

MODELING HISTORIC, CURRENT, AND AVAILABLE ABOVEGROUND FOREST  
BIOMASS ALONG THE MISSOURI RIVER CORRIDOR

---

A Dissertation

Presented to

the Faculty of the Graduate School  
at the University of Missouri-Columbia

---

In Partial Fulfillment

of the Requirements for the Degree

Doctorate of Forestry

---

by

CHRISTOPHER W. BOBRYK

Dr. Hong S. He and Dr. Shibu Jose, Dissertation Supervisors

JULY 2014

© Copyright by Christopher W. Bobryk 2014

All Rights Reserved

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

MODELING HISTORIC, CURRENT, AND AVAILABLE ABOVEGROUND FOREST BIOMASS ALONG THE MISSOURI RIVER CORRIDOR

presented by Christopher W. Bobryk,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

---

Dr. Hong S. He

---

Dr. Shibu Jose

---

Dr. Stephen R. Shifley

---

Dr. Cuizhen Wang

---

Dr. Robert Jacobson

## ACKNOWLEDGMENTS

This research was made possible with help from many wonderful people and organizations. I would like to thank my advisors, Dr. Hong He and Shibu Jose, for attaining the financial support to sustain this research, which was provided by the University of Missouri Geographic Information System Mission Enhancement Program and The Center for Agroforestry, and for their encouragement that enabled me to be the best student, researcher, and professional I could be. I would also like to thank my committee members, Dr. Steve Shifley, Dr. Cuizhen Wang, and Dr. Robert Jacobson, who were tough, thorough, and enlightening. I thank all of them for taking the time to visit with me, helping me to think critically, and ultimately guiding me in the right direction.

The individuals of the GIS and Spatial Analysis Laboratory deserve many special thanks. I would not have made it through if it were not for people like Jacob Fraser, Jeff Schneiderman, Mike Sunde, Wenchi Jin, Dr. Wenjuan Wang, and Dr. Brice Hanberry. All the members of my family have also played a large role in my success and I thank them all for their love, support, and encouragement. Lastly, I would like to thank my wife, Shannon Marie Bobryk, the most important person in my life: *Cogli l'attimo*.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	ii
LIST OF FIGURES .....	vi
LIST OF TABLES .....	viii
ABSTRACT.....	ix
CHAPTER 1: INTRODUCTION.....	1
1.1 IMPETUS.....	1
1.2 REVIEW .....	3
1.3 VIEWPOINTS .....	6
1.4 JUSTIFICATION.....	6
1.5 OBJECTIVE.....	7
CHAPTER 2: Estimating aboveground forest biomass potential along the Missouri River corridor, Missouri USA .....	9
2.1 ABSTRACT.....	9
2.2 INTRODUCTION .....	11
2.3 MATERIALS AND METHODS.....	15
2.3.1 Study Area .....	15
2.3.2 Aboveground Forest Biomass .....	17
2.3.3 Landscape Variables.....	19
2.3.4 Biomass Estimation .....	21
2.3.5 Random Forest.....	22
2.4 RESULTS .....	23
2.4.1 AFB Estimation .....	23
2.4.2 AFB Differences.....	24
2.4.3 Variable Importance .....	24
2.5 DISCUSSION .....	25
2.5.1 Aboveground Forest Biomass Estimates.....	25
2.5.2 Variable Importance .....	27
2.6 CONCLUSION.....	29

CHAPTER 3: Determining environmental controls driving discrepancies between historic and current aboveground forest biomass along the Missouri River corridor, Missouri, USA.....	48
3.1 ABSTRACT.....	48
3.2 INTRODUCTION .....	50
3.3 METHODS .....	54
3.3.1 Study Area .....	54
3.3.2 Modeling Framework .....	56
3.3.3 Explanatory Variables .....	57
3.3.4 Historic Aboveground Biomass.....	58
3.3.5 Current Aboveground Biomass .....	59
3.4 RESULTS AND DISCUSSION .....	59
3.4.1 Aboveground Forest Biomass Estimates .....	59
3.4.2 Variable Importance .....	60
3.4.3 Percent Changes .....	61
3.4.4 Abiotic controls .....	62
3.5 CONCLUSION.....	64
CHAPTER 4: Geographic analysis of production suitability for available aboveground forest biomass within the Missouri River Corridor, Missouri, USA .....	78
4.1 ABSTRACT.....	78
4.2 INTRODUCTION .....	80
4.3 METHODS .....	84
4.3.1 Study Area .....	84
4.3.2 Suitability Criteria .....	85
4.3.2.1 Aboveground Forest Biomass .....	85
4.3.2.2 Landscape Variables .....	87
4.3.3 GIS Processing .....	89
4.4 RESULTS .....	92
4.4.1 Potential Biomass .....	92
4.4.2 Available Biomass .....	94
4.4.3 Suitability.....	94

4.5	DISCUSSION .....	94
4.5.1	Aboveground Forest Biomass Potential .....	95
4.5.2	Aboveground Forest Biomass Constraints .....	96
4.5.2	Aboveground Forest Biomass Suitability.....	97
4.6	CONCLUSION.....	99
	CHAPTER 5: CONCLUSION .....	116
5.1	OVERVIEW .....	116
5.3	FUTURE DIRECTIONS .....	119
	REFERENCES .....	123
	VITA.....	138

## LIST OF FIGURES

Figure 2 - 1: The Missouri River Corridor by a 120 km (75 mile) buffer centered from the Missouri River. ....	31
Figure 2 - 2: Ecological subsections within the Missouri River Corridor. ....	32
Figure 3 - 1: The Missouri River corridor located in Missouri, USA. The corridor was constructed based on a 75 mile buffer centered on the Missouri River. The entire study region is approximately 106,000 km <sup>2</sup> and contains 69 counties dissected by the Missouri River.....	66
Figure 3 - 2: Observed and predicted aboveground forest biomass model estimates from the Historic RF model where the red line indicates model fit while the blue line indicates the 1:1 slope. ....	67
Figure 3 - 3: Observed and predicted aboveground forest biomass model estimates from the Current RF model where the red line indicates model fit while the blue line indicates the 1:1 slope. ....	68
Figure 3 - 4: Estimated total aboveground forest biomass (Mg) for each forested pixel (30m) for (A) Historic and (B) Current. ....	69
Figure 3 - 5: Variable importance plots for current and historic aboveground forest biomass estimates generated from Random a Forest regression procedure. ....	70
Figure 3 - 6: Pixel-level percent change values between historic and current aboveground forest biomass in the MRC. ....	71
Figure 3 - 7: Average percent change between historic and current AFB estimates summarized by ecological subsection.....	72
Figure 3 - 8: Summarized percent changes with standard errors between historic and current aboveground forest biomass estimates generated from a Random Forest procedure.....	73
Figure 4 - 1: The Missouri River corridor is approximately 106,000 km <sup>2</sup> covers 21 counties and is divided into 21 ecological subsections.....	100
Figure 4 - 2: Flow chart illustrating processes to obtain suitability index.....	101
Figure 4 - 3: (A) Percent changes between historic and current aboveground forest biomass (AFB) estimated using a Random Forest procedure at the pixel scale. (B) Average AFB summarized by ecological subsection. ....	102
Figure 4 - 4: Variable importance plot for estimates of Current and Historic aboveground forest biomass (AFB) generated from a Random Forest procedure. ....	103
Figure 4 - 5: The biomass mask is used to constrain total aboveground forest biomass (AFB) estimated by the RandomForest model and identify areas of available AFB. The mask was generated based on a kernel density estimate of forest transition values that were ranked from 0 – 1 (i.e. after fuzzy membership algorithm was applied to values of percent forest transition. Locations that were $\geq 0.95$ based on the fuzzy membership	



analysis were masked to represent available areas. All other areas not covered by the mask were considered not available and left out of the suitability analysis. ....	104
Figure 4 - 6: Percent changes in Forest (top-bar), Agriculture (middle bar), and ‘Other’ (bottom bar) between 2001 and 2006. ....	105
Figure 4 - 7: Total current (Unconstrained – Blue) and total available (Constrained – Red) aboveground forest biomass summarized by each ecological subsection with standard errors. ....	106
Figure 4 - 8: Fuzzy membership for (A) Euclidean distance, (B) percent change, (C) forest transition, and (D) current AFB quantity covering all forested pixels (30m <sup>2</sup> ). ....	107
Figure 4 - 9: Suitability index per ecological subsection where the horizontal bar indicates the value used (0.9) to categorized subsections with the greatest suitability (Loess Hills, Mississippi, and Missouri River Hills). ....	108
Figure 4 - 10: Available aboveground forest biomass suitability index based on Fuzzy Logic processes in ArcGIS. ....	109
Figure 4 - 11: Constrained (available) aboveground forest biomass suitability categories stratified by ecological subsection. ....	110
Figure 5 - 1: Observed versus predicted AFB from the post-hoc RF model. Red line indicates model fit and blue dashed line illustrates the 1:1 line. ....	121
Figure 5 - 2: Variable importance chart from the RF model. ....	122

## LIST OF TABLES

Table 2 - 1: Ecological subsection descriptive information.....	42
Table 2 - 2: Description of environmental variables used to estimate aboveground forest biomass in a Random Forest modeling procedure.....	43
Table 2 - 3: Remote sensing image descriptions. ....	44
Table 2 - 4: Average aboveground forest biomass (AFB) estimates for the observed data from Forest Inventory and Analysis (FIA), Random Forest (RF) estimates of actual AFB (Actual), and RF estimates of theoretical AFB (Theoretical) summarized by ecological subsection.....	45
Table 2 - 5: Random Forest model statistical summaries between current (FIA) and theoretical (Upper-bound FIA) aboveground forest biomass estimates. ....	46
Table 2 - 6: Paired two sample t-Test for means of average estimates of AFB summarized by ecological subsection between FIA and the Current RF Model. ....	47
Table 3 - 1: Summarization of GLO and RF AFB estimates.....	74
Table 3 - 2: Missouri River Corridor ecological subsection descriptions. ....	75
Table 3 - 3: Statistical summaries for RF Current and RF Historic estimates of AFB.....	76
Table 3 - 4: Paired two sample t-Test for means of average estimates of AFB summarized by ecological subsection between FIA and RF.....	77
Table 4 - 1: Description of environmental variables used to estimate aboveground forest biomass in a Random Forest modeling procedure.....	112
Table 4 - 2: Summary of average aboveground forest biomass estimates per ecological subsection between Potential and Available AFB values.....	113
Table 4 - 3: Area and total aboveground forest biomass summarized by ecological subsection.....	114

## **ABSTRACT**

This research presents the culmination of statistical, landscape, and geospatial analyses that examine the geographic dynamics of aboveground forest biomass (AFB) within the Missouri River corridor, Missouri USA. The Missouri River corridor is a region specifically within Missouri that encompasses 106,000 km<sup>2</sup>, and is regarded as a processing region for improving the viability of Missouri's biomass/biofuel industry. Current and historic forest inventory data coupled with remote sensing, edaphic, physiographic, and climate variables were integrated into an ensemble regression tree method, Random Forest (RF), to estimate AFB, determine external driving forces of AFB, and visualize geographic locations where the greatest deviations exist between current and historic AFB values. The applicability of constructing a hybrid modeling framework using RF was initially tested in Chapter 2 by estimating current (observed data derived from Forest Inventory and Analysis) and theoretical (based on 20% of AFB found within Missouri) AFB, and calculating the percent change to determine percent changes in AFB across the landscape. The third chapter extended the RF modeling procedure to include historical information derived from General Land Office (GLO) data to estimate a baseline measure of AFB. Current AFB was again estimated and then compared to historic values where an additional synthesis was performed to investigate the top predictors of AFB. The fourth chapter examined a fuzzy logic approach for developing a suitability index based on available AFB. Available AFB was determined by applying physical constraints onto estimated AFB from the RF model, which included forest transitions and distance to rivers. The model results failed to reject our null hypothesis that there were no differences between observed and predicted AFB, and

model accuracy was very low for all AFB estimate. Results from these investigations indicated that 1) the greatest potential for increasing AFB may be along the floodplains of the Missouri and Mississippi Rivers, 2) areas that have the greatest potential to reach historic baseline levels of AFB also occur along the Missouri and Mississippi Rivers alluvial plains, 3) climate variables were the most important factors for estimating AFB, and 4) a coarse-scale decision process using multi-criteria, fuzzy logic exemplified the Mississippi and Missouri River floodplains, and associated Missouri River Alluvial Plains, as the most suitable locations for AFB production.

# CHAPTER 1: INTRODUCTION

*A small grove massacred to the last ash,  
An oak with heart-rot, give away the show:  
This great society is going smash....  
... A culture is no better than its woods.*

– *W.H. Auden*

## 1.1 IMPETUS

The use of renewable energy resources is becoming increasingly necessary to achieve the ambitious bioenergy targets set by the United States Energy Independence and Security Act (EISA), which mandated an increase in biofuel production from the current 45 billion liters to 136 billion liters per year by 2022 (EISA 2007). Rapidly increasing energy demands, however, are becoming a serious challenge for achieving a more sustainable and secure energy future. World-wide energy consumption is expected to rise 40% by 2030 with a paralleled demand for fossil fuels estimated to rise approximately 77% (Bardhan and Jose 2012). The growth of different renewable energy sources such as wind, solar, and geothermal have varied in recent years; however, the development and advancement of wood-based bioenergy remains one of the most promising solutions for confronting the challenges of future energy demands (Zhuang et al. 2011). Total renewable energy production reached a new historic level of domestic energy supply in 2013 where, according to the U.S. Energy Information Administration's Monthly Energy Review (DOE/EIA 2014), approximately 2.1 quadrillion British thermal units (QBtu) of renewable energy was derived from wood.

Aboveground forest biomass (AFB), specifically of primary residues from trees or tree parts, is the most common form of feedstock available for use as bioenergy (Thiffault et al. 2014). Wood has historically been used as a source for energy, accounting for approximately 91% of domestic consumption in the mid 1800's (Aguilar et al. 2011), and remains a major source of energy for millions of people around the world. More than 200 million people, primarily living in poverty, rely directly on forests for energy, shelter, and maintaining their livelihoods (Berndes et al. 2003; Pan et al. 2013). In richer parts of the world, there is an increase of interest in using AFB as a major source for abating fossil fuel reliance and subsequently, meeting regional and global goals set by governmental policies (e.g. EISA). Increasing the use AFB as a renewable source of energy is attractive because it is a resource that can be locally or regionally exploited, which is especially important for regions with highly productive landscapes, such as along major river corridors.

River floodplains have the capacity to generate large-scale quantities of biomass to sustain regional biofuel production systems (Bardhan and Jose 2012). Specifically, the Missouri River corridor has large acreages of both marginal (less productive areas not traditionally used for foodgrain or fiber production) and productive lands, and is considered to have the greatest production potential in the U.S. (USDOE 2011). Previous studies have found that marginal lands along major river systems would be suitable to produce a substantial amount of AFB feedstock (Stanturf et al. 1998; Bardhan and Jose 2012; Gopalakrishnan et al. 2011). Bardhan and Jose (2012) suggest that the Missouri river has approximately 40 million ha suitable for AFB production.

In addition to being a source of energy, AFB provides many ecosystem services that include food, fiber, timber, medicine, clean water, aesthetics, spiritual values, and promotes carbon sequestration (Pan et al. 2013; Zheng et al. 2010). The increasing demand for producing energy from AFB has consequently triggered concerns about the sustainability of production and potential perturbations on ecosystems from over-exploitation (Thiffault et al. 2014). A great deal of attention is now being placed on understanding where AFB is located and what potential it has to support future energy demands. Large landscape-scale assessments are needed for identifying suitable areas where AFB may be the best source of biomass feedstock, which could ultimately help abate conventional fossil fuel energy demands without deprecating ecosystem services.

## **1.2 REVIEW**

The U.S. is pursuing several options to generate renewable energy sources from biomass in an effort to reduce dependency on imported fossil fuels and mitigate greenhouse gas emissions (Zhang et al. 2011). AFB is an attractive energy source because it is abundant, cheap, and has additional environmental qualities such as nutrient cycling and biodiversity. Additionally, previous research has identified AFB as a key indicator for promoting diversity of carbon stocks and is considered an important factor in assessing consequences of global changes in tree species' distributions (Zimmermann et al. 2007; Potter et al. 2008; Keith et al. 2009). Pan et al. (2013) provide an in-depth review of the structure, distribution, and biomass characteristics of the world's forests; concluding, simply, that biomass densities have increased over time, becoming more dynamic, and future, technologically advanced research is needed to monitor and manage forested ecosystems of the world. Rapidly changing forest systems require advanced means of investigating and monitoring changes in biomass through time. Improved monitoring of

biomass dynamics can help managers, researchers, and industry leaders to understand the causes of such changes, which will also be critical for designing effective programs and policies that structure sustainable AFB management and enhance important ecosystem services of forests.

Different approaches to estimate AFB have been developed and primarily focus on three main utilities: 1) field measurements, 2) remote sensing, and 3) methods for statistical analysis (Gallaun et al. 2010; Anaya et al. 2009; Baccini et al. 2004; Blackard et al. 2008). Remote sensing techniques have become prevalent in estimating AFB in recent years at various scales (Gasparri et al. 2010; Powell et al. 2010; Zheng et al. 2004; Anaya et al. 2009; Gallaun et al. 2010; Gleason and Im 2011). For example, strong correlations were found between spectral band information and biomass indices calculated from multi-temporal remote sensors (Gasparri et al. 2010). Powell et al. (2010) demonstrated the utility of an empirical modeling process that used satellite imagery in combination with forest inventory data to estimate and quantify trends in biomass dynamics. An early review of the potential and challenges of remote sensing in producing biomass estimations was provided by Lu (2005), which discussed how complex forest stand structures and environmental conditions cause great difficulty in estimating performance and additional analyses should be used in conjunction with remote sensing. Previous research has also shown that there is a lack of applicability in estimating AFB from each utility alone. For example, Lu et al. (2012) highlight four critical issues that are most often associated with remote sensing-based biomass estimation methods that included, an insufficient number of available sample plots, myriad metrics for biomass modeling, differences in allometric equations, and the use of suitable methods to conduct



uncertainty analysis of the estimates (Lu et al. 2012). Hybrid models that use combinations of data sources have been shown to improve model predictions (Baccini et al. 2004). Constructing hybrid modeling approaches for estimating AFB have also been shown to improve accuracy and promote greater ability to understand geographic dynamics of AFB at various spatio-temporal scales (Wulder et al. 2008; Blackard et al. 2008; Zheng et al. 2008; Powell et al. 2010).

Predicting biomass by relating field observations to environmental variables can significantly increase the predictive power of models (Guisan and Thuiller 2005); however, identification of suitable environmental variables for estimating AFB is difficult because there are myriad potential options. Guisan and Thuiller (2005) suggest that optimal characterization of environmental variables should reflect three main types of influences: 1) limiting factors, 2) disturbances, and 3) resources. Zimmermann et al. (2007) used distributional gradients to characterize and predict future distributions on the basis of ecological niches characterized based on limiting, environmental factors and available resources. Additionally, distribution models were developed by Kearney and Porter (2009) by characterizing biophysical interactions between species' functional traits and the physical environment in order to delineate species fundamental niches. A key factor specific to this dissertation research is the development of indicators for estimating AFB. There are many challenges in defining which variables may best represent potential suitability of AFB and determining how the membership of such variables to AFB potential may also be used to understand geographic availability within the Missouri River corridor.

### **1.3 VIEWPOINTS**

Increasing wood-based energy production is a contentious topic because of the duality between offsetting fossil fuels and threatening food security or increasing carbon footprints. For instance, alternative fuel mandates established through federal policies can have local implications where producing biofuels domestically may take land away from food production. Additionally, over the last 300 years the world economy has experienced an increasing separation between the locations of biomass production and actual consumption of biomass energy (Lambin and Meyfroidt 2010). The U.S., for example, has been a net importer of wood for approximately the last 90 years (Shifley 2006). Consequently, importing large quantities of AFB over long distances could have a large impact on CO<sub>2</sub> emissions into the atmosphere (Liu et al. 2013); therefore, it is becoming a critical issue to find a balance between growth, harvest, and consumption of AFB to maintain sustainability for future generations (Shifley 2006). Bardham and Jose (2012) suggest that second-generation feedstock production, specifically within river floodplains and marginal lands, have the greatest potential for increasing large landscape use of AFB feedstocks, especially along the Missouri and Mississippi River corridors.

