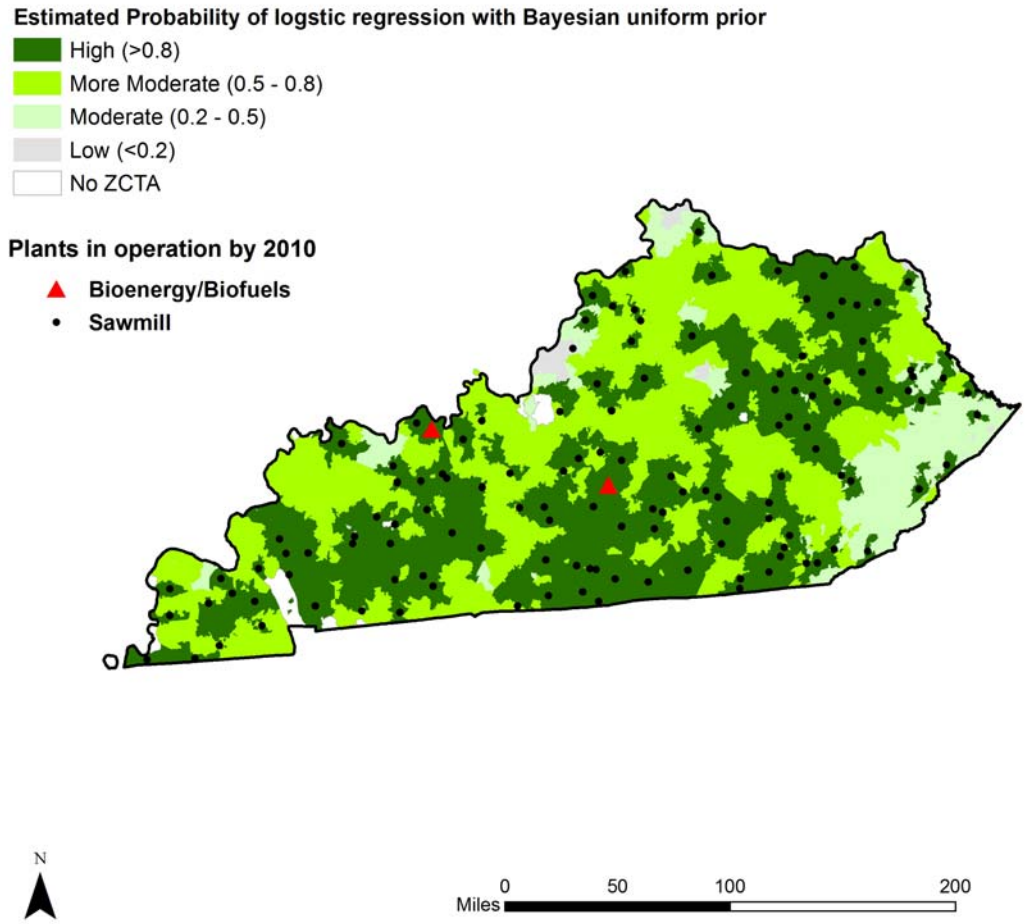


Figure A9. Estimated probability locations for Group I in Kentucky.

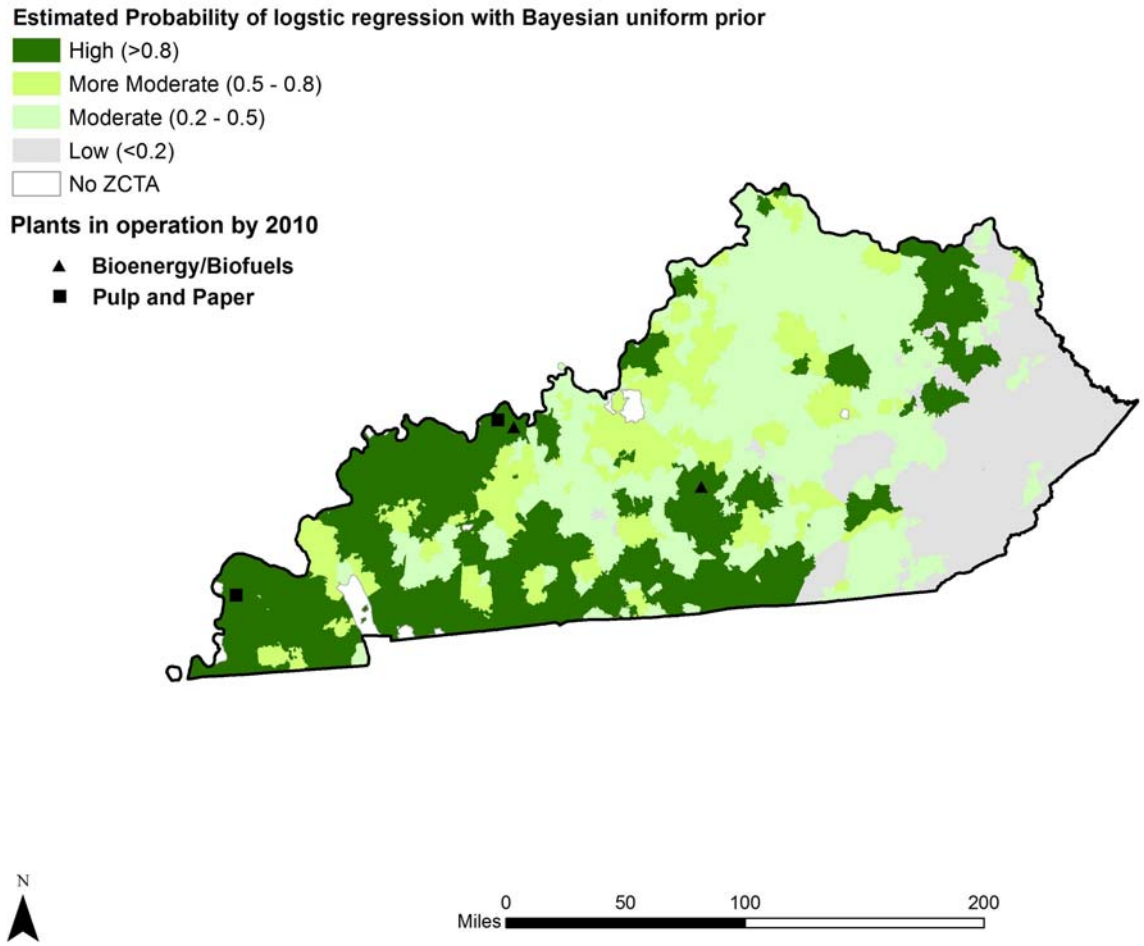


Figure A10. Estimated probability locations for Group II in Kentucky.