### **1.4 JUSTIFICATION**

In developing a strategy for managing AFB, it is necessary to acknowledge that forest ecosystems have limits for exploitation. Designing effective programs and policies that structure biomass extraction requires extensive evaluation and prioritization of suitable areas, especially at large landscape scales. Research on AFB has often been centered on modeling the physiologic characteristics of species and their ability to sequester carbon, thus offsetting fossil fuel carbon emissions (Cannell 2003). Additional research is needed to help support sequestration research, especially at broad, geographic scales. For

instance, evaluating the suitability of the landscape to maintain or increase existing AFB is important for supporting and guiding decision-making processes and to help prepare for technological advancement of lignocellulosic biofuel production. Digital maps that portray dynamics of AFB characteristics are emerging as critical tools for addressing the increasing need for carbon sequestration management, land-use monitoring, and general forestry applications (Liang and Zhou 2014). Several studies have used biomass density as a valuable baseline for predicting productivity or understanding changing distributions of species (Jose and Gillespie 1997, 1998; Zheng et al. 2004), especially after major deforestation from EuroAmerican settlement; albeit, standard, geographical methods are still needed to develop detailed spatial and environmental information for use in large landscape, ecosystem management decisions (He et al. 1998). Historical inventories and field-based studies, conducted prior to major land-use changes, may be used as a baseline measure to provide important information on AFB recovery and give insight into how much potential there is to reach or surpass historic conditions.

## **1.5 OBJECTIVE**

The objective of this research was to determine geographic discrepancies in aboveground forest biomass (AFB) with respect to historic and current conditions, and identify suitable areas where increased AFB production could be possible within the Missouri River corridor. The following chapters of this dissertation involve a comprehensive assessment of the geographic dynamics of AFB using statistical and geospatial functions.

The applicability of constructing a hybrid modeling framework using an ensemble regression tree method, Random Forest (RF), was tested in Chapter 2 for estimating current AFB quantities where observed information was obtained from the United States Forest Service Forest Inventory and Analysis (FIA) program. Aboveground forest

biomass was estimated using a series of remote sensing, edaphic, climatic, and physiographic predictor variables in the RF model. An additional analysis was performed based on the upper bounds of observed AFB for purposes of comparison, to illustrate theoretical potential of AFB, and better understand the geographic arrangement of areas with greatest differences in AFB across the landscape.

The third chapter extended the RF modeling procedure to include historical information derived from General Land Office (GLO) survey data (a type of tree inventory originally established as a cadastral survey in the nineteenth century), to estimate historical AFB. Percent changes between historical and current AFB estimates were calculated to understand where the greatest departures have occurred. An additional synthesis was performed to investigate the primary variables for estimating AFB. Finally, in the fourth chapter, geographic distributions of discrepancies and available quantities of AFB were refined by building physical constraints within a geographic information system (GIS). The resulting maps illustrated AFB availability based on physical constraints that prohibit AFB extraction under the premise of conservative exploitation for bioenergy purposes. A suitability map was then created that identified regional areas illustrating the greatest capacity for AFB production and likelihood for conservative exploitation to support bioenergy industry.

## **CHAPTER 2: Estimating aboveground forest biomass potential along the Missouri River corridor, Missouri USA**

*Existing scientific concepts cover always only a very limited part of reality, and the other part that has not yet been understood is infinite.*

– Werner Heisenberg

### **2.1 ABSTRACT**

This study used a Random Forest (RF) approach to estimate aboveground forest biomass (AFB) at 30-meter resolution within the Missouri River corridor, Missouri USA. Initial estimates were made for current conditions using data collected by the U.S. Forest Service Forest Inventory and Analysis (FIA) program. An additional modeling scenario was employed based on an upper threshold of FIA-derived AFB values to generate new estimates that reflected high AFB values potentially representative of theoretically optimal growing conditions. Percent changes between current and theoretical estimates of AFB were calculated and then summarized by ecological subsection (specific areas within Missouri that are delineated by similar ecological characteristics) to understand where current AFB had changed the most across the landscape. The RF procedure introduced forest inventory, remote sensing, and geospatial landscape datasets into an ensemble regression tree algorithm to estimate AFB and determine variable importance. Model validation was based on a comparison between RF estimates of AFB and estimates derived from FIA. The RF estimates for both current and theoretical models had very low coefficients of determination ( $R^2=0.07$  and  $0.19$ , respectively) and the predictor variables exhibited a poor statistical relationship with AFB; however, when the current AFB estimates were summarized by ecological subsection, there was no significant difference from total averages of FIA-derived estimates (FIA =  $94 \text{ Mg ha}^{-1}$

and  $RF = 97 \text{ Mg ha}^{-1}$ ,  $p = 0.06$ ). Total average AFB for theoretical AFB estimates was  $122 \text{ Mg ha}^{-1}$ . Remote sensing variables were determined by RF to be the best predictors of current AFB, while elevation, pH, and slope were the top 3 estimators in the Theoretical AFB model. The percent change between the theoretical and current AFB estimates was derived and summarized again by ecological subsection where the greatest changes in AFB were found along the floodplains of the Missouri and Mississippi Rivers. This was a first attempt to estimate AFB within the Missouri River corridor using a RF procedure, which performed poorly and resulted in inaccurate estimates. The average AFB estimates in the current model summarized by ecological subsection were not statistically different from FIA estimate, which suggests that the 30-meter resolution scale was not appropriate from estimating AFB with the given predictor variables. The data used better reflects conditions at much broader scales, such as the ecological subsection scale, and could provide more insight into the broader implications of understanding where there may be the greatest potential to increase current AFB in Missouri forests.

## 2.2 INTRODUCTION

The United States Energy Independence and Security Act (EISA 2007) has mandated the transition of the country towards a more sustainable energy future, and biomass-based alternative energy sources are being investigated as major contributors to meeting the mandated increases in bioenergy production (Bardhan and Jose 2012). Aboveground forest biomass (AFB) in particular is recognized as one of the largest sources of feedstock to use for renewable energy because there is a reliable baseload available within the continental U.S. (O'Laughlin 2009).

Forest biomass makes up approximately 90% of all terrestrial vegetative biomass (Jenkins et al. 2001) and is continually growing (Shifley 2006; Stasko et al. 2011). The use of AFB as a feedstock for energy had reached 38 million dry tons per year (MDT yr<sup>-1</sup>) in 2011 and is projected to rise over 50%, nearing 72 MDT yr<sup>-1</sup> by 2017 (U.S. Department of Energy 2011). Advancing bioenergy system technologies coupled with pressure from increasing energy demands and consumption rates all suggest that AFB has potential to be a major contributor to the bioenergy industry and may help reach a sustainable energy future (Berndes et al. 2003). Total renewable energy production reached a new historic level for domestic supply in 2013 where, according to the U.S. Energy Information Administration Monthly Energy Review (April 2014), approximately 2.1 quadrillion British thermal units (QBTU) of renewable energy was derived from wood; which is approximately 4.1% of the total primary energy consumption in the U.S. (USDOE 2011).

Many investigations focus on estimating AFB for refining measures of carbon stocks or evaluating the sequestration potential of forested systems (Mani and

Parthasarathy 2007; Soratana et al. 2013; Zimmermann et al. 2007; Potter et al. 2008); however determining the potential of forested areas to meet increasing energy demands depends on the size and suitability of land dedicated to increasing AFB. Large acreages of land are necessary to make significant impacts on woody-fuel production and CO<sub>2</sub> sequestration (Righelato and Spracklen 2007), and estimating the potential to increase AFB is the first step to understanding where the best locations are to fully take advantage of existing AFB stocks (Shifley 2006).

Concerns are continually rising about how large-scale land conversions for bioenergy development may impact food production or lead to adverse environmental externalities (Lovett et al. 2009; Havlík et al. 2011). The identification of suitable land area to potentially support a substantial feedstock among multiple competing land-use interests is an important avenue of research that employs a multitude of modeling schemes. A large portion of recent research efforts have focused on the estimation of biomass quantities specifically from energy crop plantations or monocultures (Hall et al. 2011; Puhr and Donoghue 2000; Graham and Walsh 1999), and have often neglected estimations of standing biomass from heterogeneous, forested ecosystems. Evaluating biomass production at large geographic scales, as well as, determining important variables that drive biomass potential at large scales will help develop a critical understanding of current biomass supply characteristics. This is especially important within floodplains or marginal lands along major river corridors where conditions are conducive for high biomass production but subject to frequent disturbances, such as along the Missouri River (Bardhan and Jose 2012).



In addition to a renewable energy source, AFB is a major component to the global carbon cycle and has been used as a proxy for determining the health and productivity of forested ecosystems (Brown et al. 1997; Cannell 2003; Stasko et al. 2011; Zheng et al. 2010). An additional objective within the U.S biofuel policy is to increase the use of AFB to offset greenhouse gas emissions (EISA 2007; Bardhan and Jose 2012) and research that focuses on the geographic dynamics of AFB is a critical step towards understanding the potential for particular regions to lead the country towards meeting the mandated increases. The Missouri River corridor (MRC) and associated central hardwood forests within Missouri are exemplary ecosystems for investigating AFB dynamics because this region is uniquely positioned within the U.S. to support a thriving biofuel industry due to availability to large acreages of both marginal and productive lands that have a variety of soil types, favorable climate, and growing seasons, as well as, a wide range of competing land-use interests (agriculture, recreation, industry). In addition to favorable landscape characteristics, this region is considered to have a high biomass production potential because the forested systems along the MRC are still slowly recovering from widespread clearing after European settlement in the 19<sup>th</sup> century. Variations in forest recovery rates from land-use changes have created regional disparities in the potential supplies of AFB (Nigh and Schroeder 2002; Becker et al. 2009) and evaluation of current geographic distributions may benefit management decisions that rely on landscape-level analyses. Estimating and spatially determining where AFB exists at a large landscape begins with developing a model that is capable of estimating AFB at larger scales (Righelato and Spracklen 2007).

Approaches for estimating AFB primarily focus on three main utilities: 1) field measurements, 2) remote sensing, and 3) landscape modeling; however, there is a lack in accuracy when estimating AFB from each utility alone (Powell et al. 2010; Potter et al. 2008; Lu 2006; Zimmermann et al. 2007). For instance, forest inventory typically lacks the spatial resolution necessary to produce reliable, fine-grained information and remains insufficient as a single estimate of AFB over large landscapes (Zimmermann et al. 2007). Estimates of AFB from remotely sensed data contain fine resolution information but often fail to match estimates statistically derived from ground inventory (Zheng et al. 2008); whereas hybrid models that use combinations of data sources have been shown to improve AFB estimates (Baccini et al. 2004; Wulder et al. 2008). Guisan and Thuiller (2005) suggested that optimal characterization of environmental variables should reflect three main types of influences: limiting factors, disturbances, and resources (Guisan and Thuiller 2005). The statistical models used to create AFB estimates at landscape scales also vary in type, performance, and ease of use. Specifically, Random Forest (RF), an ensemble regression tree method from the family of categorical and regression tree (CART) analyses, has gained much attention as a robust method for data mining and is capable of producing accurate predictions through its randomizing process. Prasad et al. (2004) examined the predictive performance of RF against three other statistical models, Regression Tree Analysis (RTA), Bagging Trees (BT), and Multivariate Adaptive Regression Splines (MARS) based on the differences between model performances when variables are randomly permuted throughout the modeling process. The results of the study reported RF to have superior predictive capability and interpretability of variable importance (Prasad et al. 2004).

Anthropogenic disturbances, such as those from European settlement, have had large impacts on forested ecosystems, leaving legacy effects that continue to influence current AFB conditions (Liu et al. 2007). Estimations of AFB potential using historical information can be used as a baseline to compare current AFB estimates and generate a chronological perspective on the dynamics of AFB between two periods of time (Rhemtulla et al. 2008). The basis for understanding the geographic distributions of the availability of biomass for energy production relies on two very important types of potential: 1) what would exist without major anthropogenic activities (theoretical), and 2) what is actually obtainable (available).

The objective of this study was to test the applicability of RF for estimating AFB within the Missouri River corridor in Missouri, USA and to determine geographical distributions of biomass potential at regional scales. The first specific objective was to devise a hybrid modeling approach that introduced forest inventory, remote sensing, and geospatial landscape datasets into a RF model to estimate current AFB, determine variable importance, and map the available potential AFB. The second specific objective was to devise an approach that would best represent theoretical AFB characteristics that reflect no major anthropogenic disturbances, such as those from European settlement, and represent optimal growing conditions. The final objective was to compare the theoretical and available AFB estimates and assess the geographic locations of potential to increase AFB production.

## **2.3 MATERIALS AND METHODS**

### **2.3.1 Study Area**

The Missouri River corridor (MRC) is approximately 106,000 km<sup>2</sup>, located within Missouri, USA, and includes the main channel of the Missouri River, associated

floodplain, alluvial terraces, and includes all rolling uplands (Figure 2-1). The MRC represents a shifting mosaic of portions of the hardwood forest that can be attributed to changes in land use over time (Nigh and Schroeder 2002). The associated geography consists of transitional ecotones comprised of two major landforms, Central Dissected Till Plain and Ozark Highland. The two major landforms are separated into 21 subsections that represent similar ecological characteristics and landforms (Figure 2-2; Table 2-1).

Land cover consists of a mosaic of cropland, pasture, and secondary growth woodlands with extensive past grazing influences (Nigh and Schroeder 2002). Mixed bottomland hardwood forests with oak (*Quercus spp.*), hickory (*Carya spp.*), and maple (*Acer spp.*) are limited to areas less subject to flooding, whereas near the riverfront, fluctuating water levels have created aggregations of vegetative communities that range from sandbars to stands of cottonwood (*Populus spp.*), willow (*Salix spp.*), maple (*Acer spp.*), and elm (*Ulmus spp.*). Upslope from the river valley are bluff forests dominated by mesophytic species such as maple and basswood (*Tilia americana*) (Kucera and McDermott 1955; Pallardy et al. 1988) while other land-cover types in the region include oak savannas and woodlands, mixed-hardwood forests, and occasional prairie and glade openings (Nigh and Schroeder 2002).

The premise of the MRC is to define a potential processing region that can be used to guide the biofuel industry. Biomass feedstock collection and transportation are integral aspects of biomass use (Kumar et al. 2006), and understanding the viability of resources and potential transportation avenues are crucial components in achieving sustainable AFB exploitation. For instance, the per-ton transport costs of biomass are

primarily calculated on the basis of travel distances and fuel prices where the transportation costs are directly proportional to the distances that biomass sources are hauled from towards potential processing facilities (J. Leboireiro, A.K. Hilaly 2011). To maintain sustainable bioenergy production and delivery systems, it is advantageous to minimize distances between source and facility. For example Hacetoglu et al. (2011) conducted a feasibility study in the Great Lakes regions and estimated that approximately 100 km<sup>2</sup> of forested land area may harbor a sufficient amount of material in the upper Great Lakes region to help displace fossil material for energy production and capable of producing lignocellulosic biomass with minimal adverse impacts on food and fiber production (Hacetoglu et al. 2011). We defined a processing region specific to Missouri based on a 120 km buffer centered from each side of the Missouri River. This buffer potentially represents a feasible distance to create a positive effect on biomass financial competitiveness and signifies the most economic transportation distance that necessary to make the use of standing biomass for bioenergy purposes economically viable.

### **2.3.2 Aboveground Forest Biomass**

We define aboveground forest biomass (AFB) as the total amount of living organic matter in trees present above the soil surface with a diameter at breast height (DBH) greater than 2.5 cm (1 inch). This definition is adopted by the United States Forest Service (USFS) and includes the minimum DBH observed by the Forest Inventory and Analysis (FIA) program. The FIA program of the USFS produces comprehensive data on the status, changes, and trends in the conditions of all forest ecosystems within the U.S. (Woudenberg et al. 2010). The FIA program provides plot-level information on 5-year cycles that includes, but is not limited to, species, size, growth, mortality, site index,

carbon, and biomass. FIA information is designed for use in regional monitoring of forest lands and is suitable for landscape level studies (Moser et al. 2010).

The spatial locations of FIA plots are confidential and data acquisition was required to be coordinated with personnel from FIA Spatial Data Services, and a confidential agreement form signed. The latest cycle of FIA inventory information (2006-2011) was requested from FIA personnel. To further maintain FIA plot spatial discrepancy, FIA personnel also requested that no more than 20 variables could be used to construct a model for estimating AFB. We gathered all geospatial layers (explanatory variables used in the RF model to estimate AFB) and transferred them to multiple digital versatile discs (DVD) to send via U.S. postal service to the FIA main headquarters (Northern Research Station, Newtown Square, PA). An overlay analysis was performed by FIA where all plots and associated AFB values were attached to the explanatory variable values and a tabulated document with all AFB values was sent back via email.

We used the total number of AFB values calculated by FIA personnel as the observed dataset to train and validate the actual RF model. We used a Student's T-test for means ( $\alpha = 0.05$ ) to validate the actual RF model with FIA. In addition to extracting AFB values for all plots, we requested that FIA also summarize the top 20% of the greatest AFB values from each ecological subsection. We used the top 20% of the greatest AFB as a theoretical baseline to simulate elevated biomass conditions representative of historic AFB levels prior to European settlement (circa 1800s). We made the assumption that AFB levels were historically greater than currently based on previous literature that suggested activities from European settlers severely modified bottomland and floodplain vegetation where up to 95% of forests were eliminated along the Missouri River basin

(He et al. 2007) from river channelization, forest clearing, and urban sprawl (Moser et al. 2006; Rhemtulla et al. 2008; Hanberry, Kabrick, et al. 2012). Thus, the top 20% of the greatest AFB values was deemed a reasonable threshold to use as the theoretical potential to reflect higher biomass values than are observed currently. We also only investigated total AFB (not focusing on specific species or functional groups) and treated all AFB plots as static in time; not including any possible influences from disturbance, age structure, or diameter distributions of trees. This information was not available and could not be implemented into this modeling scheme.

### **2.3.3 Landscape Variables**

Twenty variables that reflected 4 categories: 1) remote sensing (vegetation), 2) soil, 3) terrain, and 4) climate, were used to estimate AFB within the MRC (Table 2-2). Five remote sensing variables (Normalized Difference Vegetative Index, Normalized Difference Water Index, Tasseled-Cap Brightness, Tasseled-cap Greenness, and Tasseled-cap Wetness) were derived from a composite of 11 30-meter resolution Landsat Thematic Mapper (TM) satellite images (Table 2-3). We acquired the images from the Multi-Resolution Land Characteristics Consortium (<http://www.mrlc.gov/>, accessed May 2011), which provided at-sensor reflectance data. Normalized Difference Vegetation Index (NDVI) is a popular band ratio often used for determining biomass distributions because it has been shown to have a good relationship with vegetation characteristics indicative of biomass, such as leaf area and chlorophyll content (Gasparri et al. 2010). NDVI was calculated based on the combination of red and near-infrared bands according to the following equation:

$$NDVI = \left( \frac{NIR-R}{NIR+R} \right)$$

eq. 1

where NDVI is the ratio between the difference (numerator) and summation (denominator) of near-infrared (*NIR*) and red (*R*) spectral bands. The Normalized Difference Water Index (NDWI) was also used because it is a good indicator of healthy vegetation and is based on the following equation:

$$NDWI = \left( \frac{G - NIR}{G + NIR} \right)$$

eq. 2

where NDWI is the ratio between the difference (numerator) and summation (denominator) of near-infrared (*NIR*) and Green (*G*) spectral bands, which are within the water absorption band centered at 1200 nm (Gao 1996). A tasseled-cap transformation of the Landsat reflectance values was performed to extract 3 additional variables: brightness, greenness, and wetness (TCB, TCG, and TCW, respectively), where recently disturbed forest land exhibit high values of TCB, low TCG, and low TCW in relation to an undisturbed forested area (Healey et al. 2005). All remote sensing transformations and index calculations were performed in ERDAS IMAGINE® 2011.

We used annual mean temperature and precipitation seasonality (i.e. seasonality as measured by the coefficient of variation in precipitation) (BIO\_01 and BIO\_15, respectively) as climate variables to estimate AFB, which were acquired from the WorldClim database. The Worldclim database provides freely available global climate data for spatial modeling ([www.worldclim.com](http://www.worldclim.com), accessed May 2011). We derived 6 topographic variables from a 30-meter resolution Digital Elevation Model (DEM), which



is a cartographic dataset of elevation values. The DEM was obtained from Missouri Spatial Data Information Services where, in addition to elevation, the following metrics were derived in a Geographic Information System (GIS): aspect, slope, topographic wetness, ruggedness, and topographic position. We also extracted 7 soil variables that were representative of the top 20 cm of the soil surface (sand, silt, clay, coarse fragment, pH, hydraulic conductivity, and organic matter). These variables were obtained from the United States Department of Agriculture Natural Resource Conservation Service's Soil Data Viewer v6.0 (NRCS 2011). The Soil Survey Geographic Database (SSURGO) is part of the National Cooperative Soil Survey of Missouri where all soil information collected for the state ranges in scales from 1:12,000 to 1:63,360.

We resampled the soil information to fit a 30m resolution, and stratified by ecological subsections (Table 2-2). Ecological subsections are natural divisions within Missouri that are categorized based on similar environmental characteristics (soil, climate, topography, human-alteration) and integrate numerous physical and biological components that are used to establish boundaries for landscape-scale ecosystem studies (Nigh and Schroeder 2002).