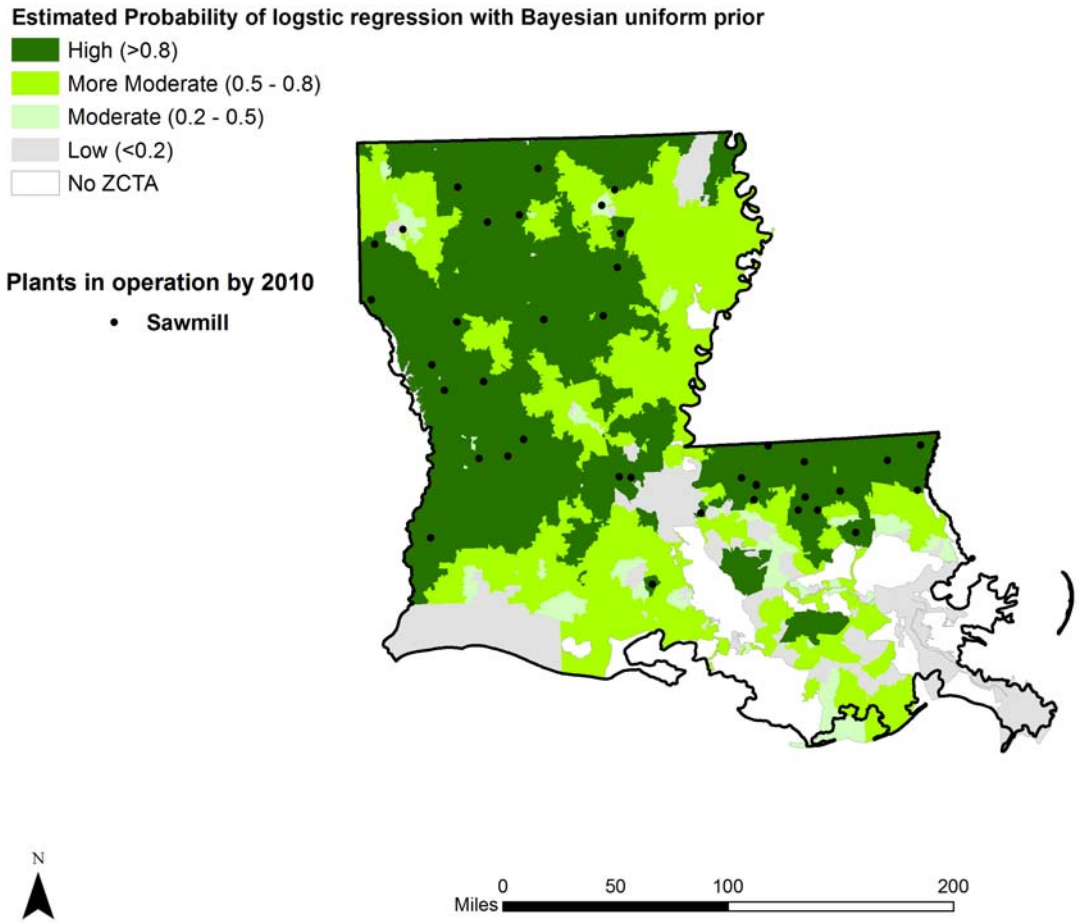


Figure A11. Estimated probability locations for Group I in Louisiana.

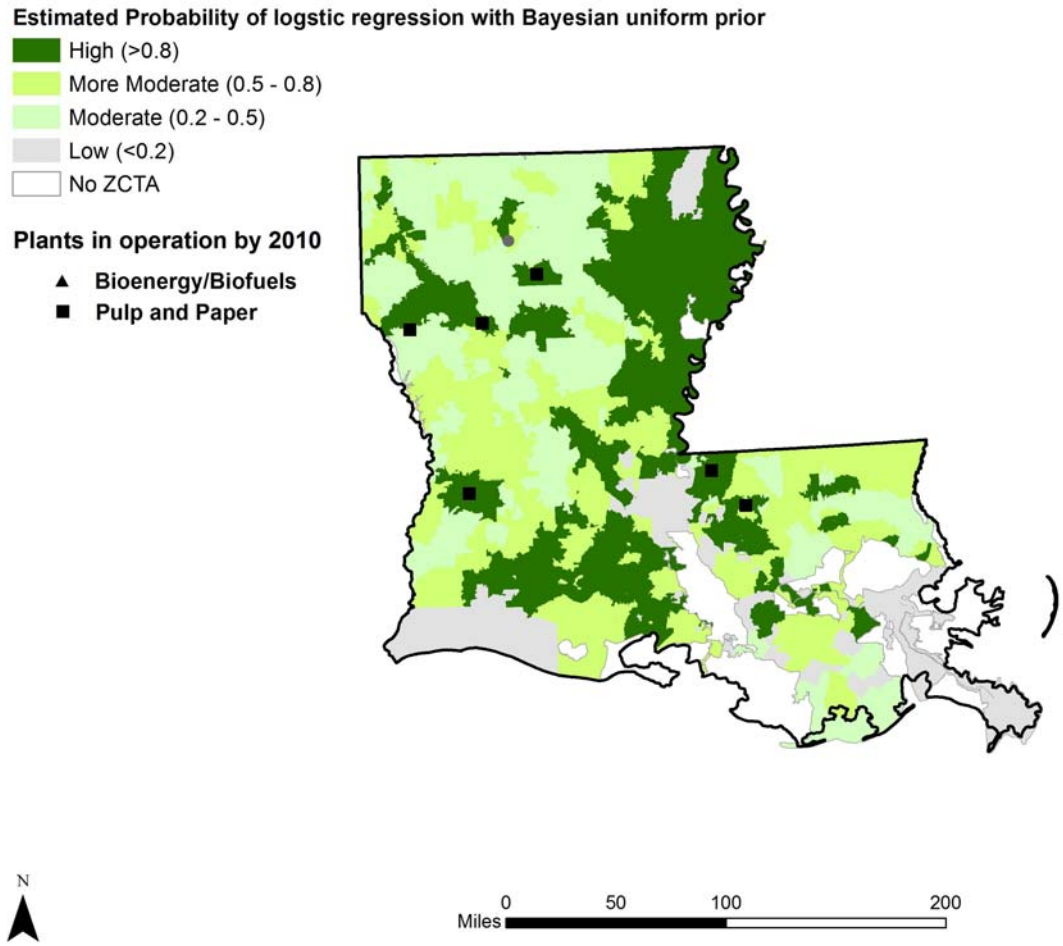


Figure A12. Estimated probability locations for Group II in Louisiana.

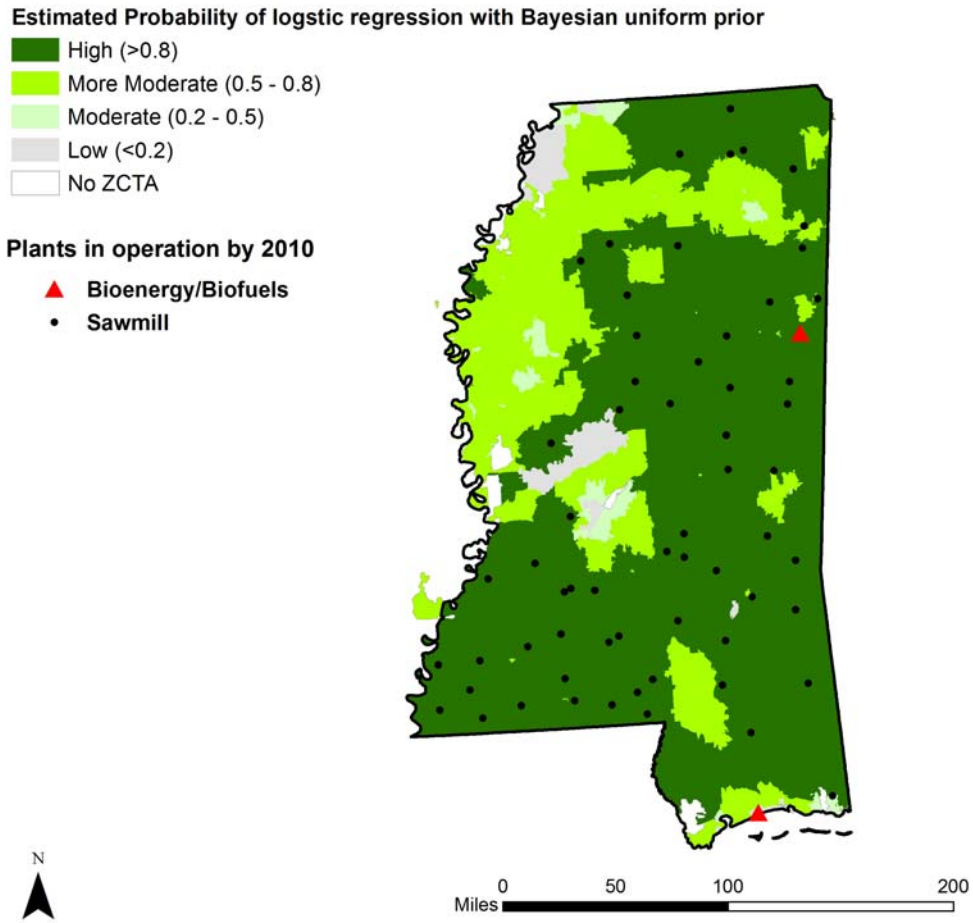


Figure A13. Estimated probability locations for Group I in Mississippi.

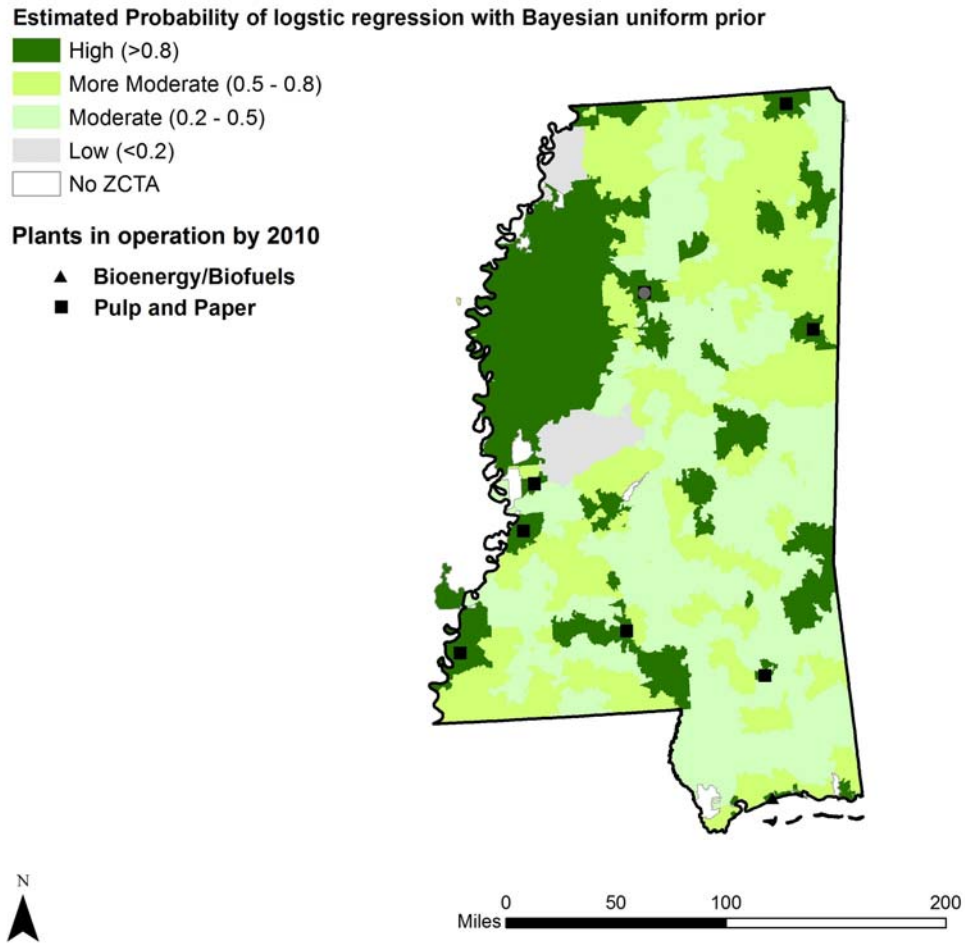


Figure A14. Estimated probability locations for Group II in Mississippi.

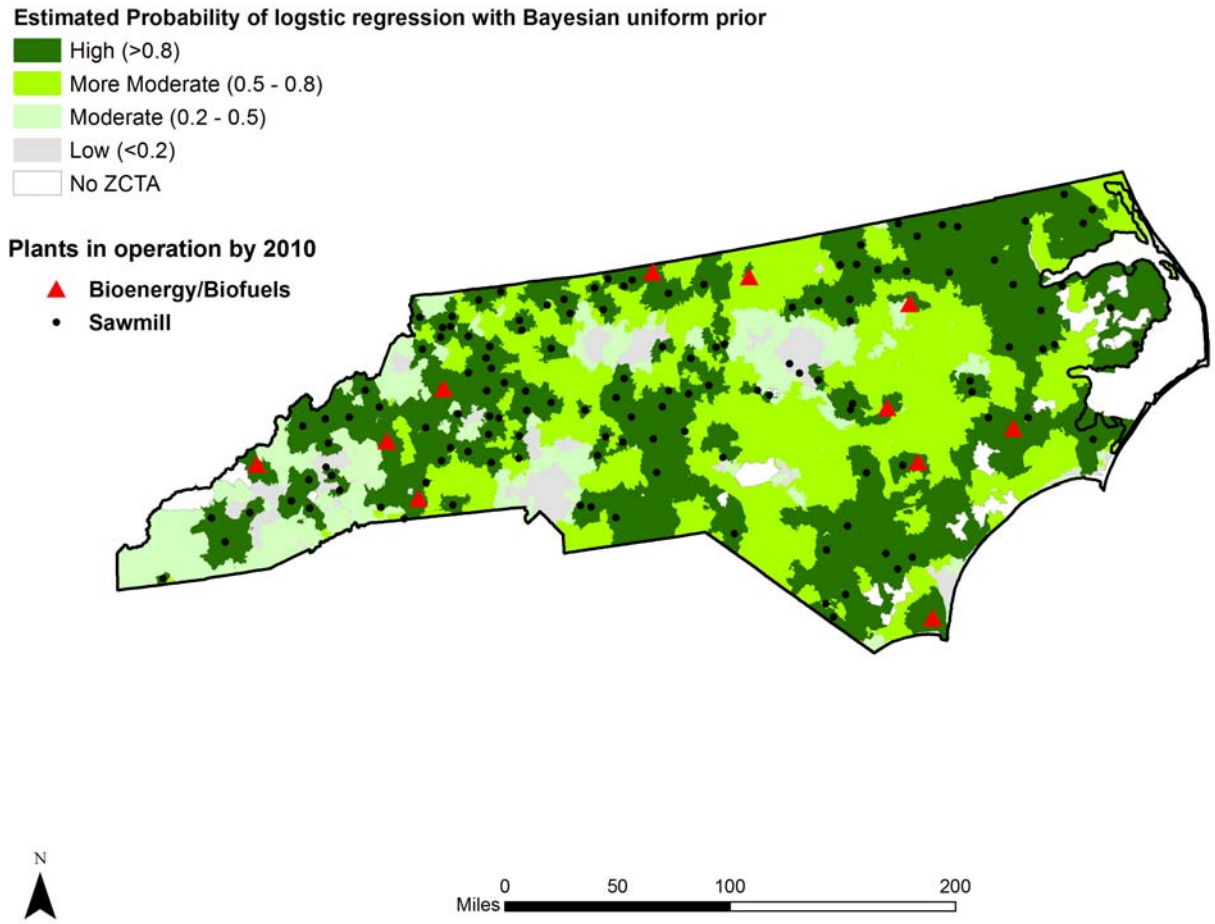


Figure A15. Estimated probability locations for Group I in North Carolina.