#### **2.3.4 Biomass Estimation**

We used Random Forest (RF) (Breiman 2001) in R statistical software (Liaw and Wiener 2002; R Development Core Team 2011) as the modeling platform to predict AFB and determine variable importance. Parameterization of the model included the number of bootstrap samples to initiate ( $n_{tree} = 1500$ ), and the number of randomly chosen predictors to split along each regression tree ( $mtry = 4$ ). The coefficient of determination ( $R^2$ ) was used as the statistical measure for model fit and the mean absolute error (MAE) was used

to assess the size of the prediction error per model (Miehle et al. 2006). Ecological subsection boundaries were used to summarize, compare, and validate AFB estimates with observed FIA AFB data (Figure 2-2).

### **2.3.5 Random Forest**

Random Forest is a non-parametric, ensemble regression tree method that grows branches and nodes based on random subset sampling from a pool of variables (Boulanger et al. 2013). Regression trees are constructed using different bootstrap samples of the data, which changes how the regression trees are constructed per iteration. In standard classification and regression tree analyses, after a single regression has been completed, a node is created where a split is determined based on the best predictor. When trees are grown in RF, randomly selected subsets of variables are used to split at each node and the process is repeated using new bootstrap samples to create a large number of trees. The number of variables and iterations used in the model is decided by the user (*mtry* and *ntree* parameters in the RF code) and the final prediction is an average across all regression trees (Breiman 2001; Cutler et al. 2007; Prasad et al. 2004).

Variable importance is determined by the model using a permutation process of the original (training) and out-of-bag (OOB) data. OOB data are defined as those variables that were left out of the random selection process during the construction of each tree. When regression trees are generated, the prediction error for both the training and the OOB variables are recorded and the difference, normalized by the standard deviation of the differences, is averaged over all trees. The variables that create large changes in model predictions result in an increase in mean square error (MSE) and are considered to contribute more to the model (Prasad et al. 2004). The importance value indicates how much a variable increases the MSE when randomly permuted, which

means when a variable is less important, it will have limited contribution to the prediction and results in very small changes in MSE. On the other hand, variables that are important cause larger changes in predictions when that variable is removed; this results in larger MSE values. Variable importance output from RF can be used to determine which variables best represent the response variable and may be used to narrow down the number of variables sufficient to make parsimonious predictions.

RF has many advantages over other statistical models because it provides well-supported predictions and classifications (Prasad et al. 2004; Cutler et al. 2007; Boulanger et al. 2013), and enables the user to generate rapid, straightforward output. Additionally, RF is capable of handling large non-linear data sets or variables that suffer from collinearity (Prasad et al. 2004; Cutler et al. 2007; Boulanger et al. 2013).

## **2.4 RESULTS**

### **2.4.1 AFB Estimation**

The initial AFB estimates from the RF model was processed for every pixel across the study area. To acquire total quantities of AFB, we used the 2011 National Landcover Database (NLCD 2011) to select only forested pixels used for summarizing AFB. Total AFB estimated by FIA inventory data was  $3.0 \times 10^8$  Mg. There was a poor statistical relationship between observed FIA and estimated current AFB ( $R^2 = 0.07$ ). (Figure 2-3). There was also no statistical relationship between observed FIA AFB and theoretical AFB estimates ( $R^2 = 0.19$ ) (Figure 2-4). For the current RF model, a total of 1300 FIA plots was used for training where AFB was estimated at  $3.4 \times 10^8$  Mg (RMSE = 14.4 Mg ha<sup>-1</sup>, MAE = 11 Mg ha<sup>-1</sup>) (Table 2-5). The theoretical RF model used 168 FIA plots to train the model which estimated AFB at  $4.2 \times 10^8$  Mg (RMSE = 13 Mg ha<sup>-1</sup>, MAE = 10

Mg ha<sup>-1</sup>), respectively. We then quantified average AFB per ecological subsection where the total averages of AFB for the current, theoretical, and FIA models was 93 Mg ha<sup>-1</sup>, 121 Mg ha<sup>-1</sup>, and 94 Mg ha<sup>-1</sup>, respectively (Figures 2-5, 2-6, 2-7 and Table 2-5). There was no statistical difference between average AFB estimated by RF and that originally observed by FIA ( $p = 0.06$ ) (Table 2-6). There was a good statistical relationship between the summarized averages of FIA and RF AFB per ecological subsection ( $R^2 = 0.6$ ,  $N=21$ ) (Figure 2-8).

#### **2.4.2 AFB Differences**

To illustrate the geographic distribution of potential to increase AFB production, the percent change between theoretical and current AFB estimates were calculated and then summarized by ecological subsection where AFB differences ranged between 16 to 53 Mg ha<sup>-1</sup> (Figure 2-9). The greatest differences between model estimates were found along the Missouri River Alluvial Plains subsection and associated tributaries surrounding the Missouri River. The lowest differences between the current and theoretical model estimates of AFB were found within the Missouri Ozark region.

#### **2.4.3 Variable Importance**

The variable importance plot generated from the RF procedure revealed interesting patterns with regards to the contribution of specific variables in estimating AFB (Figure 2-10). For instance, the five remote sensing indices included in the model represented the most important variables for current model estimates of AFB where the tasseled-cap brightness (TCB) index was the best predictor followed by normalized difference water index (NDWI), tasseled-cap greenness (TCG), normalized difference vegetation index (NDVI), and tasseled-cap wetness (TCW). This was completely opposite to the

theoretical model estimates where the top predictors were elevation, pH, slope, and then temperature. All remote sensing indices contributed minimally to the model estimates of theoretical AFB and were placed near the bottom of the variable importance plot. Additionally, the second climate variable (precipitation) did not contribute to the theoretical model as much as was seen in the current AFB estimates. An additional 8 soil and physiographic variables separated the two climate variables in the theoretical model. The least important variable for both estimates was topographic wetness index (TWI).

## **2.5 DISCUSSION**

### **2.5.1 Aboveground Forest Biomass Estimates**

Missouri ecological subsections are landmasses that were delineated based on similar ecological characteristics and are used primarily for landscape-scale assessments. Landscape scale analyses are important for informing forest planning decisions because they provide big-picture information about ecosystem characteristics and landscape-scale patterns. Defining geographic areas that may be more important than others is a critical element in landscape planning processes, especially when multiple management objectives exist, such as along the Missouri River corridor. In the context of this study, Random Forest did not perform well in estimating current AFB at the pixel scale; however, a more informative relationship was observed when average AFB estimates were compared at the ecological subsection scales, where the total AFB averages between current RF and FIA were not statistically different. This may suggest that the predictor information may not be appropriate for such small-scale estimating; rather it may be more appropriate for estimating AFB at the subsection scale.

There were approximately 1400 FIA plots placed throughout the 106,000 km<sup>2</sup> study area, which were used to train both RF models (Current and Theoretical). FIA information is most often used for coarse-scale assessment and certainly was not sufficient for creating a relationship between AFB and associated predictor variables at the 30-meter resolution pixel scale. By calculating the averages of AFB estimates for a larger area, we were able to capture enough information from all AFB estimates found within a single subsection to produce a statistically relevant relationship between average AFB values. As a result, average AFB estimates at the subsection-scale has potential to be used to inform only ecosystem-management decisions that focus on landscape-scale activities within Missouri.

Theoretical AFB estimates from the RF model resulted in subjectively reasonable averages per ecological subsection. We constructed the theoretical model using training data that consisted of the top 20% of the greatest AFB values from FIA information but we were unable to know the geographic locations of the top 20% due to FIA confidentiality restrictions. We chose the top 20% of the greatest biomass values to simulate historic conditions to construct a baseline to compare current estimates with. This decision was based on the assumptions that historic biomass quantities were once much greater than what is observed today, which could only be validated using published information on ranges of old-growth biomass estimates. The averages calculated for theoretical AFB estimates per ecological subsection fell below published averages from more notably high biomass regions of the U.S. For instance, Johnson and Bell (1976) documented that old growth forest in Illinois reached quantities between 150-300 Mg ha<sup>-1</sup> (Johnson and Bell 1976) while the lower biomass fraction of old growth forests found in

the Pacific Northwest have been estimated to reach 400 Mg ha<sup>-1</sup> (Smithwick et al. 2002). The values reported from previous studies provided a range of values to fit our model to and compare our current AFB estimates with, thus providing information on how much and where AFB had changed.

We calculated the percent differences between current and theoretical AFB estimates for all forested pixels in order to ascertain information regarding how much and possibly where the greatest changes in AFB have occurred. Our calculations indicated that the greatest differences in AFB were within alluvial plains of the Missouri and Mississippi Rivers. These types of systems exhibit highly fertile and well-drained soil that is favorable for growing biomass, and this is where the majority of agriculture exists for Missouri (Nigh and Schroeder 2002). These areas quickly transitioned from forests to agricultural systems during European settlement and have persisted since, and can be attributed to the greatest calculated differences between current and theoretical AFB.

In addition, alluvial systems are prone to disturbances from flooding, which may have also contributed to the large differences calculated between AFB estimates. The geographic locations of the calculated percent changes suggests that additional investigations should focus on using measures of disturbance as additional predictors for estimating current AFB. Disturbances are critical components of alluvial landscape and may have aided in constructing a better-fit model.

### **2.5.2 Variable Importance**

The variable importance plots for the current AFB RF model revealed that all 5 remote sensing indices were top variables for estimating AFB. These specific remote sensing indices are often used in biomass estimation studies because they are indicative of

healthy, productive forest vegetation. In particular, NDVI and NDWI are known to be good descriptors of biomass because their reflectance values saturate from high water content found in healthy vegetation. In contrast, the remote sensing indices ended up being negligible for estimating theoretical AFB. This may be a result of saturation in remote sensing values from within higher biomass areas.

For the theoretical model, elevation was the top variable for estimating AFB followed by a mixture of soil and topographic variables then a climate variable (annual mean temperature). For the current model, Precipitation Seasonality (Coefficient of Variation) and Annual Mean Temperature immediately followed the remote sensing variables in importance for estimating AFB. Missouri is regarded as having an inland (continental) type of climate with strong seasonality due to having a relatively flat topography (<http://www.crh.noaa.gov/>); therefore, elevation may be a factor affecting temperature; however, given the low accuracy of the model, we suggest that elevation only plays a small role in explaining the variation in AFB and that additional factors such as disturbance should be used in future research methods for estimating AFB. Precipitation, which has a distinct northwest to southeast gradient, was ranked just under the remote sensing variables, which may suggest that precipitation has a small role in driving AFB distributions within Missouri (Figure 2-11); albeit minimal due to the poor model estimation.

Overall, climate can be a major stressor for forested ecosystems due to increased variability in extreme weather events (IPCC 2007), which could explain the relative importance of both variables for estimating current AFB. The importance of both climate variables could also be due in part to interactions between regional-scale distributions of



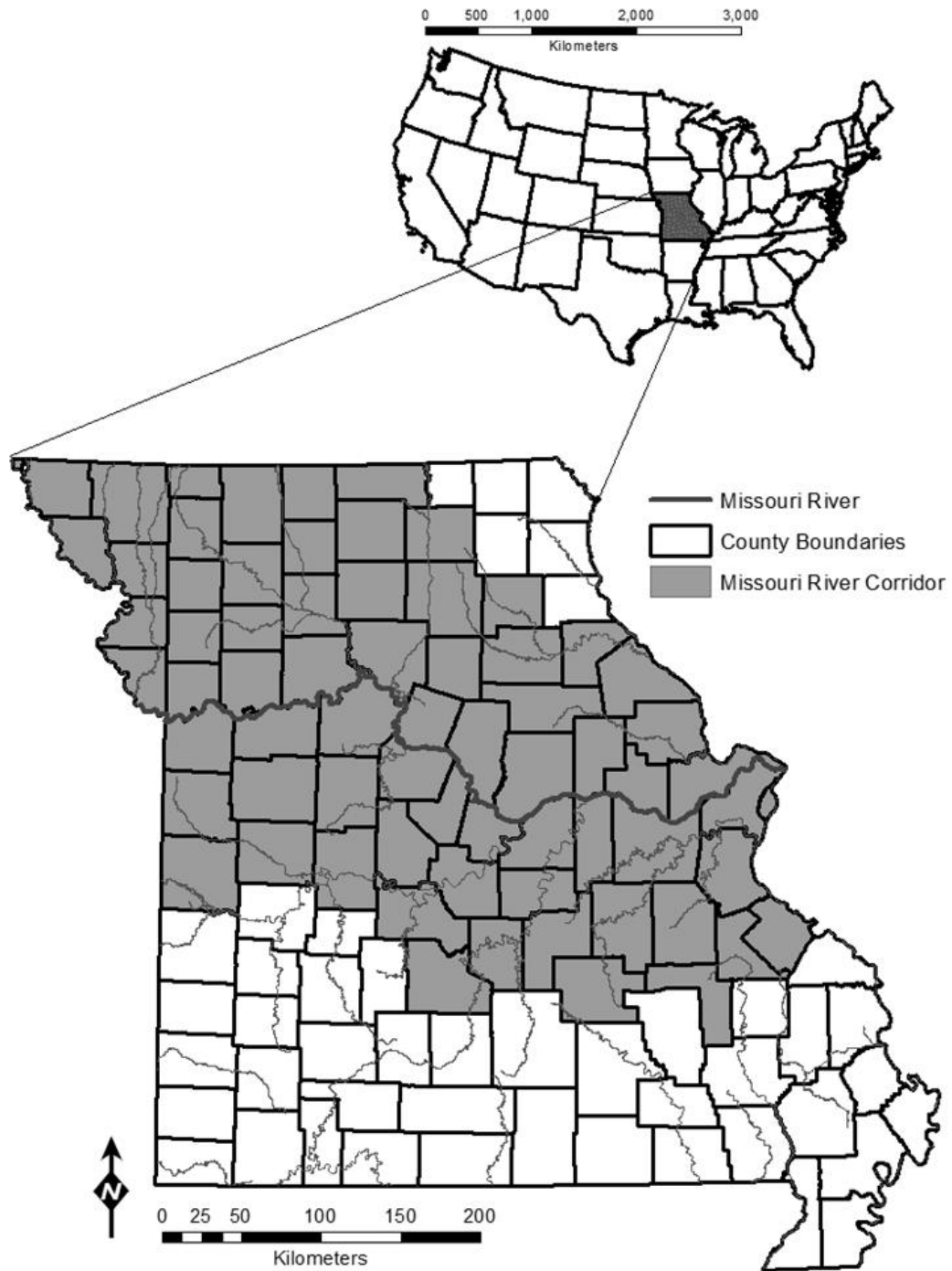
AFB and potential local-scale responses of AFB to multiple additional environmental drivers that we could not add into the model (Brzostek et al. 2014).

The hybrid RF procedure did not perform well, and the results were shown to be inconclusive for identifying major environmental mechanisms (soil, topography, or climate) driving AFB estimations. Top predictors may be discussed; however, they were top predictors for a model that was not able to explain very much variation. Estimating AFB at 30-m resolution had shown to be an ineffective strategy for identifying concentrated areas where major differences in AFB occurred. A relationship between observed and predicted AFB estimates was formulated after summarizing AFB estimations at the ecological subsection scale. An improved relationship was formulated most likely because the scales of FIA data and predictor variables were only adequate at this larger landscape-scale. Additionally, the theoretical AFB estimates were within other published ranges of high biomass values; however, this information does not provide enough information to fully realize this as an accurate baseline measure of AFB within Missouri. Rather, the top 20% of AFB used for estimating should be considered experimental. Future research needs to include true historical data, such as GLO tree information, as well as, additional factors like disturbance to establish a quality baseline measure and understand true differences between potential and available AFB.

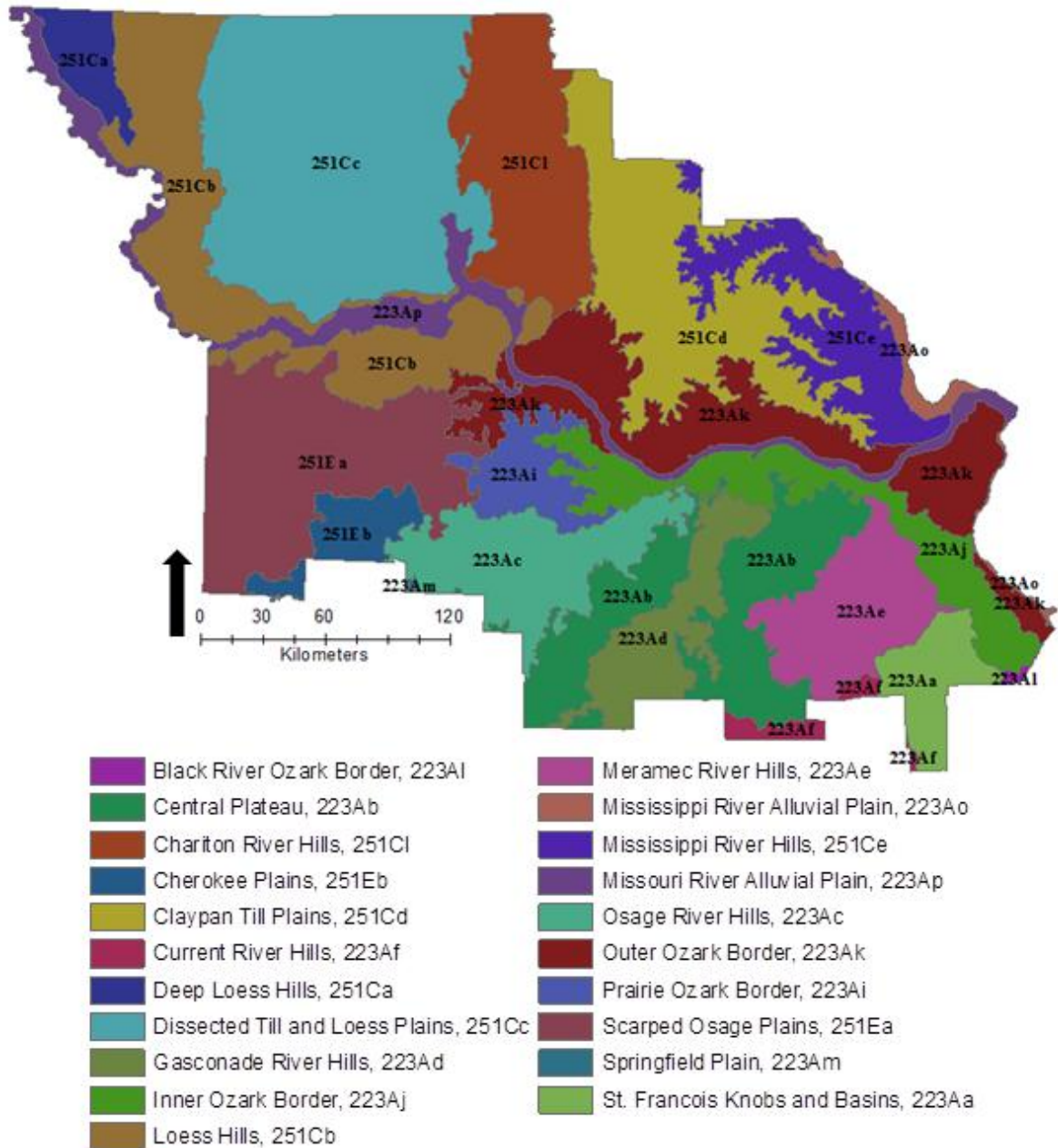
## **2.6 CONCLUSION**

This study provided a first, regional assessment of AFB along the Missouri River corridor using a RF, hybrid modeling approach. Random Forest did not perform well at the 30m pixel resolution but a small statistical relationship could be formed after summarization of estimated AFB values at ecological subsection scales. The RF approach was capable of identifying primary drivers of AFB estimation; albeit with a very low relationship to

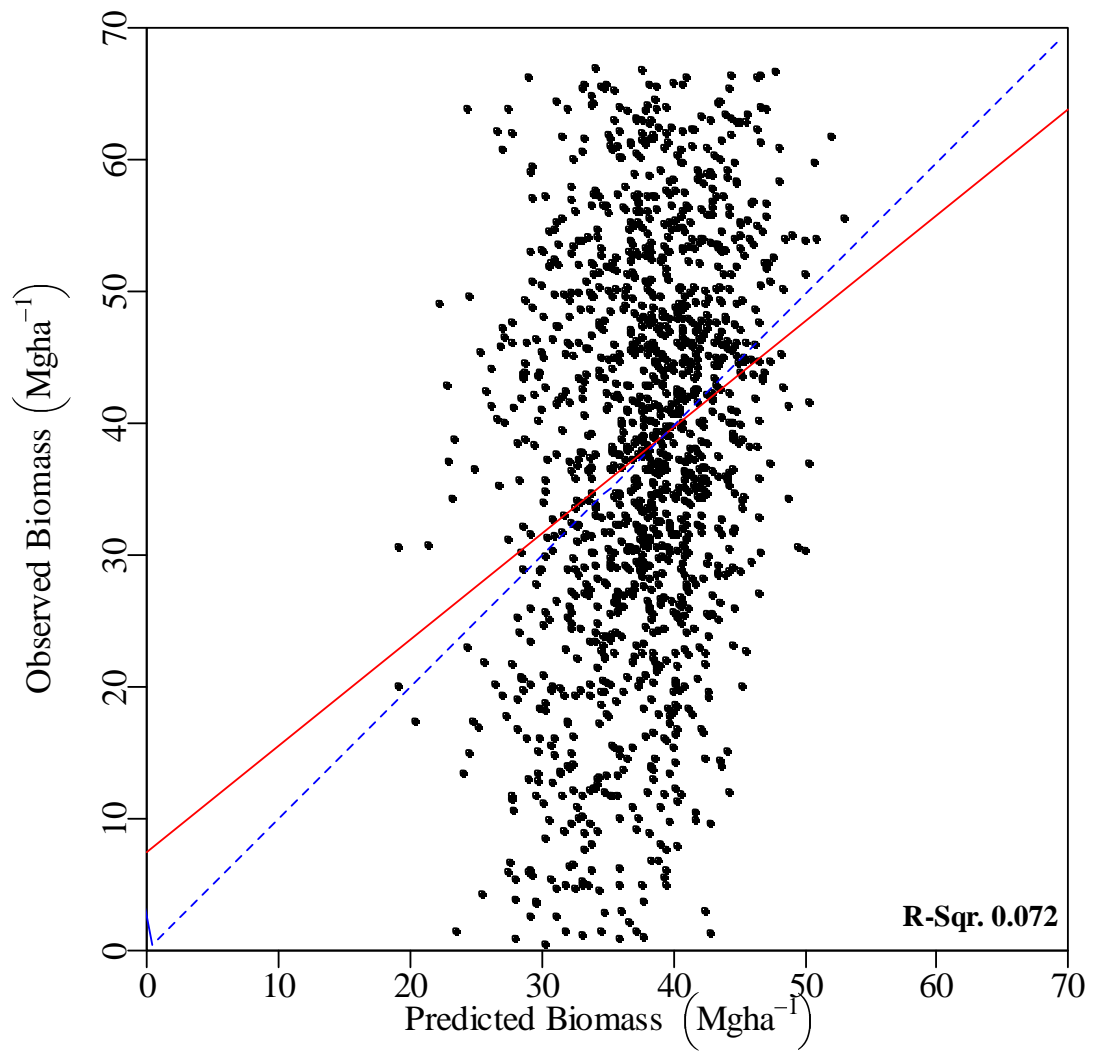
observed data. This type of modeling procedure may only be used for large landscape scale assessments and should be used with caution if applied to help guide ecosystem management objectives.