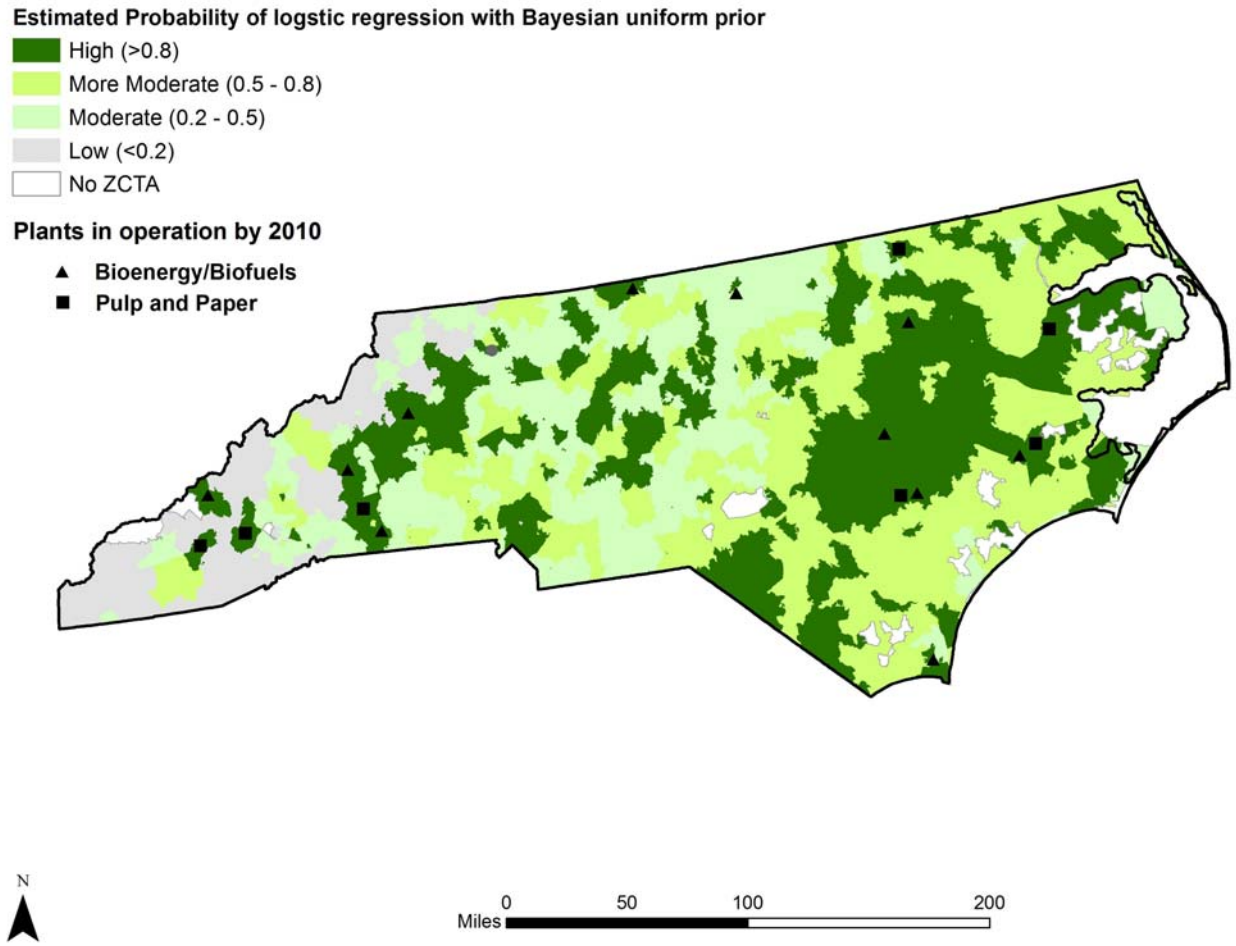


Figure A16. Estimated probability locations for Group II in North Carolina.

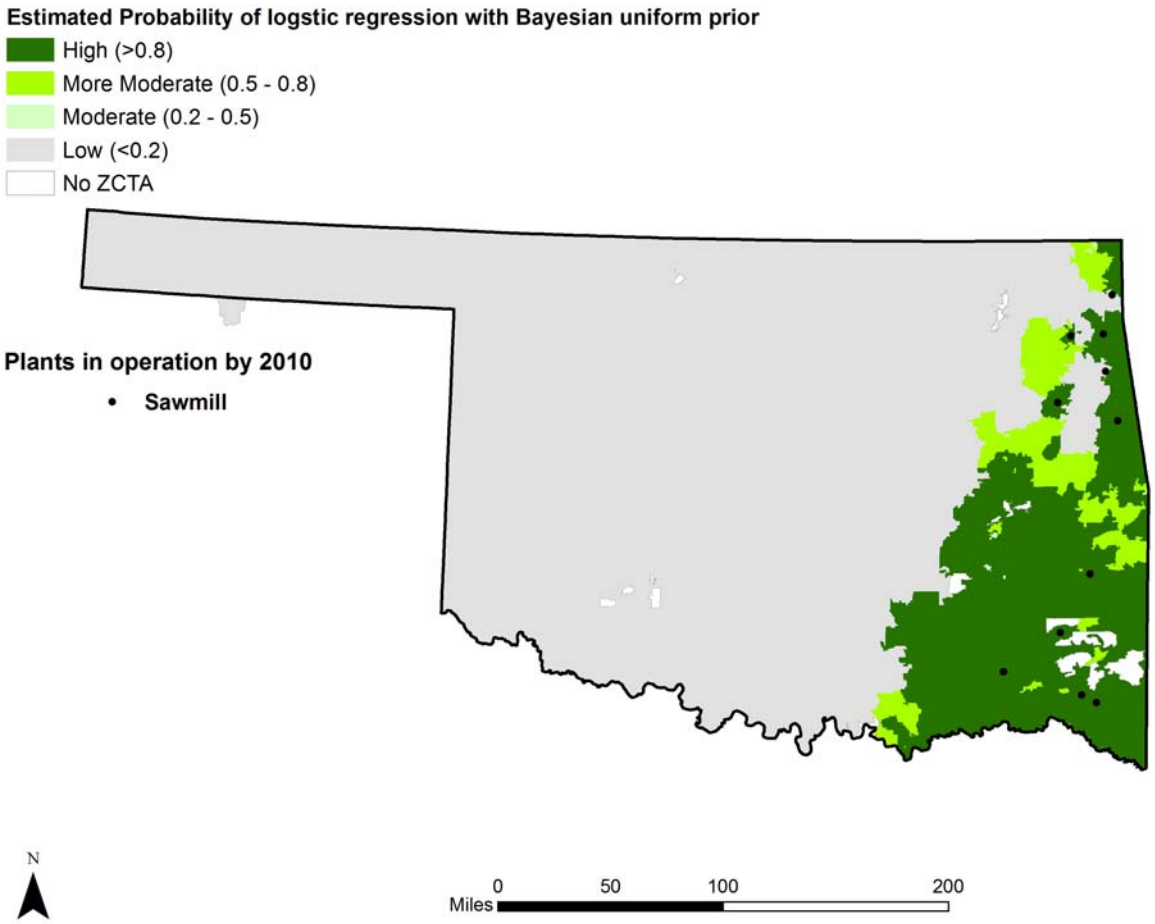


Figure A17. Estimated probability locations for Group I in Oklahoma.

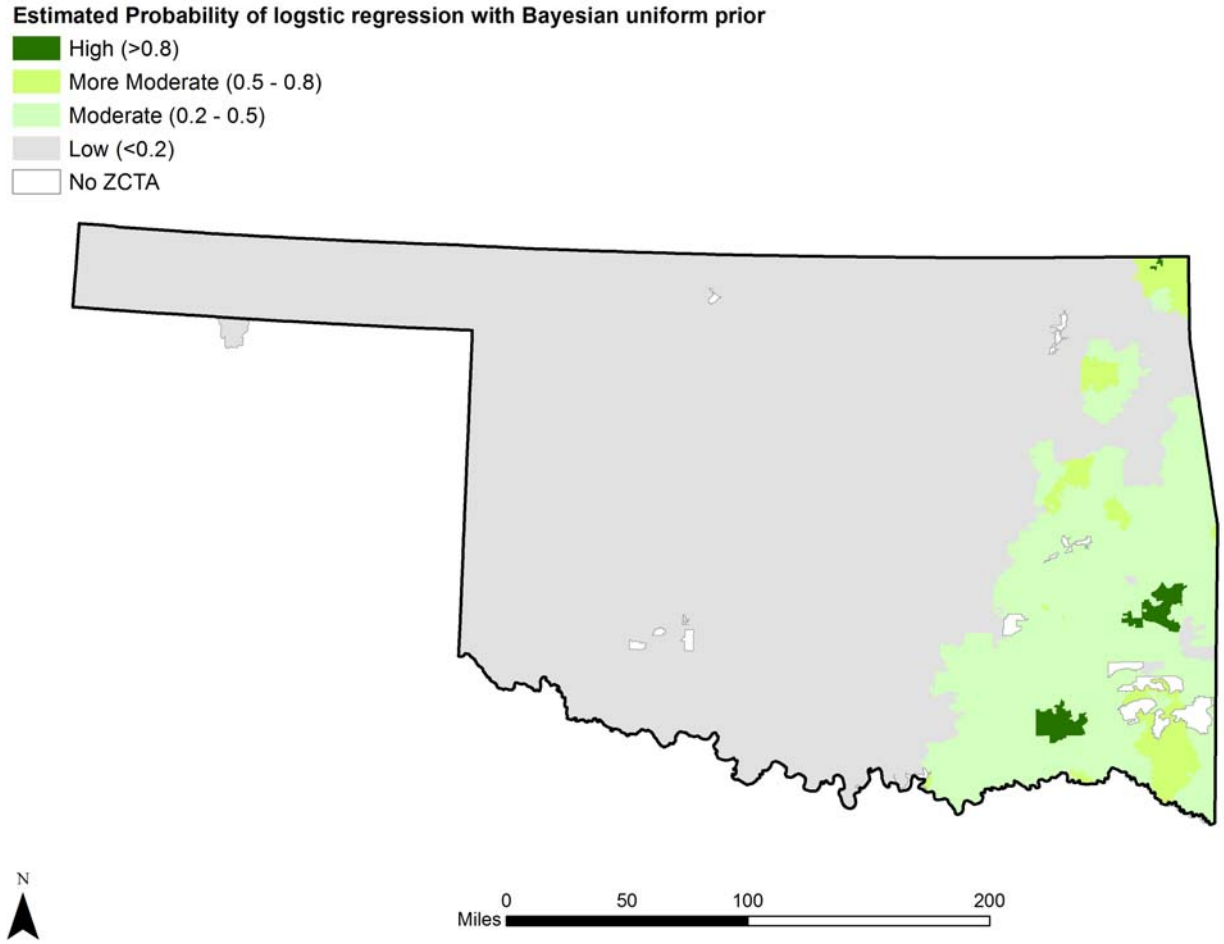


Figure A18. Estimated probability locations for Group II in Oklahoma.

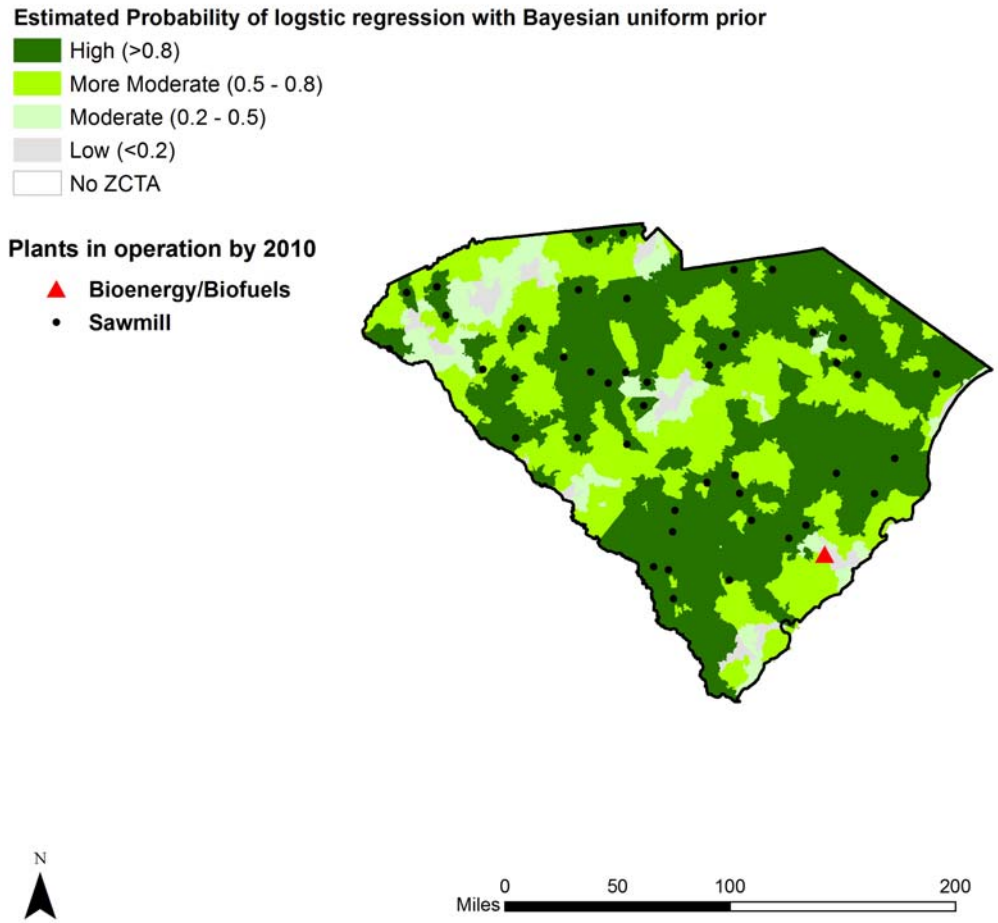


Figure A19. Estimated probability locations for Group I in South Carolina.

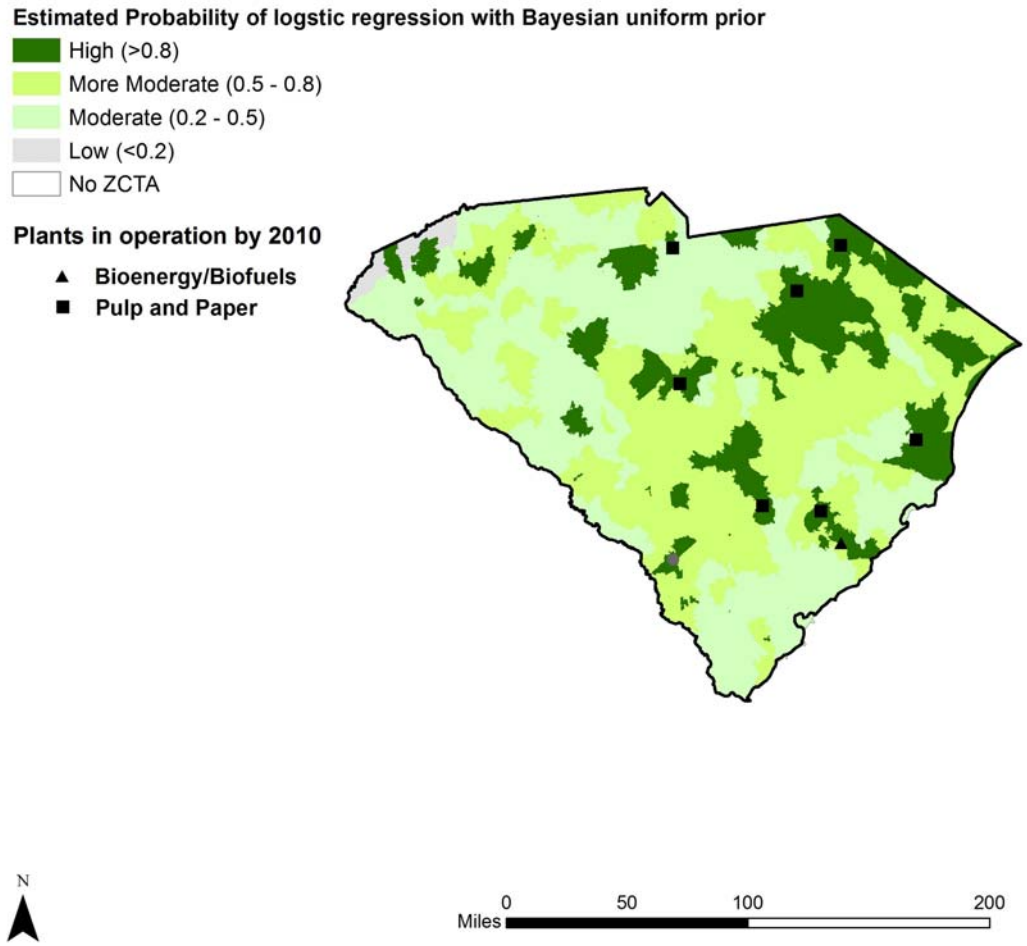


Figure A20. Estimated probability locations for Group II in South Carolina.