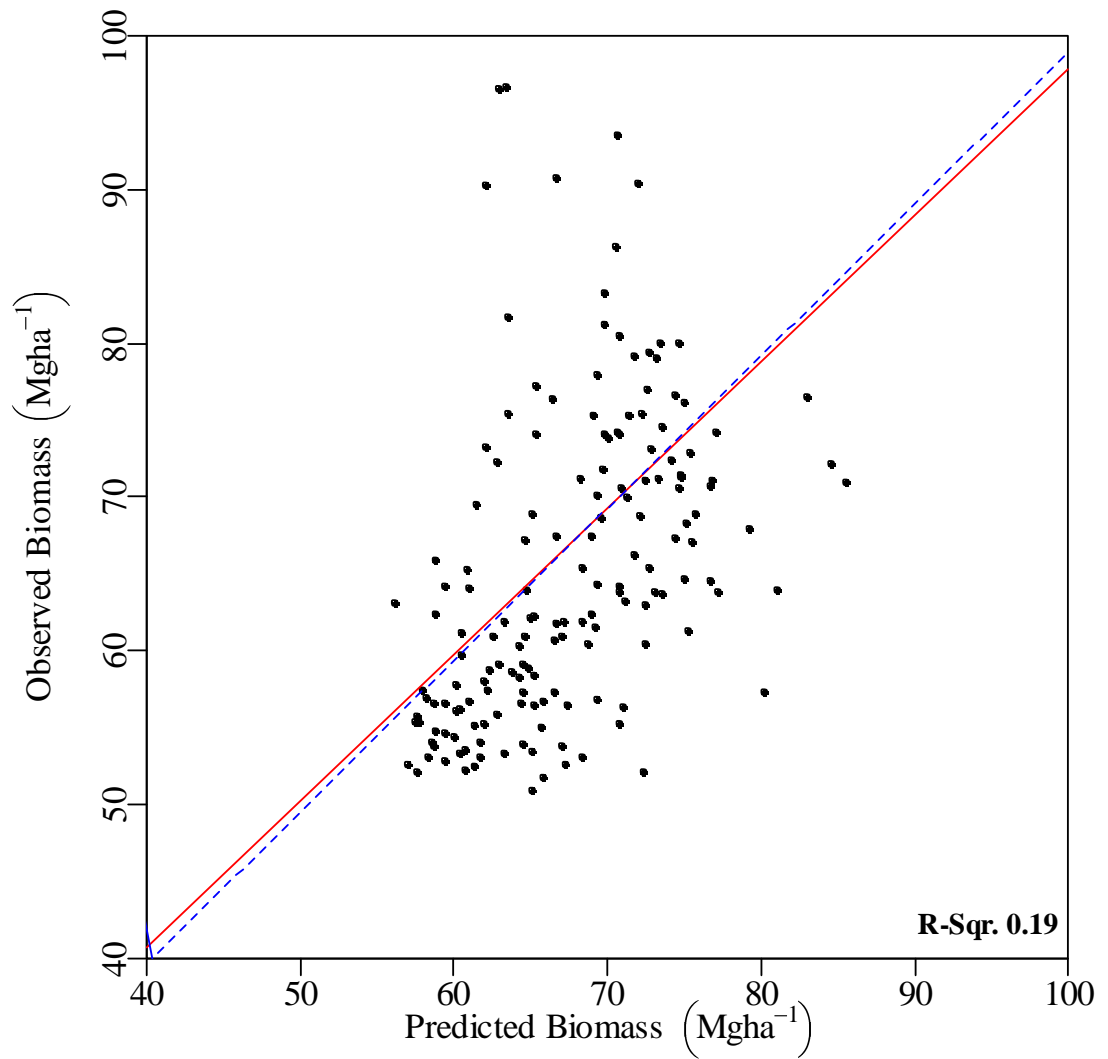
**Figure 2 - 1: The Missouri River Corridor by a 120 km (75 mile) buffer centered from the Missouri River.**



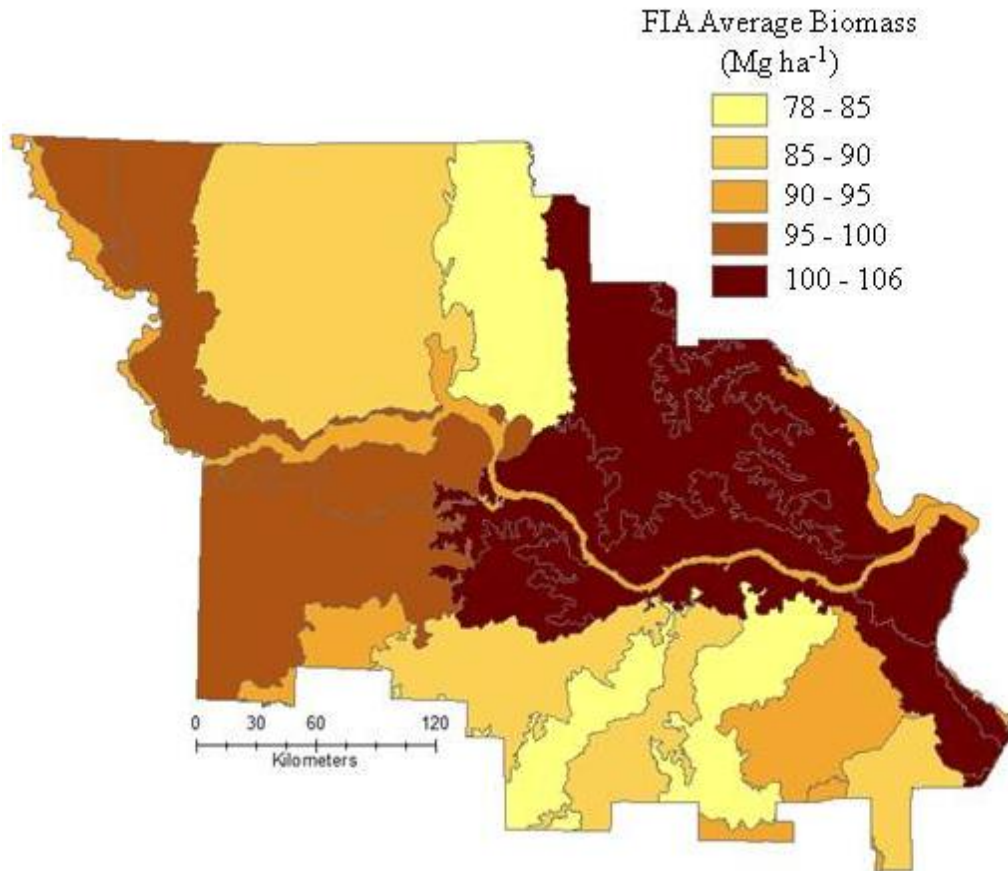
**Figure 2 - 2: Ecological subsections within the Missouri River Corridor.**



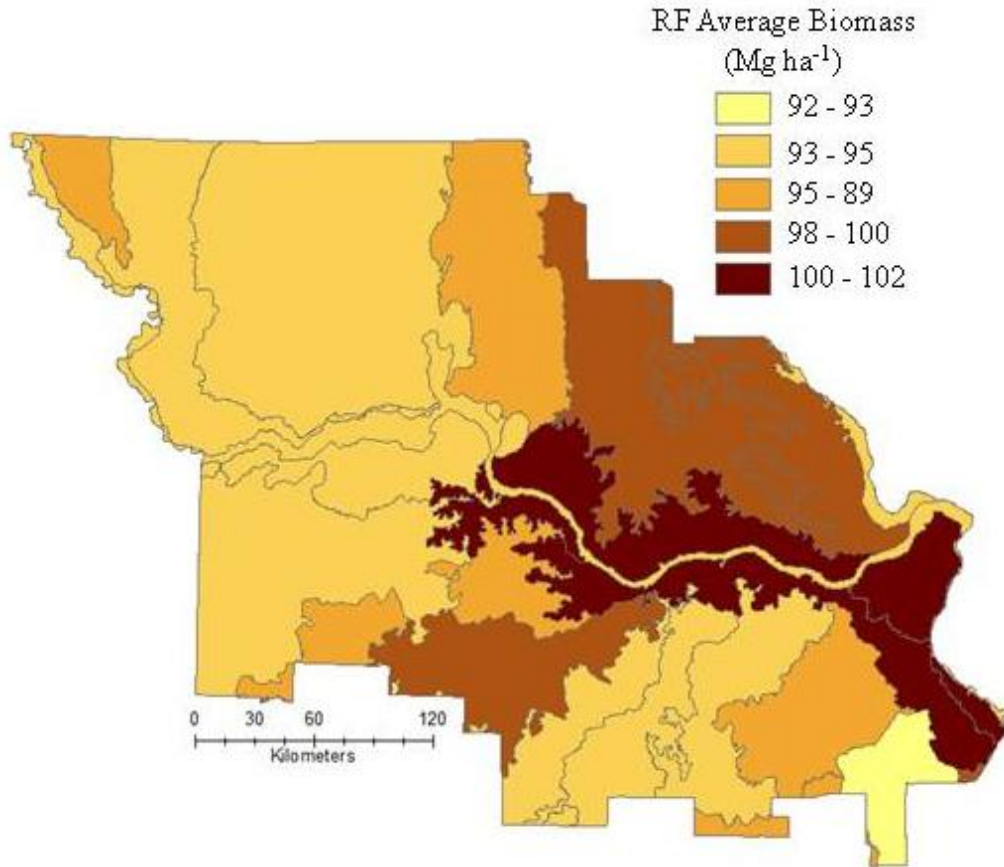
**Figure 2 - 3: Model fit (red line) between observed FIA and Random Forest AFB estimates. The blue dashed line indicates the 1:1 relationship.**



**Figure 2 - 4: Model fit (red line) between the observed upper-bound ranges of FIA data and Random Forest AFB estimates. The blue dashed line indicates the 1:1 relationship.**

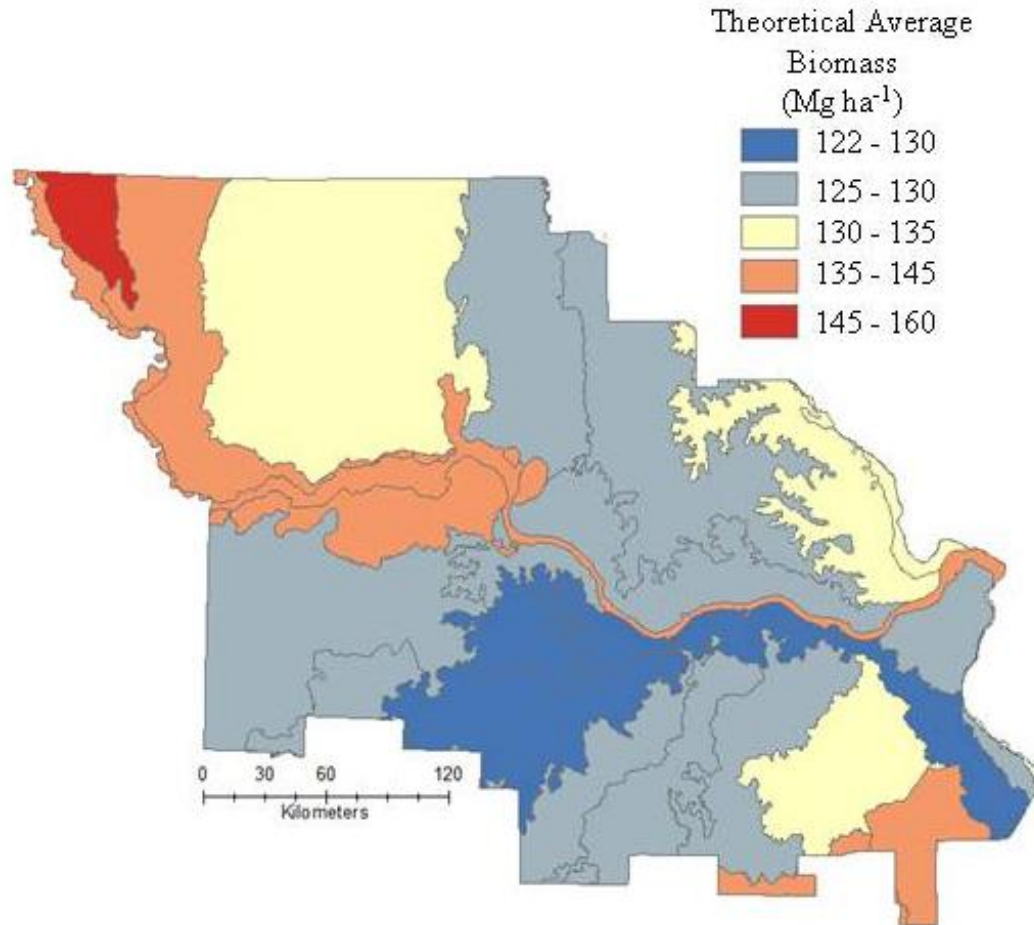


**Figure 2 – 5: Average aboveground forest biomass estimates from FIA summarized by ecological subsection.**

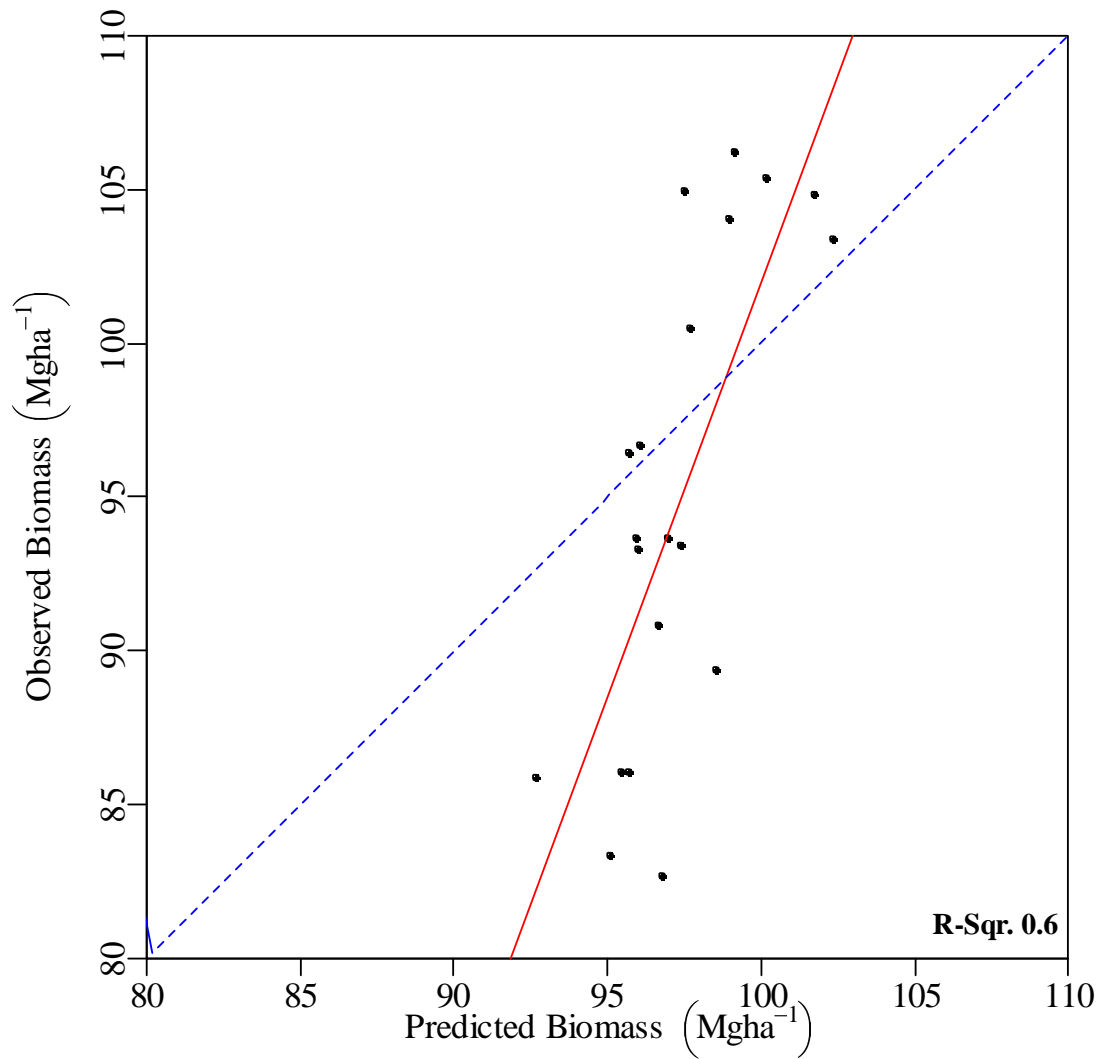


**Figure 2 - 6: Average aboveground forest biomass estimates for the current Random Forest model summarized by ecological subsection. The theoretical estimates were derived from upper bounds of the original FIA data, defined as the top 20% AFB for each ecological subsection.**

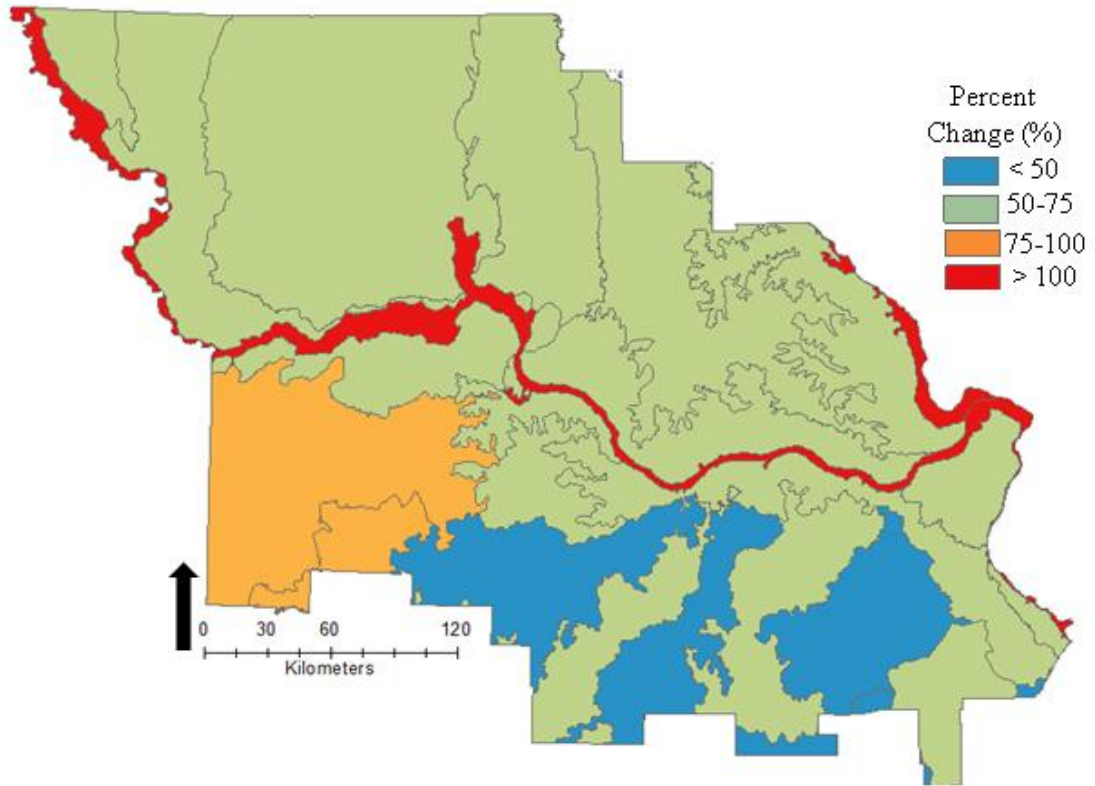




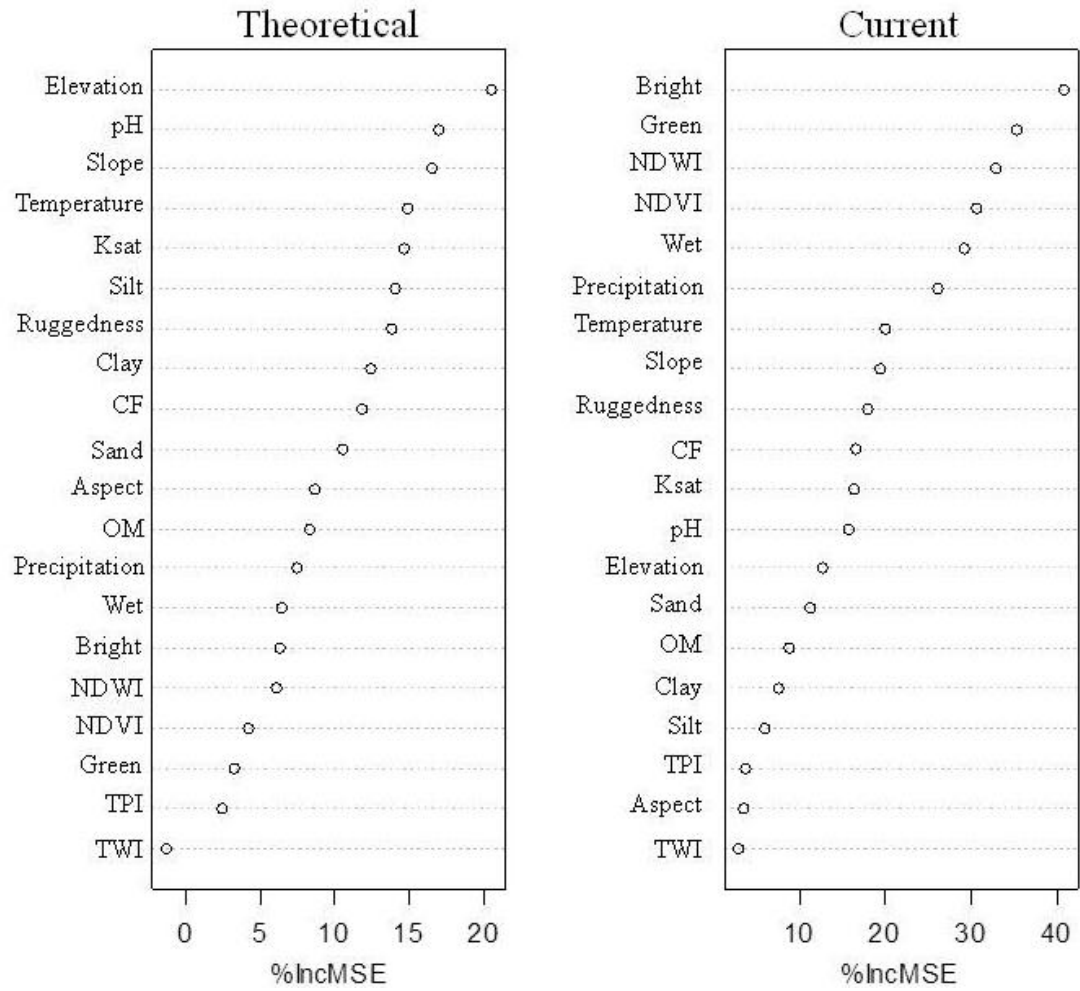
**Figure 2 - 7: Average aboveground forest biomass estimates for the Theoretical Random Forest model summarized by ecological subsection. The theoretical estimates were derived from upper bounds of the original FIA data, defined as the top 20% AFB for each ecological subsection.**



**Figure 2-8: Model fit (red line) between observed (FIA) and predicted (RF) average estimates of AFB summarized by ecological subsection (N=21). The blue dashed line indicates the 1:1 relationship.**



**Figure 2 - 9: Percent change between upper-bound (theoretical) and actual (FIA) aboveground forest biomass estimates from a Random Forest procedure.**



**Figure 2 - 10: Variable importance plot of all variables used to estimate aboveground forest biomass in a Random Forest model procedure.**

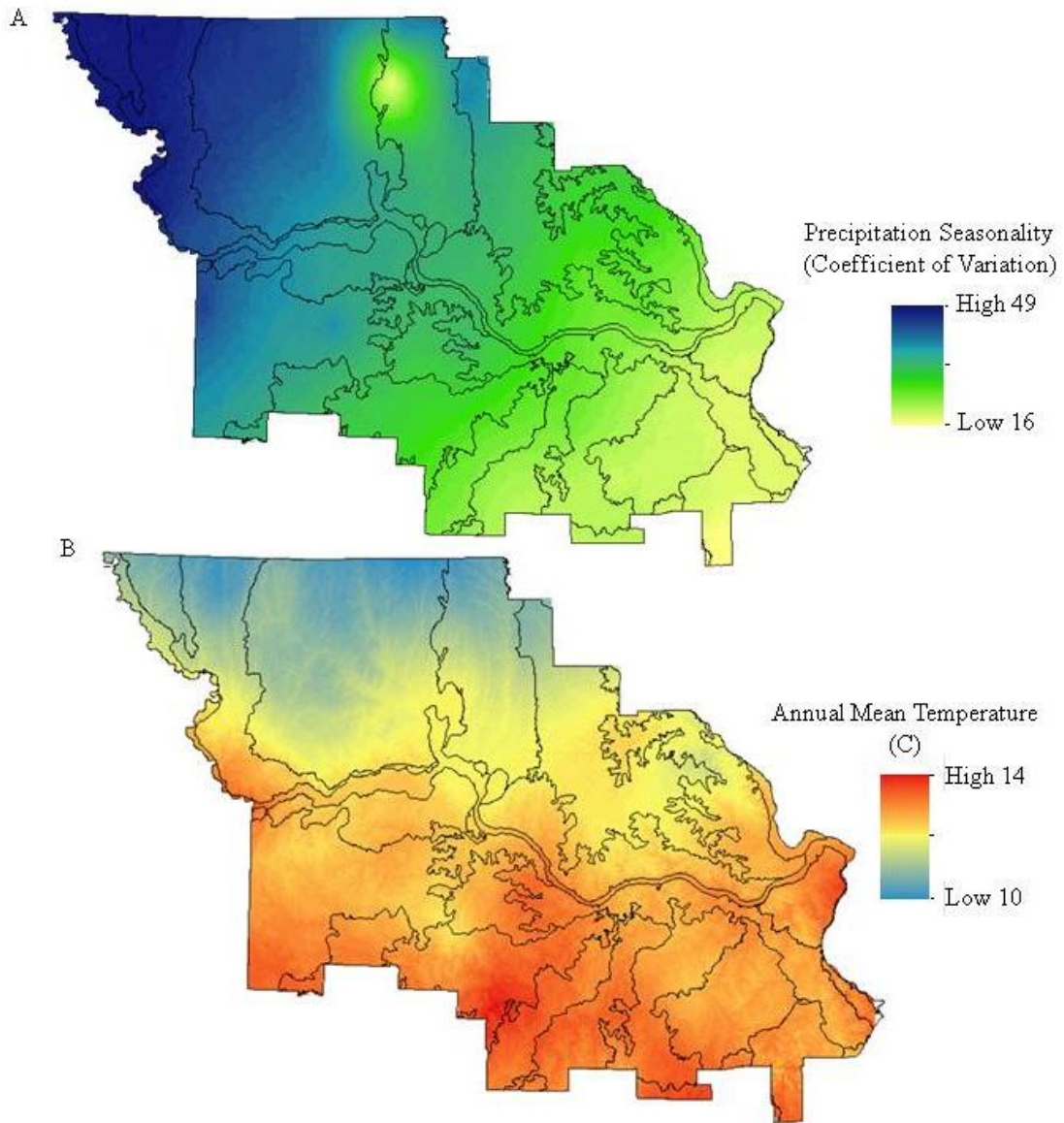


Figure 2-11: maps illustrating (A) The extent of variability in Precipitation seasonality (Coefficient of variation) and (B) distributional gradients of annual mean temperature.

**Table 2 - 1: Ecological subsection descriptive information.**

<b>Subsection Name</b>	<b>Total Area (ha)</b>	<b>Forested Area (ha)</b>	<b>Proportion Forested</b>
Black River Ozark Border	7.6x10 <sup>4</sup>	6.0x10 <sup>4</sup>	78.4
Central Plateau	7.7x10 <sup>5</sup>	3.6x10 <sup>5</sup>	47.5
Chariton River Hills	7.0x10 <sup>5</sup>	2.0x10 <sup>5</sup>	27.9
Cherokee Plains	1.7x10 <sup>5</sup>	3.9x10 <sup>4</sup>	22.5
Claypan Till Plains	9.2x10 <sup>5</sup>	1.3x10 <sup>5</sup>	14.7
Current River Hills	6.7x10 <sup>4</sup>	5.9x10 <sup>4</sup>	88.1
Deep Loess Hills	1.6x10 <sup>5</sup>	1.0x10 <sup>4</sup>	6.7
Dissected Till and Loess Plains	1.7x10 <sup>6</sup>	3.3x10 <sup>5</sup>	18.0
Gasconade River Hills	3.6x10 <sup>5</sup>	2.2x10 <sup>5</sup>	66.3
Inner Ozark Border	5.0x10 <sup>5</sup>	2.3x10 <sup>5</sup>	53.4
Loess Hills	99.8x10 <sup>5</sup>	1.6x10 <sup>5</sup>	16.3
Meramec River Hills	4.6x10 <sup>5</sup>	3.6x10 <sup>5</sup>	78.3
Mississippi River Alluvial Plain	8.1x10 <sup>4</sup>	1.3x10 <sup>4</sup>	15.6
Mississippi River Hills	5.0x10 <sup>5</sup>	1.8x10 <sup>5</sup>	36.8
Missouri River Alluvial Plain	4.3x10 <sup>5</sup>	4.0x10 <sup>4</sup>	9.3
Osage River Hills	5.3x10 <sup>5</sup>	3.4x10 <sup>5</sup>	64.0
Outer Ozark Border	8.7x10 <sup>5</sup>	4.0x10 <sup>5</sup>	45.8
Prairie Ozark Border	2.3x10 <sup>5</sup>	5.6x10 <sup>4</sup>	24.6
Scarped Osage Plains	9.4x10 <sup>5</sup>	1.7x10 <sup>5</sup>	18.4
Springfield Plain	2.0x10 <sup>3</sup>	9.9x10 <sup>2</sup>	52.4
St. Francois Knobs and Basins	22.2x10 <sup>5</sup>	1.5x10 <sup>5</sup>	68.2

**Table 2 - 2: Description of environmental variables used to estimate aboveground forest biomass in a Random Forest modeling procedure.**

<b>Environmental Group</b>	<b>Variable Code</b>	<b>Description</b>
<b>Climate</b>	BIO_01	Mean annual temperature (Celsius)
	BIO_15	Covariation annual precipitation (inches)
<b>Soil</b>	SAND	Sand content (%)
	SILT	Silt content (%)
	CLAY	Clay content (%)
	OM	Organic matter content (%)
	CF	Coarse fragments (%)
	PH	The sum of NH <sub>4</sub> OAc extractable bases (pH 7.0)
	KSAT	Hydraulic Conductivity
<b>Topography</b>	ELEV	Elevation (ft)
	ASP	Aspect (Degree)
	SLOPE	Slope (%)
	TPI	Topographic Position Index
	TWI	Topographic Wetness Index
	VRM	Vector Ruggedness Index
<b>Remote Sensing</b>	NDVI	Normalize Difference Vegetation Index
	NDWI	Normalized Difference Wetness Index
	B*	Brightness
	G*	Greenness
	W*	Wetness

*\*Derived from tasseled-cap transformation.*

**Table 2 - 3: Remote sensing image descriptions.**

**Spacecraft ID: Landsat 5**

**Sensor ID: TM**

**Product Type: L1T**

<b>Path Number</b>	<b>Row Number</b>	<b>Acquisition Date</b>
23	34	Aug-10
24	33	Jul-11
24	34	Jul-11
25	32	Jun-10
25	33	Jul-11
25	34	Aug-11
26	32	Aug-10
26	33	Jul-11
26	34	Jul-11
27	32	Jul-11
27	33	Jun-11



**Table 2 - 4: Average aboveground forest biomass (AFB) estimates for the observed data from Forest Inventory and Analysis (FIA), Random Forest (RF) estimates of actual AFB (Actual), and RF estimates of theoretical AFB (Theoretical) summarized by ecological subsection.**

<b>SUBSECTION</b>	<b>FIA (Mg ha<sup>-1</sup>)</b>	<b>Actual (Mg ha<sup>-1</sup>)</b>	<b>Theoretical (Mg ha<sup>-1</sup>)</b>
Black River Ozark Border	105.33	96.65	107.79
Central Plateau	83.27	91.44	107.99
Chariton River Hills	82.62	88.08	130.82
Cherokee Plains	90.79	91.28	130.82
Claypan Till Plains	104.03	91.07	130.01
Current River Hills	93.64	95.04	117.57
Deep Loess Hills	100.44	97.45	124.97
Dissected Till and Loess Plains	86.01	88.39	133.70
Gasconade River Hills	86.04	92.75	111.40
Inner Ozark Border	104.84	95.48	120.25
Loess Hills	96.42	94.16	122.06
Meramec River Hills	93.39	95.00	121.19
Mississippi River Alluvial Plain	93.66	88.22	145.61
Mississippi River Hills	106.18	93.40	126.38
Missouri River Alluvial Plain	93.30	90.25	135.59
Osage River Hills	89.34	93.56	114.06
Outer Ozark Border	103.37	96.40	126.48
Prairie Ozark Border	104.93	94.75	114.59
Scarped Osage Plains	96.64	93.48	123.72
Springfield Plain	78.30	90.84	100.07
St. Francois Knobs and Basins	85.81	93.05	115.68
<b>AVERAGE</b>	<b>94.21</b>	<b>92.89</b>	<b>121.94</b>

**Table 2 - 5: Random Forest model statistical summaries between current (FIA) and theoretical (Upper-bound FIA) aboveground forest biomass estimates.**

<b>MODEL METRIC</b>	<b>CURRENT</b>	<b>THEORETICAL</b>
Total AFB Estimates	3.4 x10 <sup>8</sup> Mg	4.2 x10 <sup>8</sup> Mg
Total Average AFB Estimates	97 Mg ha <sup>-1</sup>	121 Mg ha <sup>-1</sup>
Mean Absolute Error (MAE)	11.38 Mg ha <sup>-1</sup>	9.94 Mg ha <sup>-1</sup>
Root Mean Square Error (RMSE)	14.39 Mg	12.86 Mg
Coefficient of Determination (R <sup>2</sup> )	0.07	0.19

**Table 2 - 6:** Paired two sample t-Test for means of average estimates of AFB summarized by ecological subsection between FIA and the Current RF Model.

	<i>RF Current</i>	<i>FIA</i>
Mean (Mg ha <sup>-1</sup> )	97.13	94.13
Standard Deviation (Mg ha <sup>-1</sup> )	24.63	8.58
Observations (n)	21	21
Pearson Correlation	0.78	
df	20	
t Stat	2.00	
P(T<=t) two-tail	0.06	
t Critical two-tail	2.09	

## **CHAPTER 3: Determining environmental controls driving discrepancies between historic and current aboveground forest biomass along the Missouri River corridor, Missouri, USA**

*[T]he way to deal with large-scale complexity is to search for overriding simplicity. Sometimes, this turns out to be old-fashioned common sense.*

*– Eugene Odum*

### **3.1 ABSTRACT**

Aboveground forest biomass (AFB) is an integral component in assessing the status of forested ecosystems, and changes in distributions are considered a critical proxy for determining effects of global climate change. Additional information on how the abiotic environment may influence variations in biomass remains a critical point of research, especially at large landscape scales. Landscape models are important tools for ecologists to geographically visualize feedbacks between AFB and abiotic controlling factors. This investigation used a series of abiotic factors consisting of soil, topography, and climate characteristics to estimate current and historic AFB, and spatially determine percent changes between the two estimates of AFB at large landscape scales along the Missouri River corridor, located in Missouri, USA. Random Forest (RF), an ensemble regression tree method, was used as the main procedure to estimate AFB across the landscape and determine variable importance. All estimates were conducted at a 30-meter spatial resolution and then summarized by ecological subsections, which are natural divisions within Missouri categorized by similar environmental characteristics. The RF model produced very poor estimates of AFB for both the historic ( $R^2 = 0.1$ ) and current ( $R^2 = 0.02$ ). Despite inaccurate model estimates, the best predictors of AFB determined by the RF model were climate variables (Annual Mean Temperature (Bio\_01) and Precipitation

Seasonality (Coefficient of Variation) (Bio\_15)) for both historic and current estimates of AFB. We also calculated the percent changes between historic and current models, which were an average of - 63% decline in AFB from historical values along the Mississippi, Missouri River alluvial plains, and St. Francois Knobs and Basin subsections. The smallest percent changes that were summarized by ecological subsection (< 50%) occurred within the Central Plateau subsection, along the southern portion of the study area and the Deep Loess Hills subsection, found in the upper northeast corner of Missouri. This investigation could not account for recent harvests within FIA plots, which may have led to poor estimates of current AFB. Although the model did not perform well, what we can ascertain from the given results illustrate was that major there was a major decline in AFB, which may due to large-scale variations in recovery rates from major landscape clearing for agriculture during European settlement. The two climate variables (BIO\_01 and BIO\_15) were the top predictors of AFB for both models; however, because of the poor model performance for both models, the variable importance results cannot be used as sufficient evidence for determining specific drivers of AFB; albeit it has been noted that changes in both precipitation and temperature are increasing the vulnerability of forests to changes in AFB disturbances.

### **3.2 INTRODUCTION**

The burning of fossil fuels is the primary source of atmospheric CO<sub>2</sub>, which makes up approximately 60 percent of total greenhouse gas emissions in the U.S. (IPCC 2007). Concentrations are expected to continue rising unless major social and environmental changes are made to help limit dependency on fossil fuels. To lessen this dependency and strengthen energy independence, the U.S. Energy Independence and Security Act of 2007 (EISA) mandated an increase in production of renewable fuels from 9 billion gallons in 2008 to 36 billion gallons by 2022 (EISA 2007). The demand to meet alternative fuel needs, however, remains a contentious topic because of potential environmental, political, and cultural implications associated with competing land-use interests. An avenue that is currently under investigation to help meet alternative energy needs and abate conflicting interests is using existing stands of biomass, as well as forest stands on marginal lands, for wood production and investment in the wood-based bioenergy industry.

Aboveground forest biomass (AFB) is receiving a great deal of attention as an alternative source for energy within the United States (Bardhan and Jose 2012; Smeets and Faaij 2007; Houghton et al. 2009). One reason for the attention is that AFB has several environmental advantages compared to fossil fuels. First of all, carbon is released into the atmosphere from both fossil fuels and AFB; however, fossil fuels release carbon into the atmosphere that was buried for millennia, whereas burning AFB releases carbon that was sequestered in plants for hundreds of years (Goerndt et al. 2012). Additionally, reforestation or planting dedicated crops to replace harvested wood will be able to sequester an amount of carbon equivalent to that released by burning woody biomass (Hoogwijk et al. 2003). In fact, under current conditions, AFB is capable of capturing and

storing more than 270 million tons of carbon per year (USDOE 2011).

Forest biomass has steadily been recovering from major disturbances from European settlement during the late 18<sup>th</sup> century (Hanberry, Palik, et al. 2012). Approximately 33 percent (304 million hectares) of U.S. land area is now forested (Smith et al. 2007) and is projected to continually increase leading to potential abundance of biomass feedstock available to use within the biofuel industry. The sources of wood are likely to come from two sources, residues left over from logging activities and individual trees from overstocked forest stands (Goerndt et al. 2012). The need for alternative sources of energy rising and Missouri harbors a considerable base load of AFB that could possibly be used to some capacity for energy production but the question of how much and where remains open. Research is now focusing on assessing the spatial locations and changes in AFB over large landscapes to understand where and in what quantities exploiting AFB for bioenergy may be possible (Berndes et al. 2003; Shifley 2006; O'Laughlin 2009).