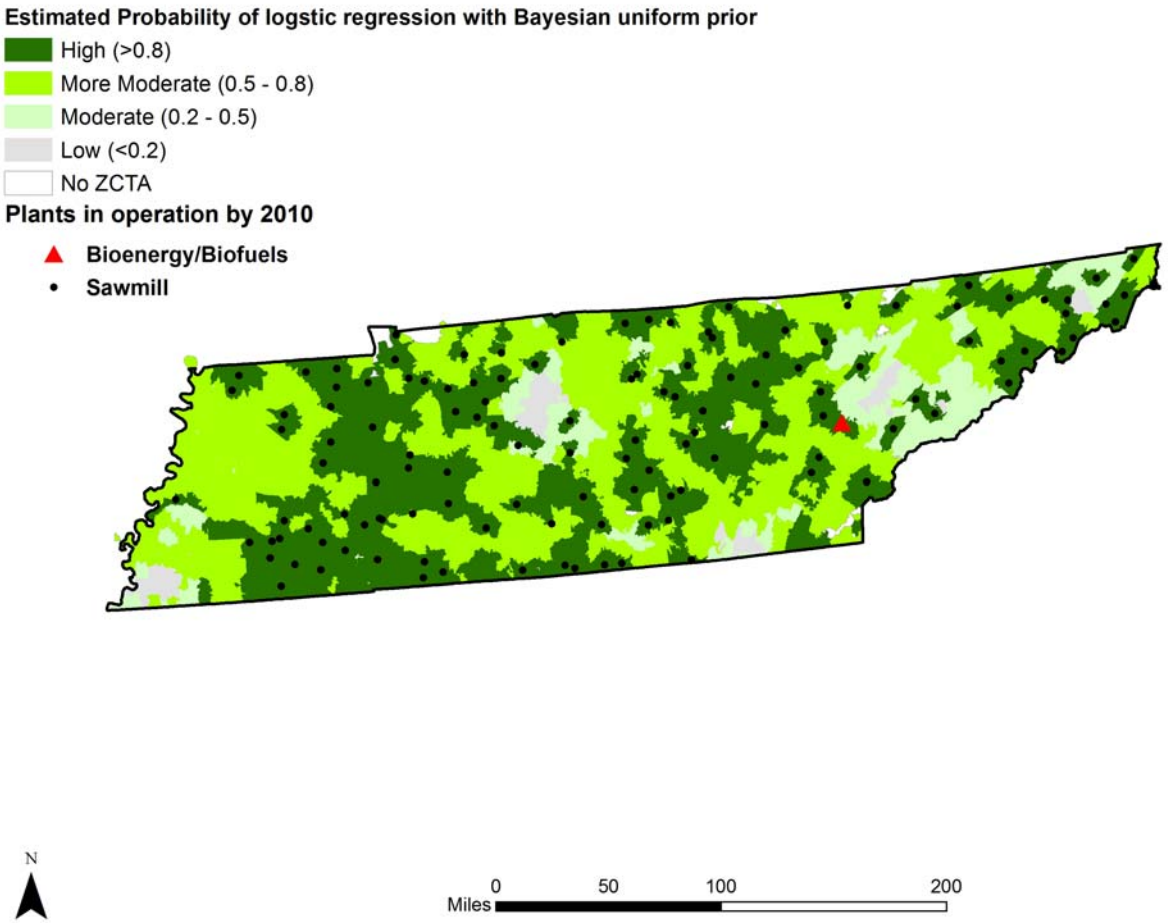


Figure A21. Estimated probability locations for Group I in Tennessee.

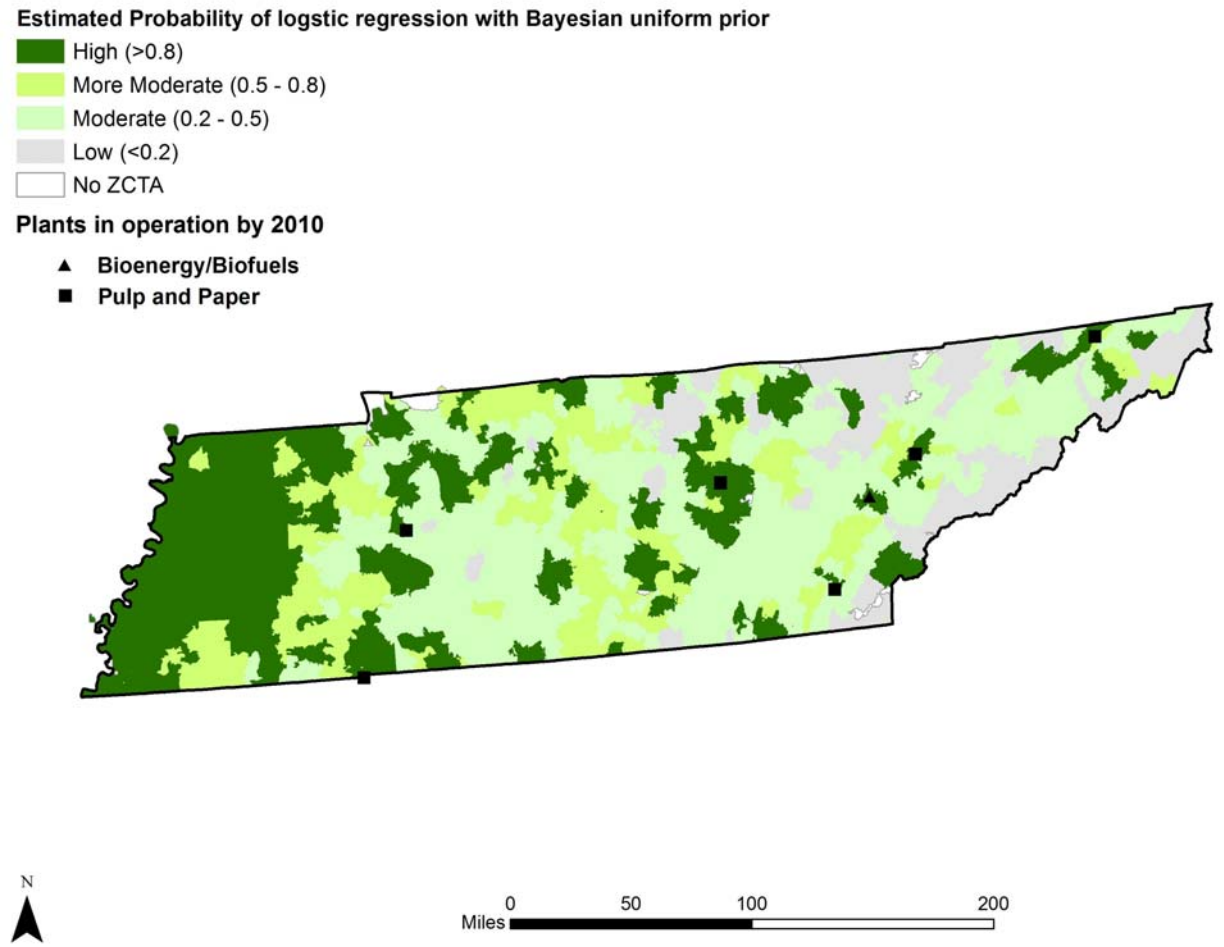


Figure A22. Estimated probability locations for Group II in Tennessee.

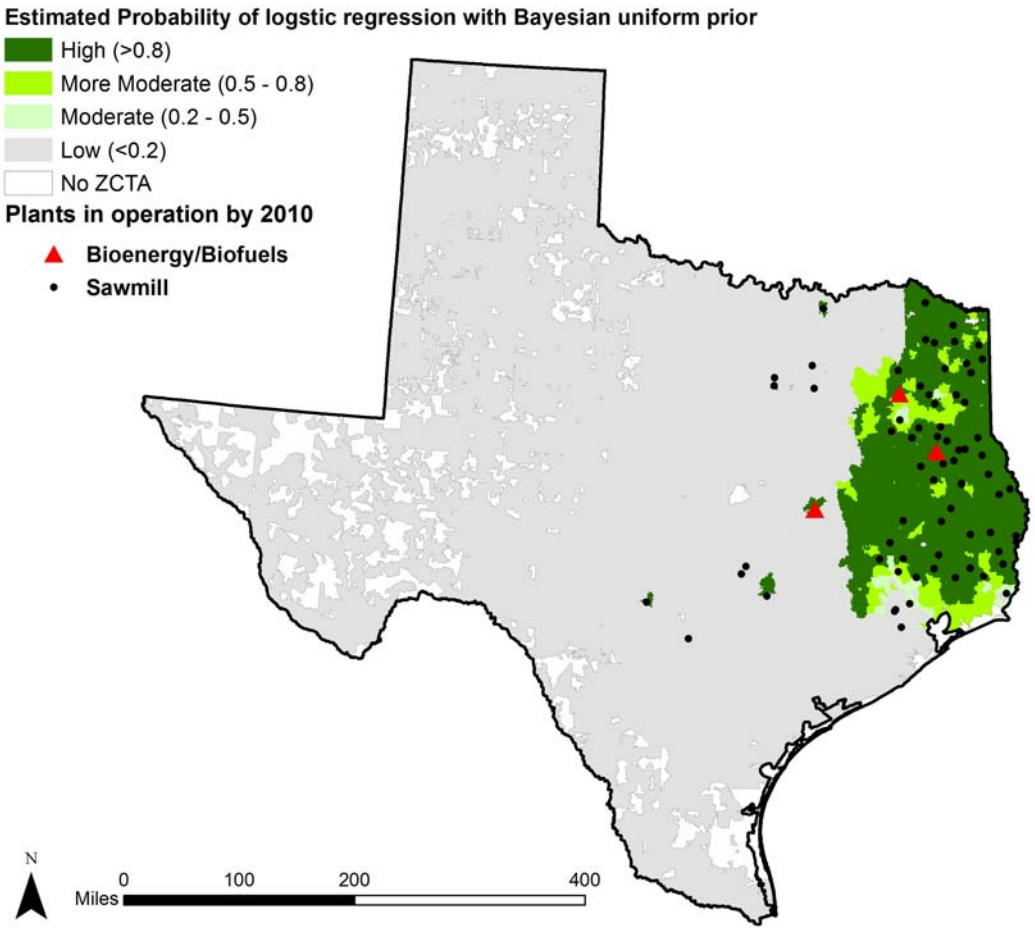


Figure A23. Estimated probability locations for Group I in Texas.

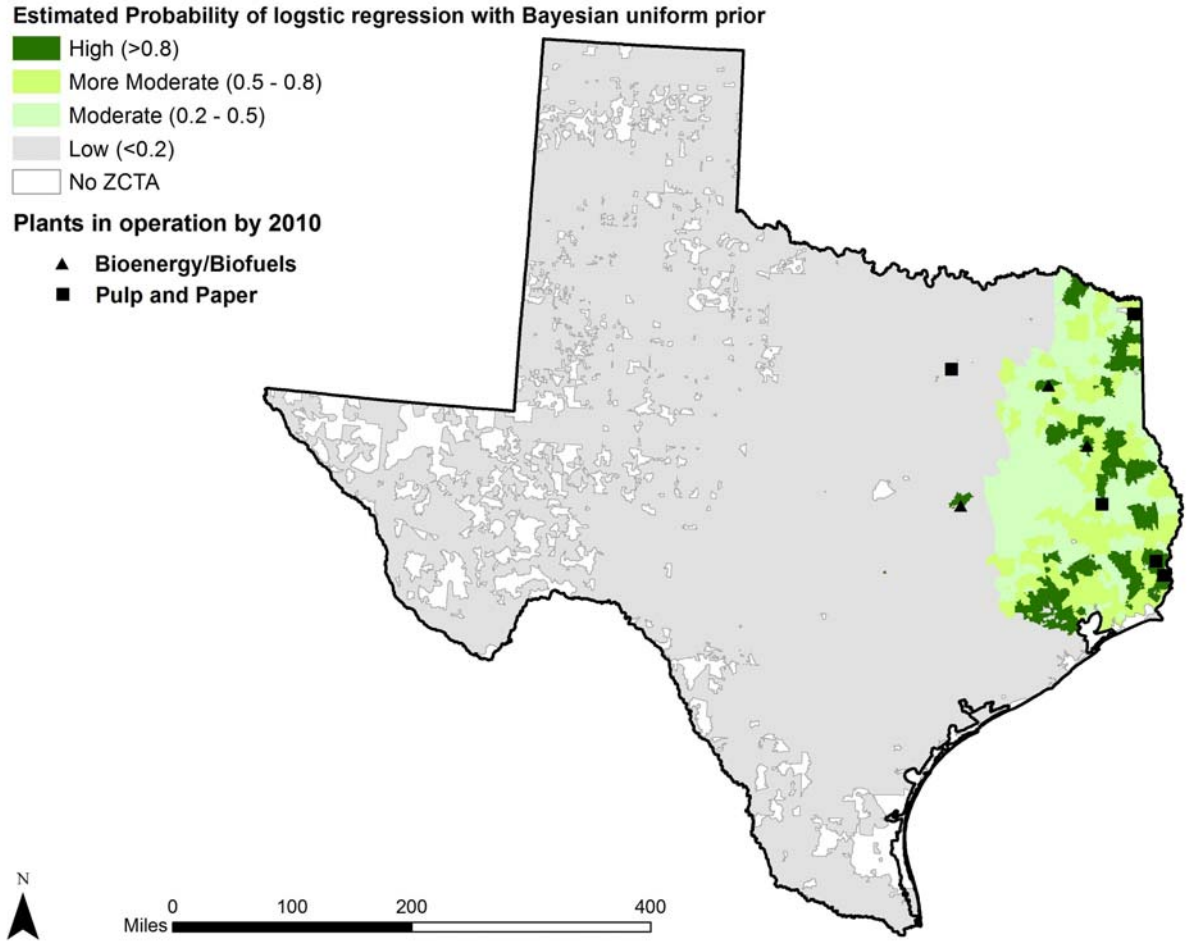


Figure A24. Estimated probability locations for Group II in Texas.

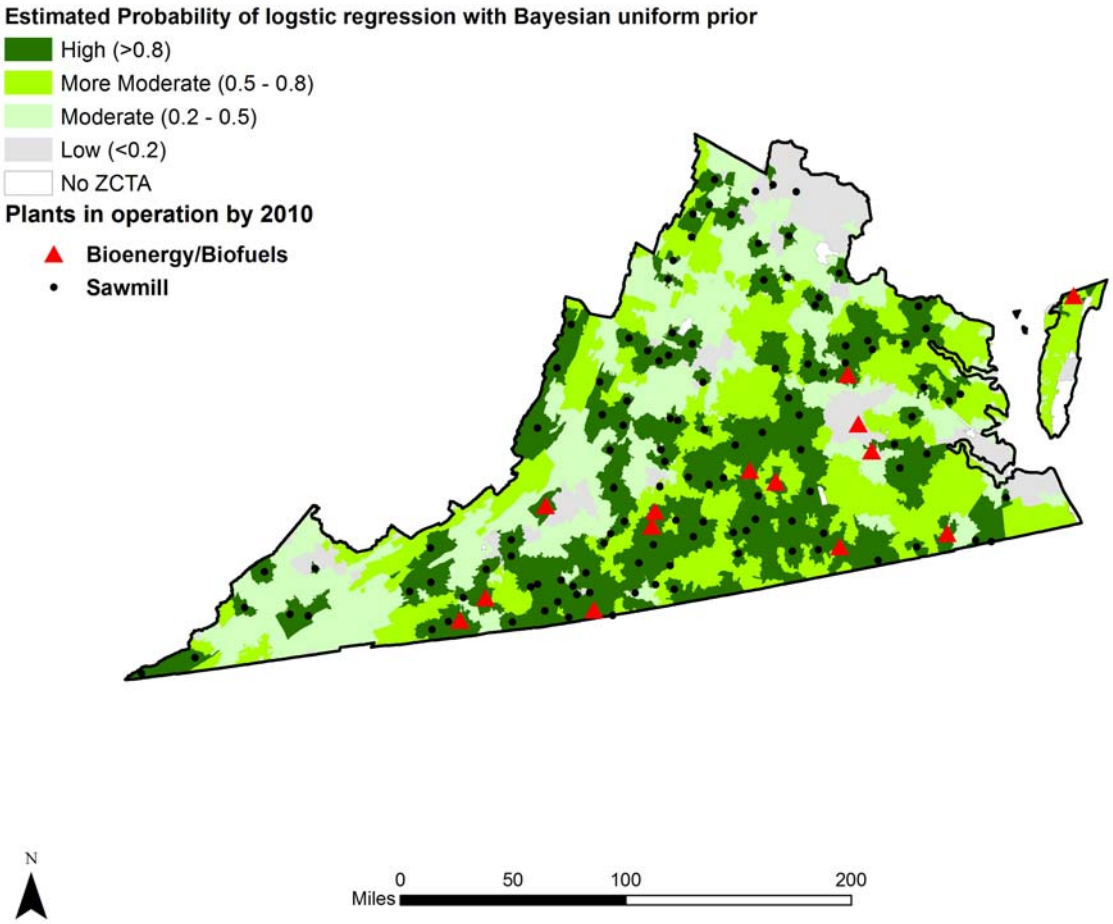


Figure A25. Estimated probability locations for Group I in Virginia.