Considerable efforts have focused on remote sensing technology to estimate AFB at various scales using passive (Lu 2006; Anaya et al. 2009) and active (Zolkos et al. 2013; Kellndorfer et al. 2010) sensory information. Models generated from remotely sensed data contain fine resolution information to define structural attributes of forest stands, but lack the ability to accurately match robust estimates statistically derived from ground-based forest inventory (Zheng et al. 2008). Remote sensing information is also unable to address historical factors that are considered to be a prominent causality in current AFB patterns (Hanberry et al. 2012). Hybrid models, an additional method for estimating AFB, rely on a combination of remote sensing and inventory data. This hybrid

approach has illustrated an improvement in predictions compared to inventory or remote sensing alone (Baccini et al. 2004; Wulder et al. 2008; Yeo and Huang 2013; Gallaun et al. 2010); however, hybrid modeling approaches still lack a historical context, which neglects legacy effects (Liu et al. 2007) from previous anthropogenic impacts on forest ecosystems (Zheng et al. 2008; Powell et al. 2010).

Modeling forest biomass over large areas is a complex process with high variability because of inherent differences in environmental controls underlying biomass distributions (Yeo and Huang 2013). Subsequently, there are various techniques to model biomass and critically examine the potential factors controlling distributions across the landscape. Historical data are an effective avenue for determining baseline measures of change in forest landscapes (Grumbine 1994; Heinemann 2010). Historical inventory records, such as those from the General Land Office (GLO), can be used to model AFB conditions prior to major anthropogenic disturbances (Hanberry, Kabrick, et al. 2012; Rhemtulla et al. 2008; Hanberry, Palik, et al. 2012). Landscapes indicative of pre-European settlement are a source of reference to which current forested landscapes can be compared and changes spatially identified (Moser et al. 2006). Combining two snapshots in time that cover approximately 150 years can provide an efficient baseline to use for spatially identifying where and how much change in AFB has occurred. The controlling features that dictate AFB distributions can be characterized based on limiting factors known to drive AFB distribution and success (Zimmermann et al. 2007). Combining all these elements into one model may provide the information necessary to identify major changes in AFB and suggest where there is potential to improve AFB at the ecosystem scale (Cannell 2003).



Forestry-based energy programs are focusing on river corridors, such as the Missouri and Mississippi Rivers, as viable options for woody biomass energy production. River corridors are central areas with unique functional characteristics that include accessibility to biomass, existing transportation networks, and industrial infrastructure. Additionally, marginal lands (areas unsuitable for traditional agriculture due to frequent flooding, erosion, or poor soil quality) are often found along river corridors and are considered to be important areas to implement afforestation, thereby avoiding competition or degradation of agriculture lands. Furthermore, transportation costs are considered a limiting economic factor in transitioning to wood-based bioenergy. The expense of transporting biomass is considerably greater than fossil fuels on a per unit energy basis due to low bulk density (Shabani and Sowlati 2013). The procurement of biomass on dedicated lands nearer the source may allow for easier transportation using existing barge traffic that is already carrying various industrial and agricultural products.

The Missouri River corridor is a particularly important area for advancing biomass feedstock for alternative energy production. About one-third of the state is forested, which ranks Missouri one of the top states for forest land area (MDC 2009). The Missouri River corridor also has an existing industrial infrastructure comprised of a dense network of roadways, railroads, and barge traffic that enables this area to provide affordable ways to transport biomass. Various studies have focused on the life cycle analysis of biomass (Cherubini and Strømman 2011), but the issue of how to sustainably manage standing AFB for future procurement remains a contentious topic and needs considerably more research, especially with respect to the potential utility of river corridors as transportation routes in the biofuel industry. This investigation attempts to

use a statistical modeling approach to estimate biomass along the Missouri River corridor and identify what environmental factors may be driving particular patterns of AFB at landscape scales.

The overall objective of this investigation is to estimate historic and current AFB, and examine what primary abiotic controls may be driving AFB along the Missouri River corridor. Additionally, we determine discrepancies between historic and current model estimates to understand where differences between historic conditions may exist. This study uses historical and contemporary tree survey data in combination with climate and geologic variables to examine the distributions of AFB, and assess the main factors that may control AFB patterns. This modeling procedure is static in time, focusing on two critical periods, historic (1800's) and current (2011), and uses fundamental controls on characteristics of plant growth (climate, topography, and soil conditions) as the modeling variables. This study investigates the underlying properties of abiotic covariates in order to offer resolution to primary management questions for Missouri, such as: (i) how much AFB is there on a regional scale in the Missouri River corridor, (ii) how AFB varies across the landscape, and (iii) what external factors control AFB patterns?

### **3.3 METHODS**

#### **3.3.1 Study Area**

This investigation was conducted within the Missouri River corridor (MRC), located within Missouri, USA. The MRC is 106,000 km<sup>2</sup> area defined by a 120 km (75 mi) buffer that encompasses 69 counties along the lower Missouri River (Figure 3-1). This area is described as a central processing region that has the capacity to guide the biofuel industry by promoting energy crop competitiveness using existing infrastructure (rail, roads, and

river) and utilizing economically viable transportation distances (i.e. 120 km buffer). The region offers large acreages of both marginal and productive agricultural land with a variety of soil types, water sources, and biomass options (residues from forest and agricultural harvesting, as well as, standing biomass). The geography of the MRC consists of a mosaic of floodplains, crop, pasture, and secondary-growth forestlands with extensive past grazing influences (Nigh and Schroeder 2002). Elevations along the MRC range between 102-540 m (334-1771 ft.). Land cover is diverse and includes large swaths of productive farmland adjacent to the main channel and regions of oak savannas, woodlands, mixed deciduous forests, and occasional prairie and glade openings at higher elevations. Upslope from the river valley are bluff forests dominated by comparatively mesophytic red oak (*Quercus ruba*), sugar maple (*Acer saccharum*), basswood (*Tilia americana*) and black walnut (*Juglans nigra*). Bottomland hardwood forests stands exist in areas less subject to flooding and are dominated with cohorts of oak (*Quercus spp.*), hickory (*Carya spp.*), and maple (*Acer spp.*). The riverfront has fluctuating water levels that create aggregations of sandbars and forest stands dominated by cottonwood (*Populus spp.*), willow (*Salix spp.*), maple (*Acer spp.*), and elm (*Ulmus spp.*).

The MRC harbors varying degrees of soil and landscape quality that ranges from flat floodplains with deep, fertile alluvium and loess soils to moderately rolling and steeply dissected hills with shallow, unconsolidated soils, which is mostly characteristic of the southern Ozark region. Regional and Landscape scale ecological classifications were generated by Nigh and Schroeder (2003) as a framework to categorize ecosystem diversity and generate spatial datasets for broad-scale ecosystem management. At intermediate spatial scales, the MRC is divided into sections and subsections defined by

changes in topography, geomorphology, soil characteristics, and vegetation type (Nigh and Schroeder 2002). The MRC lies within two major landform sections, Central Dissected Till Plain and Ozark Highland. This area is further divided into an additional 21 smaller subsections with areas ranging from  $1.0 \times 10^3$  -  $4.0 \times 10^5$  ha with an average of  $2.0 \times 10^5$  ha. All AFB descriptive statistics were stratified by ecological subsection and reported based on summarizations of each subsection (Table 3-1).

### **3.3.2 Modeling Framework**

Random Forest (RF) (Breiman 2001), accessed using R statistical software (Liaw and Wiener 2002; R Development Core Team 2011), was the modeling platform used to predict AFB quantities and determine variable importance. The modeling procedure was conducted at the 30 meter pixel resolution. Parameterization of the model used 1500 iterations randomly sampling 4 parameters from the total pool of 15 variables.

RF is a non-parametric, ensemble regression tree method that grows branches and nodes based on random subset sampling from a pool of explanatory variables (Boulanger et al. 2013). A split is determined at each node and the process is repeated based on new bootstrap samples. The final prediction is a weighted average across all regression trees (Breiman 2001; Cutler et al. 2007; Prasad et al. 2004). Variable importance is determined by permuting the covariate values, where an out-of-bag (OOB) error is generated. The permuted OOB error is then compared to the original (or training data). OOB error and the variables with larger differences (i.e. increases in mean square error) contribute more to the model prediction and are weighted accordingly (Prasad et al. 2004). Variable importance is used to find critical parameters related to the response variable and used for interpretation of why the parameter is chosen. Also, variable importance may be used to

narrow down the number of parameters that are sufficient to make a parsimonious prediction.

RF has many advantages over other statistical models. It provides well-supported predictions and classifications (Prasad et al. 2004; Cutler et al. 2007; Boulanger et al. 2013) and enables the user to generate rapid, straightforward output. Additionally, RF is capable of handling large non-linear data sets or variables that suffer from collinearity (Prasad et al. 2004; Cutler et al. 2007; Boulanger et al. 2013).

### **3.3.3 Explanatory Variables**

We used a set of 15 variables to predict AFB. These variables were classified into three main categories: 1) climate, 2) topography, and 3) soil. We used ArcMap™ v10.2 (ESRI 2013) as the main platform for visualization and summarization of AFB estimates.

Climate data included annual mean temperature (Bio\_01) and precipitation seasonality (Coefficient of Variation) (Bio\_15). The data were acquired from WorldClim database (Hijmans 2005) ([www.worldclim.com](http://www.worldclim.com)). We used 6 variables to represent physiographic characteristics of the landscape, which included: 1) elevation, 2) aspect, 3) slope, 4) topographic position, 5) topographic ruggedness, and 6) topographic wetness. All features were extracted from a single digital elevation model (DEM) using ArcGIS embedded spatial analyst tools for creating surface components. The DEM was obtained from Missouri Spatial Data Information Services. Seven soil variables were extracted using the United States Department of Agriculture Natural Resource Conservation Service's Soil Data Viewer v6.0 (NRCS 2011), which is a graphical user interface that extracts soil information from the United States Department of Agriculture's Natural Resource Conservation Service (NRCS) Soil Survey Geographic (SSURGO) database.

We used 7 soil variables as predictors in the RF model: 1) Sand, 2) Silt, 3) Clay, 4) Coarse Fragments, 5) Hydraulic conductivity (Ksat), 6) Alkalinity (pH), and 7) Organic matter. All soil variables were extracted from the most detailed map unit level defined and projected as polygonal boundaries with scales ranging from 1:12,000 to 1:63,360. The data were transformed to a raster dataset using the same 30m<sup>2</sup> resolution.

### **3.3.4 Historic Aboveground Biomass**

Historic biomass values were quantified from General Land Office (GLO) data. GLO data collection occurred between 1815 and 1900 and was sampled based on a rectangular survey grid that generated a system of townships, composed of 36.1 km<sup>2</sup> sections (White 1983). Inventory measurements (species, DBH, distance, and bearing from surveying boundaries) from 2-4 trees were taken at each corner and center of section lines.

The average DBH from the group of trees was calculated and assigned to each point location. Historical forest density estimates were quantified at the subsection level (Hanberry et al. in preparation) and assigned ubiquitously to all points residing in the respective subsection boundaries. Aboveground forest biomass was calculated using Jenkins et al. 2003 DBH-based equation for mixed deciduous hardwood species:

$$bm = \text{Exp}[(\beta_0 + \beta_1) \ln (dbh)]$$

eq. 3

where the total weight of aboveground biomass (*bm*) in kg of dry weight for trees  $\geq 2.5$  cm is the exponential function of the addition of regression coefficients ( $\beta_0 + \beta_1$ ), which are -2.4800 and 2.4835, respectively, multiplied by the natural log (2.718282) of diameters at breast height (*dbh*) (Jenkins et al. 2003). This equation is representative of many species, locations, and the equation should be expected to have a larger range in

predicted component biomass as compared to other equations from studies with a smaller scope. Thus, the biomass equation should be used for large-scale application.

### **3.3.5 Current Aboveground Biomass**

We used the total number of AFB values calculated by FIA personnel as the observed dataset to train and validate the actual RF model. We used a Student's T-test for means ( $\alpha = 0.05$ ) to validate the actual RF model with estimates derived from FIA only at the ecological subsection scale. There was no opportunity to validate at 30m-resolution.

The FIA program of the USFS produces comprehensive data on the status, changes, and trends in the conditions of all forest ecosystems within the U.S. (Woudenberg et al. 2010). The FIA program provides plot-level information on 5-year cycles and is designed for landscape-scale studies and regional monitoring of forest lands (Moser et al. 2010). We were only allowed up to 20 variables (please see the previous chapter) and the spatial locations of FIA plots were kept confidential. Thus, data acquisition (FIA cycle 2006-2011) was required to be coordinated with personnel from FIA Spatial Data Services where a confidential agreement form was signed. Of the 20 variables obtained from with AFB values from FIA, we had to eliminate 5 remote sensing variables in order to use the same variable structure with both the current and historic models. Consequently, we used only 15 explanatory variables in the RF model to estimate AFB.

## **3.4 RESULTS AND DISCUSSION**

### **3.4.1 Aboveground Forest Biomass Estimates**

The historic and current RF models performed very poorly with respect to estimating AFB. Historic AFB was derived using a total of 10,857 inventory points where each

location had  $\geq 2$  trees sampled and the coefficient of determination was  $R^2 = 0.09$  (Figure 3-2). Current AFB was derived using a total of 1,100 inventory plots and also had a very low coefficient of determination ( $R^2=0.02$ ) (Figure 3-3).

The RF model estimated AFB for all pixels across the landscape where we then selected pixels that were categorized as forest based on the 2006 National Land Cover Dataset (Figure 3-4). Based on only forested pixels, the total historic AFB within the MRC was estimated at  $8.1 \times 10^8$  Mg and the current model estimated total AFB at  $3.4 \times 10^8$  Mg (Table 3-2). We then summarized both historic and current AFB estimates by ecological subsection where the average historic and current AFB was  $234 \text{ Mg ha}^{-1}$  with standard deviation of  $26 \text{ Mg ha}^{-1}$  and  $99 \text{ Mg ha}^{-1}$  and standard deviation of  $25 \text{ Mg ha}^{-1}$ , respectively (Table 3-3). For the current model only, average AFB summarized by ecological subsection was not statistically different from FIA estimates per ecological subsection ( $p=0.34$ ,  $\alpha = 0.05$ ) (Table 3-4). The historic baseline estimates of AFB resulted in a reasonable range of AFB values that are similar to previous studies reporting historical or old growth biomass estimates around  $300 \text{ Mg ha}^{-1}$  (Saatchi et al. 2007; Johnson and Bell 1976). The current model of AFB also had a reasonable biomass average as compared to the most recent FIA estimates (2011). The extent of variability with respect to the mean values of AFB estimates at the subsection scale exhibited a low coefficient of variation for both historic and current model (0.09 and 0.02, respectively).

### **3.4.2 Variable Importance**

Variable importance was derived from the RF model for the historic and current estimates of AFB, and illustrated as a graph of importance in descending order (Figure 3-5). The mean square error (MSE) values for the historic model ranged between 0 – 90%



while the current model exhibited a much tighter, MSE range, between 0 – 35%. Both climate variables, Annual Mean Temperature (Bio\_01) and Precipitation Seasonality (Coefficient of Variation) (Bio\_15), were the top 2 variables for estimating AFB for historic and current RF models; albeit the position of the these two variables with respect to the remaining variables yields minimal information because of the poor estimating power of the RF model. Thus, the two climate variables are considered to be the top estimators of AFB under a very inaccurate model foundation that explains minimal variation.

Elevation was ranked as the third best variable for both the historic and current RF models but, again, due to the poor model performance, this explains very little information and climate and elevation variables must be regarded as top predictors within a very inaccurate model. It should also be noted that the variables that contributed the least to estimating AFB were also the same for both models and consisted of topographic roughness, topographic wetness, and aspect.

### **3.4.3 Percent Changes**

We calculated the percent change between historic and current estimates of AFB (Figure 3-6) and then summarized them by ecological subsection (Figure 3-7). The largest percent changes (i.e.  $\geq 65\%$ ) occurred within three different subsections: St. Francois Knobs and Basin (67%), Missouri River Alluvial Plain (66%), and Mississippi River Hills (65%) (Table 3-2). In all cases, the summarized mean historic AFB levels were greater than current AFB levels, which resulted in negative percent changes. The smallest percent changes between historic and current AFB estimates ( $< 45 \text{ Mg ha}^{-1}$ ) occurred in the Central Plateau subsection, which is located along the southern portion of the study

area and is part of the Ozark region. An additional subsection that had small percent changes was along the Deep Loess Hills subsection, which is found in the upper northeast corner of Missouri (Figure 3-8).

#### **3.4.4 Abiotic controls**

Climate variability can be a major stressor on forested ecosystems and it is noted that extreme weather events may become more frequent, severe, and isolated in the near future (IPCC 2007). In addition, the frequency and intensity of severe weather events within the last decade has been raising concerns about how to adapt to future climate conditions (Spittlehouse and Steward 2003; IPCC 2007). This modeling approach illustrated the 2 climate variables as being the best predictors of AFB; however, an inaccurate RF model undermines the validity of climate driving AFB within the MRC. Despite the inaccuracy of the model, we suggest that future research still look closer into large landscape variations in weather events, gradients, and possible micro-climate interactions with physiographic factors such as elevation, and associated effects on forested systems.

The methods used for this investigation quantified AFB on a large landscape scale; however, the results represent the best estimates of AFB from a model with very poor predictive power. The introduction of errors is likely due to many reasons such as estimating biomass from stem diameter, and region-specific allometric equations (i.e. AFB equation from Jenkins et al. (2003)). Errors associated with ground inventory will remain uncertain until more accurate coefficients (from destructive sampling) or more regionally accepted allometric equations are developed (Saatchi et al. 2011). There was also great uncertainty in AFB estimates at the pixel level as described by the extremely

low coefficient of determination for both the historic and current models. This is likely due to the scale and types of variables used to estimate AFB. We recommend that future research using the RF methodology with FIA inventory data should conduct multiple analyses using a greater variety of data that are more generalized (i.e. averages at the ecological subsection scale) because the 30m resolution datasets used here were not sufficient in capturing any variation in estimating AFB.

The best estimates of AFB were achieved when we summarized AFB estimates based on ecological subsections, which resulted in statistically relevant estimates that were validated based on FIA summaries of AFB at the same scale. Thus, using ecological subsections as the focus for future investigations to understand abiotic controls may be most appropriate, especially for ecosystem management activities.

### **3.4.5 The Role of Disturbances**

Even though climate variables were selected as the top estimators of AFB in the RF model, we cannot ascertain any relevant information from the rankings in the variable importance plots due to the very poor performance of the RF model. The poor performance may be due to the types of variables used. We were constrained by the total number of variables extracted by FIA personnel and consequently, could not attempt to run different models with different sets of variables. We believe that changing the types and numbers of variables within the RF model would have produced more accurate estimates of AFB and now we must consider what types of different variables may be used to produce better RF estimates, such as disturbances.

The areas of estimated AFB that were considered the least changed (i.e. small percent changes between historic and current AFB) occurred in the southern portion of

the study area (Ozark region), which indicates that this area was the least disturbed as compared to the alluvial floodplains of the Mississippi and Missouri Rivers; where the largest percent changes between historic and current AFB occurred. Small amounts of disturbance occur within all regions each year, but we were not able to account directly for such mortality, whether due to flooding, harvesting, disease, insects, drought, or fires.

Moreover, some of the FIA plots could have been used even though recent harvesting may have taken place. Monitoring forest disturbances is challenging for inventory programs like the FIA because of temporal issues in collecting forest information. FIA data is collected in 5-year increments, which likely results in missed disturbance events. Our investigation only considered 2 time periods where the percent changes were calculated between historical (1800s) and current (approximately 2011, which was the last year for cycle 6 in the FIA database). Not accounting for disturbances (i.e. adding an additional predictor variable to the RF model) within this FIA timeframe likely caused low coefficient of determination in estimating AFB.