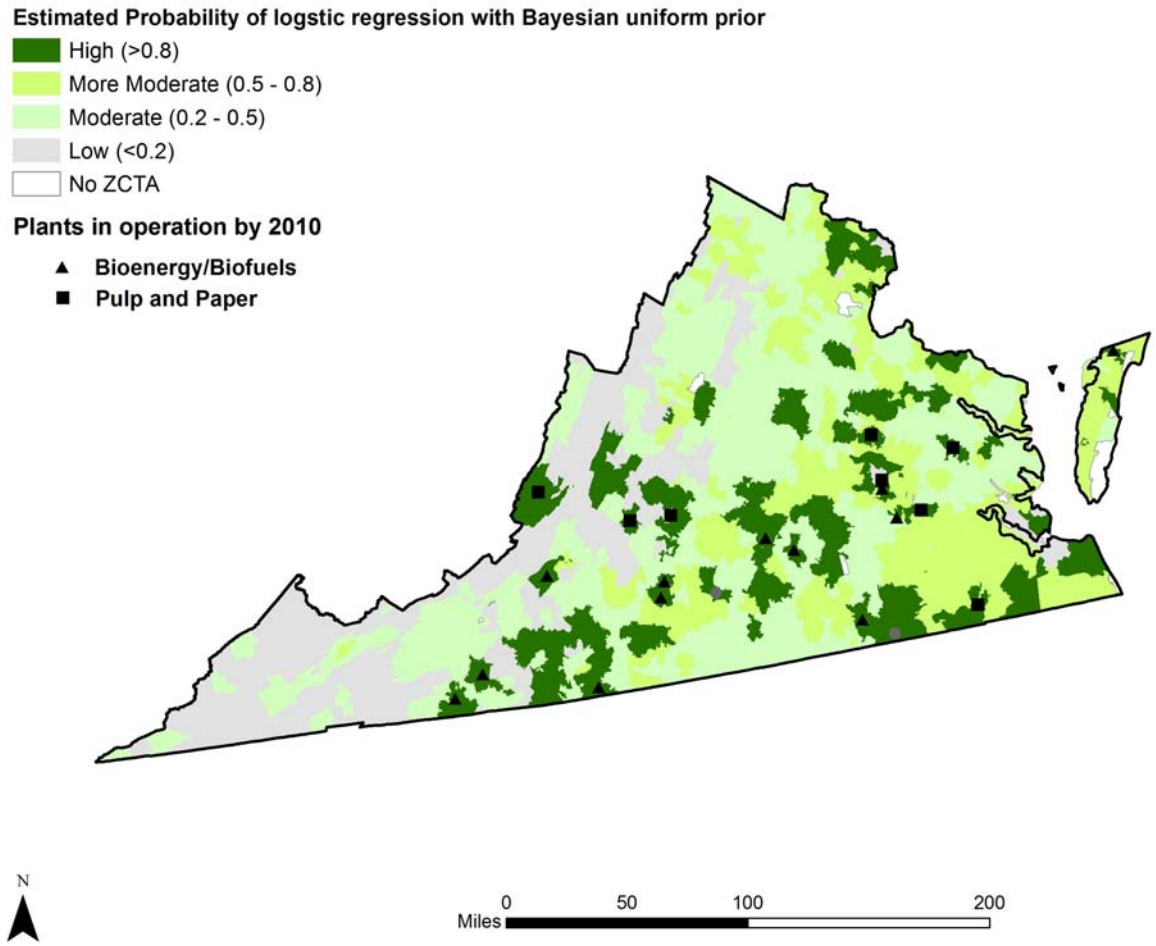


Figure A26. Estimated probability locations for Group II in Virginia.

Appendix B.

Queues

5: Greenback Road & 72 3/7/2013

Lane Group	EBL	EBT	WBL	WBT	WBR	NBT	SBT	SBR
Lane Group Flow (vph)	566	393	5	414	576	15	347	223
v/c Ratio	0.94	0.19	0.01	0.19	0.48	0.04	1.01	0.38
Control Delay	40.2	6.2	5.4	6.2	2.0	18.5	82.6	5.8
Queue Delay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total Delay	40.2	6.2	5.4	6.2	2.0	18.5	82.6	5.8
Queue Length 50th (ft)	224	37	1	39	0	4	~175	0
Queue Length 95th (ft)	#469	55	4	57	33	18	#344	50
Internal Link Dist (ft)		5440		10142		3051	600	
Turn Bay Length (ft)	250		175		600			125
Base Capacity (vph)	604	2080	618	2234	1208	427	345	580
Starvation Cap Reductn	0	0	0	0	0	0	0	0
Spillback Cap Reductn	0	0	0	0	0	0	0	0
Storage Cap Reductn	0	0	0	0	0	0	0	0
Reduced v/c Ratio	0.94	0.19	0.01	0.19	0.48	0.04	1.01	0.38

Intersection Summary

~ Volume exceeds capacity, queue is theoretically infinite.
Queue shown is maximum after two cycles.

95th percentile volume exceeds capacity, queue may be longer.
Queue shown is maximum after two cycles.

Figure B1. Sample results of travel simulation with 25MG local trucks at intersections near Vonore, TN

Queues
5: Greenback Road & 72 3/7/2013

Lane Group	EBL	EBT	WBL	WBT	WBR	NBT	SBT	SBR
Lane Group Flow (vph)	566	401	5	422	584	15	355	223
v/c Ratio	0.96	0.18	0.01	0.19	0.48	0.03	0.99	0.38
Control Delay	44.5	6.1	5.4	6.2	2.0	17.2	77.3	5.6
Queue Delay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total Delay	44.5	6.1	5.4	6.2	2.0	17.2	77.3	5.6
Queue Length 50th (ft)	216	36	1	38	0	3	165	0
Queue Length 95th (ft)	#452	54	4	56	33	17	#330	49
Internal Link Dist (ft)		5440		10150		3051	600	
Turn Bay Length (ft)	250		175		600			125
Base Capacity (vph)	591	2215	602	2218	1210	432	357	586
Starvation Cap Reductn	0	0	0	0	0	0	0	0
Spillback Cap Reductn	0	0	0	0	0	0	0	0
Storage Cap Reductn	0	0	0	0	0	0	0	0
Reduced v/c Ratio	0.96	0.18	0.01	0.19	0.48	0.03	0.99	0.38

Intersection Summary

95th percentile volume exceeds capacity, queue may be longer.
Queue shown is maximum after two cycles.

Figure B2. Sample results of travel simulation with 50MG local trucks at intersections near Vonore, TN

Queues

5: Greenback Road & 72 3/7/2013

Lane Group	EBL	EBT	WBL	WBT	WBR	NBT	SBT	SBR
Lane Group Flow (vph)	566	407	5	428	591	15	363	223
v/c Ratio	0.97	0.20	0.01	0.20	0.49	0.03	1.01	0.37
Control Delay	48.8	6.7	5.8	6.7	2.1	17.9	80.9	5.5
Queue Delay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total Delay	48.8	6.7	5.8	6.7	2.1	17.9	80.9	5.5
Queue Length 50th (ft)	241	40	1	42	0	4	~183	0
Queue Length 95th (ft)	#481	60	5	62	34	18	#354	49
Internal Link Dist (ft)		5440		10142		3051	600	
Turn Bay Length (ft)	250		175		600			125
Base Capacity (vph)	582	2039	596	2191	1202	447	361	597
Starvation Cap Reductn	0	0	0	0	0	0	0	0
Spillback Cap Reductn	0	0	0	0	0	0	0	0
Storage Cap Reductn	0	0	0	0	0	0	0	0
Reduced v/c Ratio	0.97	0.20	0.01	0.20	0.49	0.03	1.01	0.37

Intersection Summary

~ Volume exceeds capacity, queue is theoretically infinite.
Queue shown is maximum after two cycles.

95th percentile volume exceeds capacity, queue may be longer.
Queue shown is maximum after two cycles.

Figure B3. Sample results of travel simulation with 75MG local trucks at intersections near Vonore, TN

Queues

5: Greenback Road & 72 3/7/2013

Lane Group	EBL	EBT	WBL	WBT	WBR	NBT	SBT	SBR
Lane Group Flow (vph)	566	415	5	436	599	15	370	223
v/c Ratio	0.98	0.20	0.01	0.20	0.50	0.03	1.02	0.37
Control Delay	51.4	6.7	5.8	6.7	2.1	17.9	85.7	5.5
Queue Delay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total Delay	51.4	6.7	5.8	6.7	2.1	17.9	85.7	5.5
Queue Length 50th (ft)	245	41	1	43	0	4	~193	0
Queue Length 95th (ft)	#484	61	5	63	35	18	#363	49
Internal Link Dist (ft)		5440		10142		3051	600	
Turn Bay Length (ft)	250		175		600			125
Base Capacity (vph)	576	2039	590	2191	1205	446	361	597
Starvation Cap Reductn	0	0	0	0	0	0	0	0
Spillback Cap Reductn	0	0	0	0	0	0	0	0
Storage Cap Reductn	0	0	0	0	0	0	0	0
Reduced v/c Ratio	0.98	0.20	0.01	0.20	0.50	0.03	1.02	0.37

Intersection Summary

~ Volume exceeds capacity, queue is theoretically infinite.
Queue shown is maximum after two cycles.

95th percentile volume exceeds capacity, queue may be longer.
Queue shown is maximum after two cycles.

Figure B4. Sample results of travel simulation with 100MG local trucks at intersections near Vonore, TN