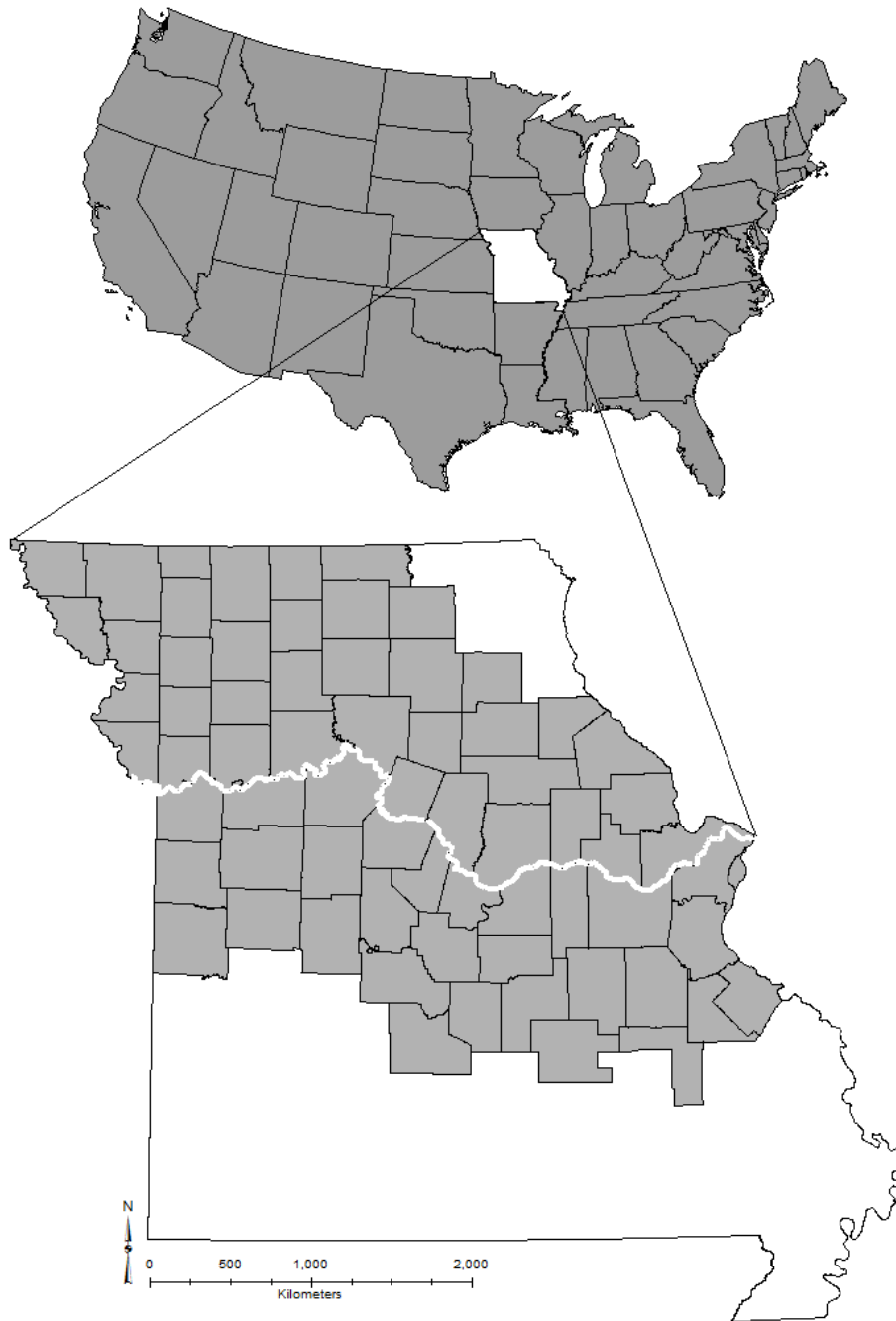
### **3.5 CONCLUSION**

This study attempted to quantify and spatially identify where and how much aboveground forest biomass has changed using a RF approach based on historical and current forest inventories. The AFB estimation results were inconclusive due to poor model accuracy. Variable importance ranking is an important component of the RF procedure, but due to the very low coefficient of variation for both historic and current models, the variables ranked best for estimating AFB should be considered with great caution. Nonetheless, estimating AFB at a 30m resolution using RF was a critical step towards understanding what an appropriate scale should be for determining the dynamics of AFB along major

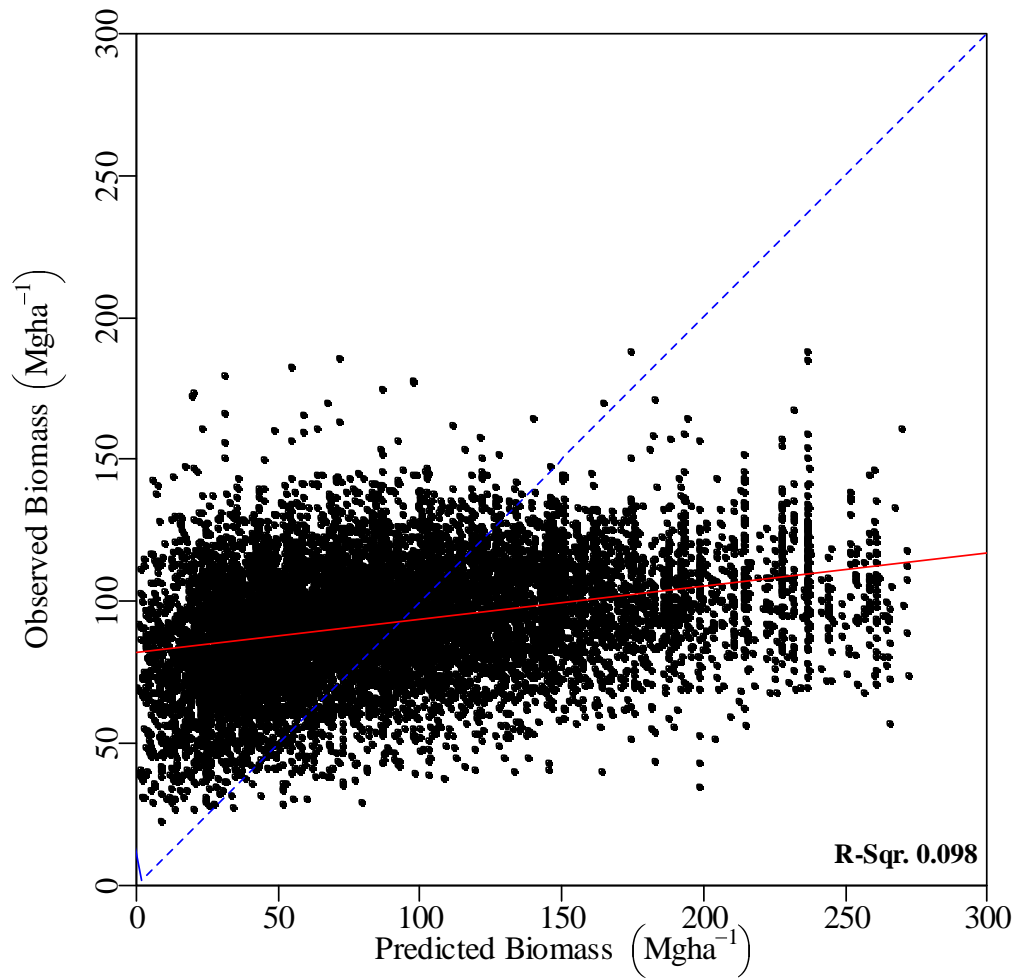
river corridors and determining the type of variables necessary to develop a successful model. Disturbance and age-structure are suggested to be critical parameters to include in future estimates of AFB within the MRC.

There were several constraints administered by FIA personnel due to confidentiality agreements, and extremely limited the scope of this investigation by control the number variables used in the RF model and restricting any spatial information about FIA plot locations. Overall, the modeling procedure was simple, produced easily interpretable output, but was ineffective in finding any relationships between AFB and the chosen predictor variables.

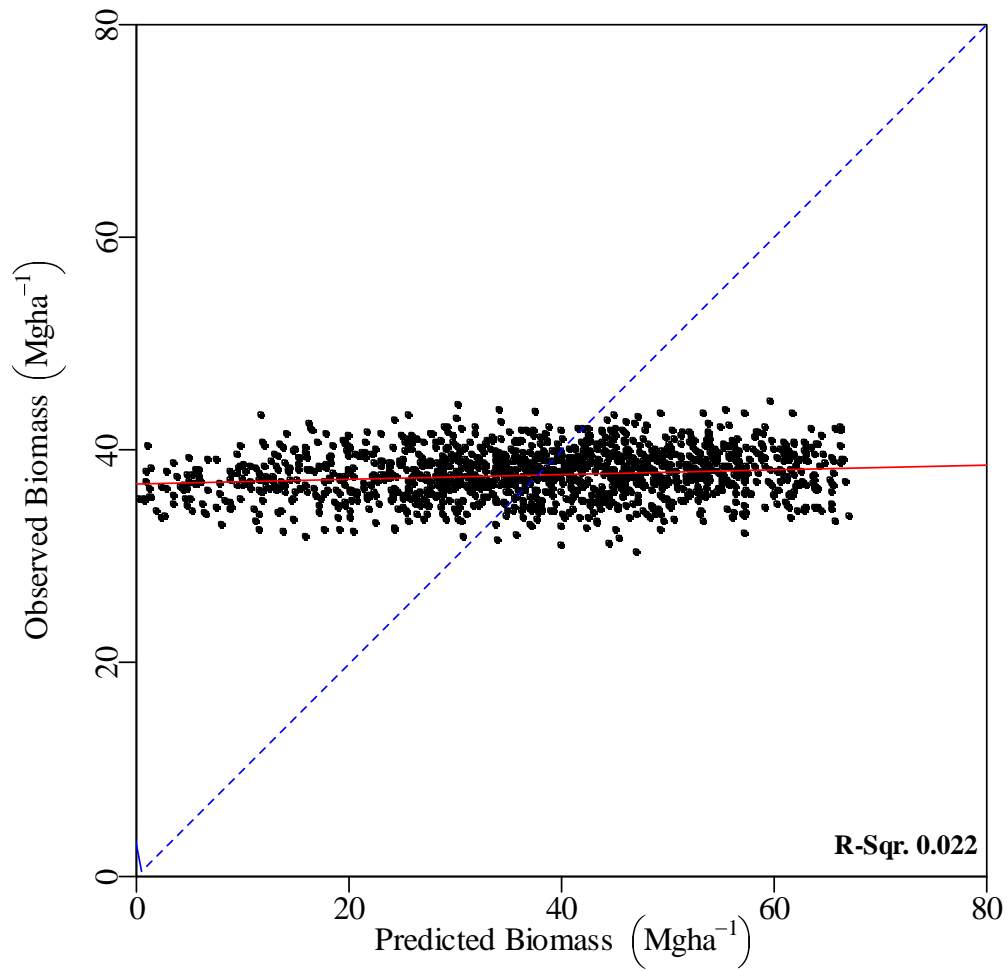
Although biases have been identified in GLO records (Hanberry et al. 2011), they represent some of the best data available for investigating historic patterns of AFB (Delcourt and Delcourt 1996) and are noted to be good baseline measures for comparing current AFB values. The variable importance output is a critical factor in the RF model that provides essential information in identifying contributing factors that best explain patterns of discrepancies in AFB. The final model illustrates where discrepancies exist across the landscape and suggests the primary controlling factors for these discrepancies; however, there is still an incomplete understanding of all the mechanisms and processes that drive distributions of AFB (Boisvenue and Running 2006; Levy et al. 2004), especially due to the poor performance of the RF model. Improving model performance is a critical issue and any information derived from these AFB estimates should be used with great caution.



**Figure 3 - 1: The Missouri River corridor located in Missouri, USA. The corridor was constructed based on a 75 mile buffer centered on the Missouri River. The entire study region is approximately 106,000 km<sup>2</sup> and contains 69 counties dissected by the Missouri River.**

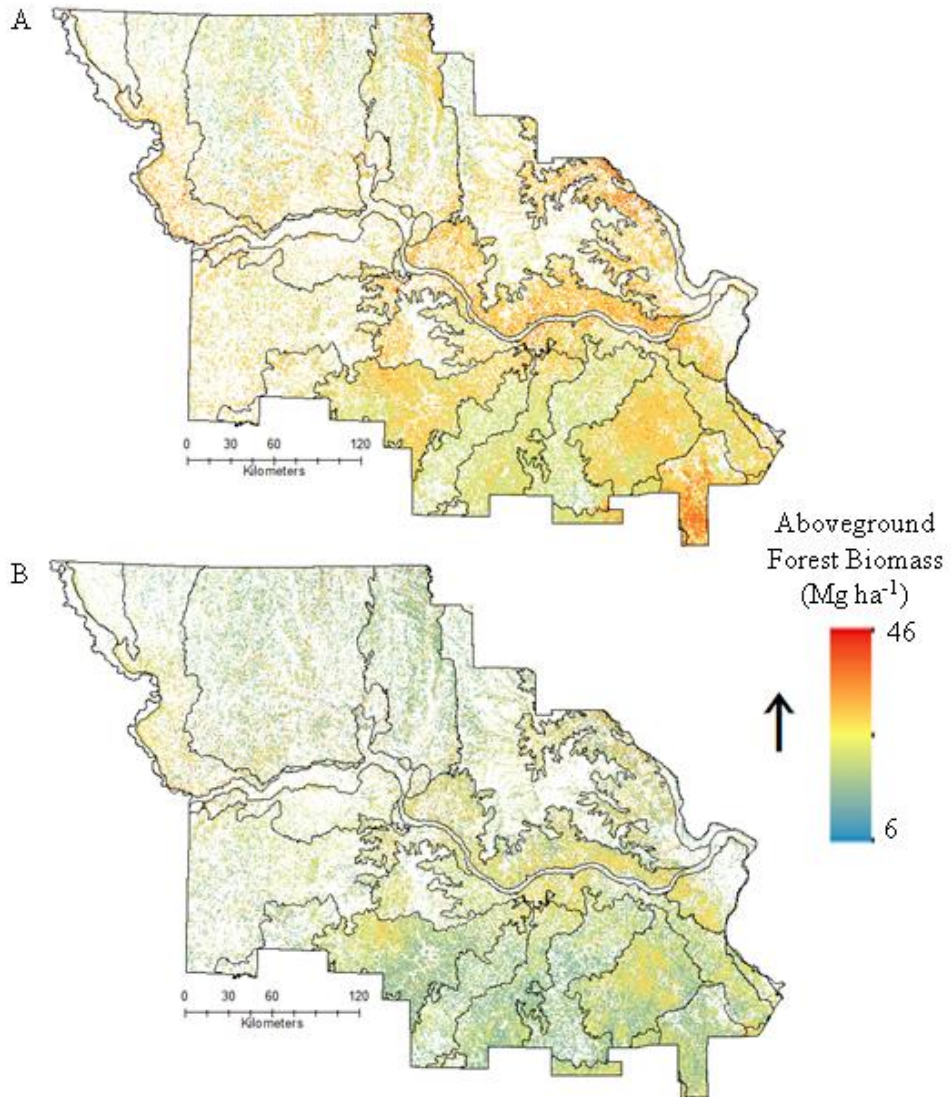


**Figure 3 - 2: Observed and predicted aboveground forest biomass model estimates from the Historic RF model where the red line indicates model fit while the blue line indicates the 1:1 slope.**

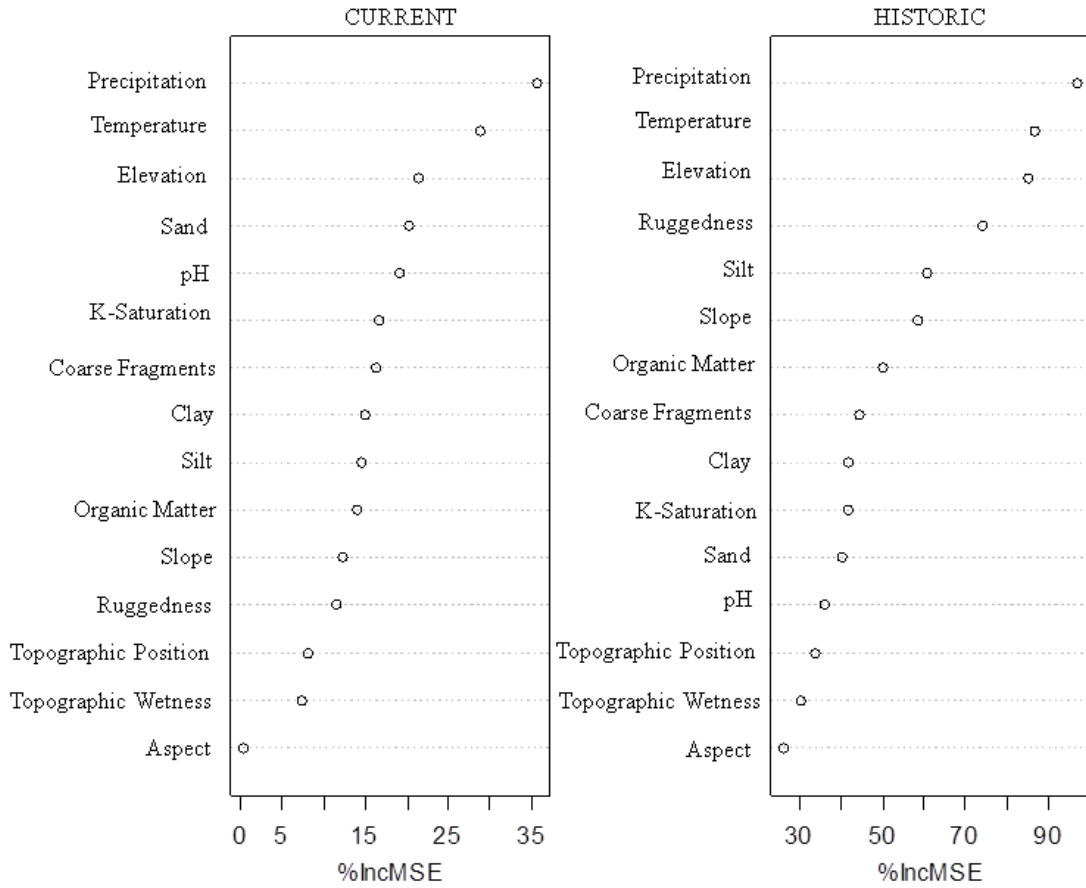


**Figure 3 - 3: Observed and predicted aboveground forest biomass model estimates from the Current RF model where the red line indicates model fit while the blue line indicates the 1:1 slope.**

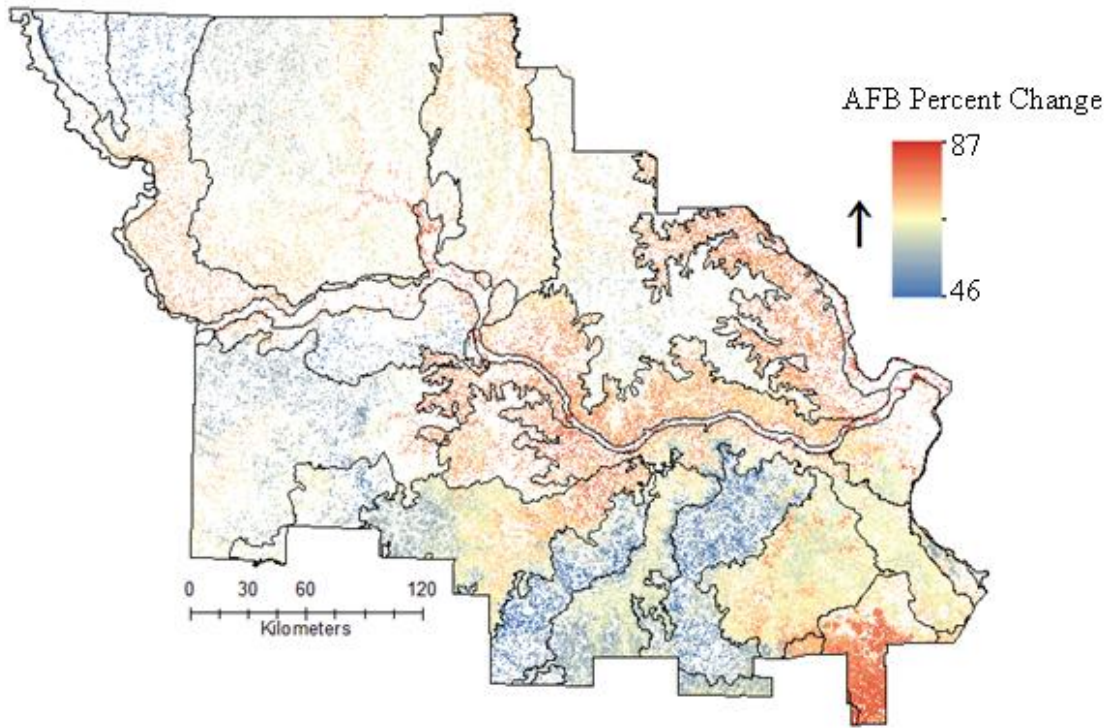




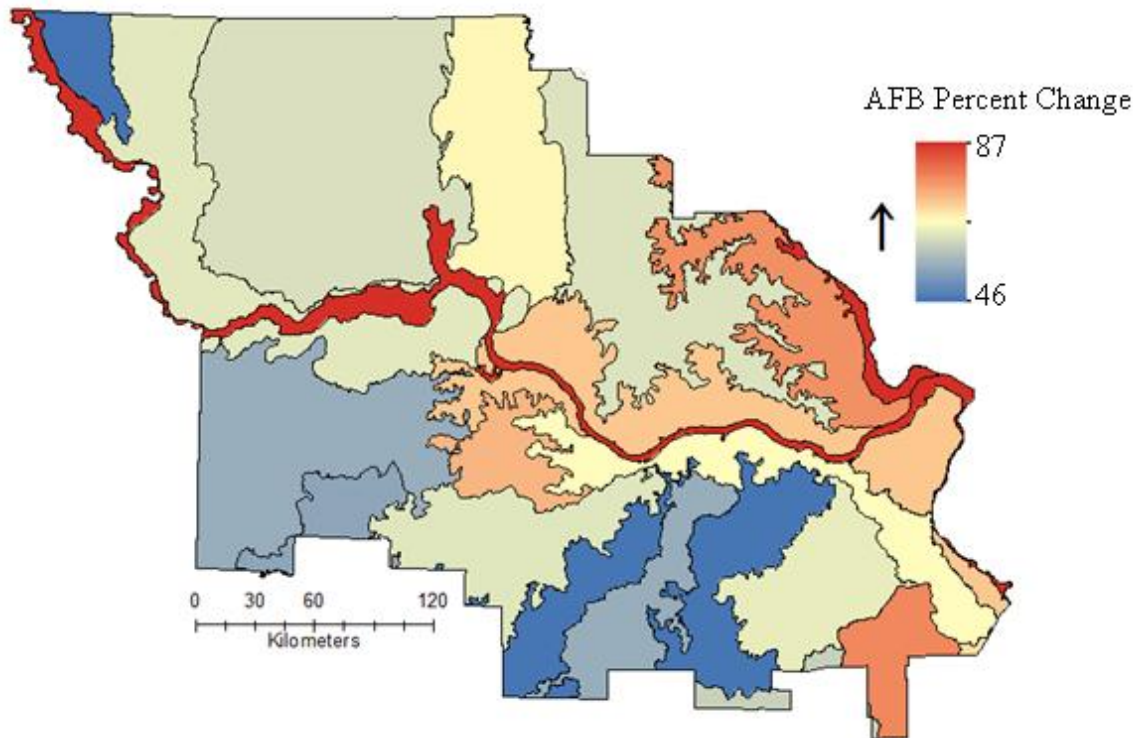
**Figure 3 - 4: Estimated total aboveground forest biomass (Mg) for each forested pixel (30m) for (A) Historic and (B) Current.**



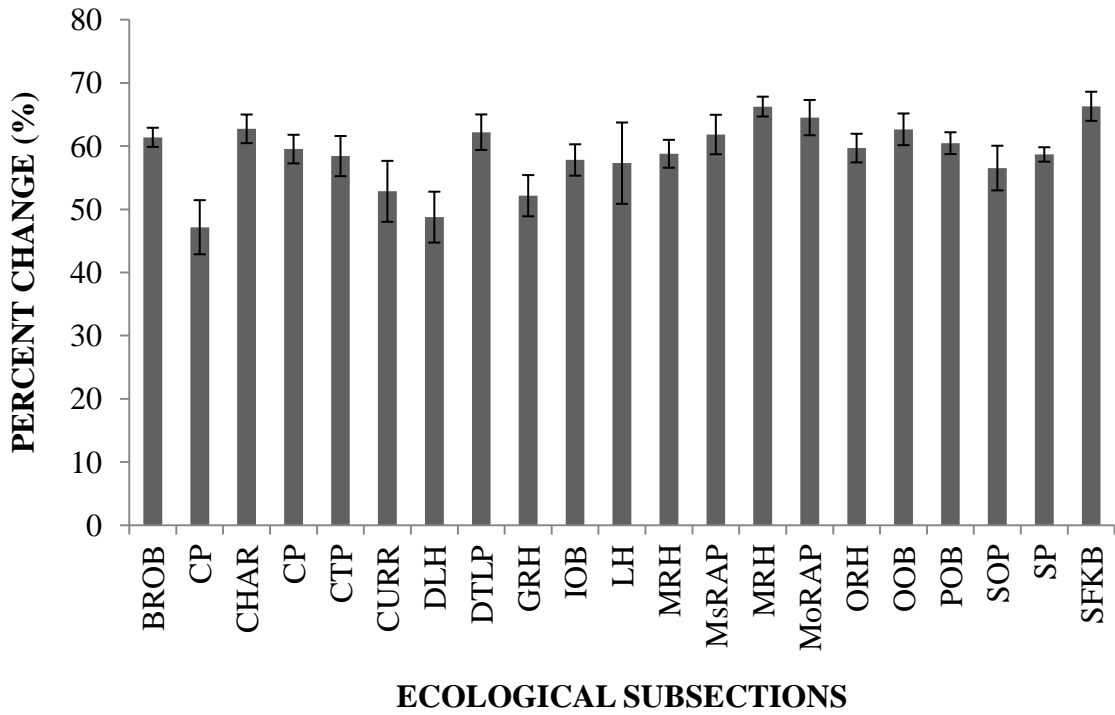
**Figure 3 - 5: Variable importance plots for current and historic aboveground forest biomass estimates generated from Random a Forest regression procedure.**



**Figure 3 - 6: Pixel-level percent change values between historic and current aboveground forest biomass in the MRC.**



**Figure 3 - 7: Average percent change between historic and current AFB estimates summarized by ecological subsection.**



**Figure 3 - 8: Summarized percent changes with standard errors between historic and current aboveground forest biomass estimates generated from a Random Forest procedure.**

**Table 3 - 1: Summarization of GLO and RF AFB estimates.**

SUBSECTION	AREA (ha)	HISTORIC		CURRENT		Percent Change
		TOTAL (Mg)	AVG (Mg ha <sup>-1</sup> )	TOTAL (Mg)	AVG (Mg ha <sup>-1</sup> )	
Black River Ozark Border	5.9.E+03	1.5.E+06	251.8	5.8.E+05	96.6	-61.62
Central Plateau	3.6.E+05	6.5.E+07	179.1	3.3.E+07	91.4	-48.93
Chariton River Hills	2.0.E+05	4.7.E+07	240.8	1.7.E+07	88.2	-63.35
Cherokee Plains	3.9.E+04	9.0.E+06	228.3	3.6.E+06	91.4	-59.96
Claypan Till Plains	1.3.E+05	3.0.E+07	225.0	1.2.E+07	91.2	-59.49
Current River Hills	5.9.E+04	1.3.E+07	213.4	5.6.E+06	95.4	-55.29
Deep Loess Hills	1.0.E+04	2.0.E+06	195.1	1.0.E+06	97.4	-50.06
Dissected Till and Loess Plains	3.0.E+05	7.3.E+07	240.4	2.7.E+07	88.5	-63.17
Gasconade River Hills	2.4.E+05	4.7.E+07	198.3	2.2.E+07	92.8	-53.23
Inner Ozark Border	2.6.E+05	6.1.E+07	230.2	2.5.E+07	95.5	-58.52
Loess Hills	1.6.E+05	3.8.E+07	240.2	1.5.E+07	94.2	-60.76
Meramec River Hills	3.6.E+05	8.3.E+07	233.6	3.4.E+07	95.0	-59.33
Mississippi River Alluvial Plain	1.3.E+04	3.0.E+06	238.7	1.1.E+06	89.1	-62.67
Mississippi River Hills	1.8.E+05	5.1.E+07	279.5	1.7.E+07	93.5	-66.54
Missouri River Alluvial Plain	4.0.E+04	1.0.E+07	260.6	3.6.E+06	90.6	-65.24
Osage River Hills	3.4.E+05	8.0.E+07	235.4	3.2.E+07	93.6	-60.25
Outer Ozark Border	4.0.E+05	1.0.E+08	263.0	3.8.E+07	96.4	-63.34
Prairie Ozark Border	5.6.E+04	1.4.E+07	241.8	5.3.E+06	94.7	-60.82
Scarped Osage Plains	1.7.E+05	3.8.E+07	221.6	1.6.E+07	93.5	-57.82
Springfield Plain	9.9.E+02	2.2.E+05	220.4	9.0.E+04	90.8	-58.78
St. Francois Knobs and Basins	1.5.E+05	4.2.E+07	282.7	1.4.E+07	93.1	-67.08

Table 3 - 2: Missouri River Corridor ecological subsection descriptions.

<b>Subsection Name</b>	<b>Total Area (ha)</b>	<b>Forested Area (ha)</b>	<b>Proportion Forested</b>
Black River Ozark Border	$7.6 \times 10^4$	$6.0 \times 10^4$	78.4
Central Plateau	$7.7 \times 10^5$	$3.6 \times 10^5$	47.5
Chariton River Hills	$7.0 \times 10^5$	$2.0 \times 10^5$	27.9
Cherokee Plains	$1.7 \times 10^5$	$3.9 \times 10^4$	22.5
Claypan Till Plains	$9.2 \times 10^5$	$1.3 \times 10^5$	14.7
Current River Hills	$6.7 \times 10^4$	$5.9 \times 10^4$	88.1
Deep Loess Hills	$1.6 \times 10^5$	$1.0 \times 10^4$	6.7
Dissected Till and Loess Plains	$1.7 \times 10^6$	$3.3 \times 10^5$	18.0
Gasconade River Hills	$3.6 \times 10^5$	$2.2 \times 10^5$	66.3
Inner Ozark Border	$5.0 \times 10^5$	$2.3 \times 10^5$	53.4
Loess Hills	$99.8 \times 10^5$	$1.6 \times 10^5$	16.3
Meramec River Hills	$4.6 \times 10^5$	$3.6 \times 10^5$	78.3
Mississippi River Alluvial Plain	$8.1 \times 10^4$	$1.3 \times 10^4$	15.6
Mississippi River Hills	$5.0 \times 10^5$	$1.8 \times 10^5$	36.8
Missouri River Alluvial Plain	$4.3 \times 10^5$	$4.0 \times 10^4$	9.3
Osage River Hills	$5.3 \times 10^5$	$3.4 \times 10^5$	64.0
Outer Ozark Border	$8.7 \times 10^5$	$4.0 \times 10^5$	45.8
Prairie Ozark Border	$2.3 \times 10^5$	$5.6 \times 10^4$	24.6
Scarped Osage Plains	$9.4 \times 10^5$	$1.7 \times 10^5$	18.4
Springfield Plain	$2.0 \times 10^3$	$9.9 \times 10^2$	52.4
St. Francois Knobs and Basins	$22.2 \times 10^5$	$1.5 \times 10^5$	68.2

**Table 3 - 3: Statistical summaries for RF Current and RF Historic estimates of AFB.**

<b>MODEL METRIC</b>	<b>CURRENT</b>	<b>HISTORIC</b>
Total AFB Estimates	3.4x10 <sup>8</sup> Mg	8.1x10 <sup>8</sup> Mg
Total Average AFB Estimates	99 Mg ha <sup>-1</sup>	253 Mg ha <sup>-1</sup>
Mean Absolute Error (MAE)	12.48 Mg ha <sup>-1</sup>	42.38 Mg ha <sup>-1</sup>
Root Mean Square Error (RMSE)	15.21 Mg	53.21 Mg
Coefficient of Determination (R <sup>2</sup> )	0.02	0.10



**Table 3 - 4: Paired two sample t-Test for means of average estimates of AFB summarized by ecological subsection between FIA and RF.**

	<i>RF Current</i>	<i>FIA</i>
Mean (Mg ha <sup>-1</sup> )	99.26	94.13
Standard Deviation (Mg ha <sup>-1</sup> )	24.64	8.58
Observations (n)	21	21
Pearson Correlation	0.24	
df	20	
t Stat	0.98	
P(T<=t) one-tail	0.17	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.34	
t Critical two-tail	2.09	

## **CHAPTER 4: Geographic analysis of production suitability for available aboveground forest biomass within the Missouri River Corridor, Missouri, USA**

*Curiosity makes me nervous because I am compelled to do something about it.*  
- Bruno Frohlich

### **4.1 ABSTRACT**

Increasing demand for aboveground forest biomass (AFB) as an alternative energy source has triggered innovative research directions investigating the spatial dynamics of AFB across large landscapes. New geographical methods are needed for assessing landscape-scale distributions of AFB that help to assess the potential of forested ecosystems to provide resources that can support rising bioenergy industries. Additionally, all AFB cannot realistically be used for bioenergy and models must be used to determine how much AFB is actually available and where. Therefore, the objective of this study was to estimate AFB using inventory data from the Forest Inventory and Analysis (FIA) and determine areas where AFB may be most suitable for use in the biofuel industry within the Missouri River corridor, located within Missouri, USA. Total AFB was estimated using a Random Forest (RF) statistical approach where two landscape-scale constraints (forest transition and distances from the Mississippi and Missouri Rivers), were used to theoretically limit total AFB and estimate available AFB. A suitability index was then constructed with a fuzzy logic procedure using total AFB and the constraining factors as input variables to derive a map illustrating areas of AFB potentially most suitable for use in the biofuel industry. The RF model performed very poorly with respect to estimating AFB where total AFB was estimated at  $3.4 \times 10^8$  Mg (N=1100,  $R^2=0.01$ ). Available AFB

was calculated after the two technical constraints were applied to the total AFB estimates. Available AFB was estimated at  $7.2 \times 10^6$  Mg. Aboveground forest biomass suitability was then calculated from the fuzzy logic overlay analysis, which indicated areas along the Missouri and Mississippi River Alluvial Floodplains as the most suitable areas for AFB production. Areas indicated to be most suitable were also compared to soil productivity maps derived from the Soil Survey Geographic information, which also illustrated the river corridors as being most productive, with respect to soil characteristics.

## 4.2 INTRODUCTION

Society depends on forests to provide renewable natural resources that help support human well-being by maintaining vital ecosystem services such as clean water, clean air, biodiversity, and prolonged carbon sequestration. Forest biomass as an energy source can also play an important role in diversifying energy sources for many regions and is an important component for socioeconomic development. Bioenergy derived from aboveground forest biomass (AFB) also has a great potential for increasing the proportion of clean, renewable energy produced nationally to help reach the 36 billion gallon per year mandate set by the United States Energy Independence and Security Act (EISA 2007). In addition to a being a renewable energy source, a reliable baseload of AFB exists at local and regions scales, which can create a positive effect on the local energy industries by increasing employment opportunities for the forestry, shipping, or industrial processing sectors (O'Laughlin 2009).

The increasing demand for forest biomass has consequently triggered concerns about the sustainability of AFB production and associated pressures on forest ecosystems from over-exploitation (Thiffault et al. 2014). For instance, soil productivity could decrease due to increases in physical disturbances on the landscape, which may limit the capacity for some areas to sustain a growing forest (Thiffault et al. 2014). Biomass exploitation is inevitable if diversification in energy sources remains a constituent of sustainability; therefore, assessing the potential in available resources will continue to play an important role in sustainable production of regionally-renewable energy sources (Freppaz et al. 2004; Hoogwijk et al. 2003). Current trends in research are now heading

in the direction of regional, ecosystem-scale assessments of geographic potentials of AFB in an effort to better manage existing and future resources.

A significant part of recent research efforts has focused on the estimation of biomass quantities in conjunction with identifying spatial arrangements of biomass. In many regions, AFB has a dispersed geographical distribution and there are gaining interests in identifying biomass supply, landscape characteristics, and where the greatest potential may be for sustainably extracting AFB. Aboveground forest biomass is geographically dependent on many physiographic and human-induced characteristics, and defining different levels of availability is a critical step to finding suitable locations to focus time and monetary investment into sustainable extraction. Legislation and policies at both the federal and state levels are actively engaged in assessments of the potential role of forest biomass in renewable energy standards and portfolios, but generally have not accounted for fully understanding the capacity of the natural resource required to meet energy demands. Increased production of biomass for energy has the potential to offset substantial use of fossil fuels, but it also has the potential to threaten ecosystem integrity. In addition, the combined effects of federal, state, and local decisions about forest use, management, or product consumption may have unforeseen impacts when accumulated across landscapes. Thus, understanding forest dynamics requires the ability to see forests at multiple spatial scales.

To evaluate resource potential of biomass, systematic approaches are needed to define production potential (i.e. what may be theoretically available versus what is actually available), especially for regions where other kinds of energy are either difficult to extract or biomass is readily available. Previous AFB studies have found that the

magnitude of AFB supply available for use in power generation may be far smaller than previously suggested because regional-scale availability is not often considered (Goerndt et al. 2012). Thus, understanding the geographic distributions of AFB coupled with identifying constraints on AFB is becoming even more important given the mandated increases in alternative energy use. Defining a systematic approach to evaluating regional availability of forest biomass is a critical step for beginning the assessment process.

Several studies have used AFB as a proxy for estimating productivity or predicting shifts in distributions of species due to climate change (Jose and Gillespie 1997, 1998; Zheng et al. 2004); albeit, methods are quite diverse and continually change with the development of new detailed spatio-temporal information. Understanding the geographical distribution of AFB over large landscapes is a necessary process for understanding the dynamic role of biomass in supporting regional energy demands while maintaining and protecting ecological processes. Further developing ways to identify geographic dynamics of AFB is needed to obtain detailed spatial and environmental information for use in local or regional forest management decisions (He et al. 1998). For instance, the Ozark forests and associated bottomland hardwoods found within the Missouri River corridor have dispersed geographical distributions, which has raised the interest of researchers in using Geographical Information Systems (GIS) for the evaluation of the biomass potential and modeling the associated characteristics of biomass availability (Voivontas et al. 2001).

The majority of problems associated with landscape evaluation require methodologies that use powerful tools designed for managing, transforming, and illustrating georeferenced data, such as in a GIS (Nyeko 2012). Spatially conceptualizing

patterns and processes enables better understanding of important characteristics of the phenomenon of interest and provides a framework to evaluate how the natural and human environment is organized and interacts (Vizzari 2011). Spatial suitability modeling with GIS is increasingly used as a technique to identify potential locations where AFB extraction may be best suited given a range of constraints. A GIS is a computer-based system that offers a convenient and powerful platform for performing land suitability analyses (Hieisig et al. 2005) and allocation using an integration of multi-criteria methods, such as fuzzy logic (Lewis et al. 2014).

GIS-based multi-criteria decision analysis can be defined as a process that transforms and combines geographical data and value judgments to obtain appropriate and useful information for decision-making (Perpiña et al. 2013). The most common GIS-based suitability models include multi-criteria evaluation such as Boolean overlay, weighted linear combination (WLC), or fuzzy logic. Boolean analyses draw from classified criterion that use logical connectives (i.e. AND, OR) to create discrete (TRUE or FALSE) boundaries that delineate suitable areas (Perpiña et al. 2013; Lewis et al. 2014; Freppaz et al. 2004; Nyeko 2012). Weighted linear combinations use a binning process to categorize input data where an importance weight can be attributed to each factor, based on statistical relationships or expert opinions. A rigid delineation representing suitable areas is created once all weighted factors are combined. Suitability models based on the fuzzy set theory assigns a continuous membership value to each variable that ranges between 0 and 1, where 0 indicates low membership and values approaching 1 are considered more suitable.

The purpose of this investigation was to generate a regional assessment of the capacity for AFB production along the Missouri River corridor. The first objective of this study was to use historic and current forest inventory information to estimate AFB across the landscape and quantify percent changes, which were then used as input criteria for the second objective, developing a suitability map for AFB production. Theoretical and absolute geographic locations of AFB were quantified using a multi-criteria process focusing on theoretical and available biomass potential. *Theoretical biomass* potential is defined as the total amount of AFB in a region that represents an unconstrained system that shows only the upper bounds of existing AFB, disregarding ownership status or position on the landscape. The *available biomass* potential represents a constrained scenario and is defined as the residual amount of standing biomass after accounting for technical constraints, such as distances to major rivers and probable expansion from existing urban centers.

## **4.3 METHODS**

### **4.3.1 Study Area**

The Missouri River corridor (MRC) is a 106,000 km<sup>2</sup> region that represents an area specific within Missouri that has the resource and infrastructural potential to lead woody biomass-based, alternative energy development (Figure 4-1). The study region ranges in elevations between 540 to 102 meters above sea level where the highest peaks are found in the southern-most region and the lowest are generally along the Missouri and Mississippi River floodplains. The central component of this region is the Missouri River where growing, harvesting, shipping, and processing of woody material for energy are all within a feasible distance to create a positive effect on biomass financial competitiveness.



The surrounding landscape consists of a shifting mosaic of land uses and land types ranging from upland hardwood forests, mixed bottomland hardwood, cropland, pasture, and secondary growth woodlands with extensive past grazing influences (Nigh and Schroeder 2002). Forest cover in the region had been dramatically altered from European settlement in the late nineteenth century and is still slowly recovering. Large diameter oak (*Quercus spp.*) and hickory (*Carya spp.*) species remain along with an increasing understory of maple (*Acer spp.*).

#### **4.3.2 Suitability Criteria**

##### **4.3.2.1 Aboveground Forest Biomass**

This investigation defined aboveground forest biomass (AFB) as the total amount of living organic matter in trees present above the soil surface with a diameter at breast height (DBH) greater than 2.5 cm. AFB data were obtained from the United States Forest Service Forest Inventory and Analysis (USFS FIA) (USDA FIA 2013). USFS FIA is the only program that collects, publishes, and analyzes data on the extent, condition, volume, growth, and depletions of timber from all ownerships on the Nation's forest land (Woudenberg et al. 2010; USDA FIA 2013). The FIA program provides plot-level information of AFB on 5-year cycles, which allows for regional monitoring of forest lands and is suitable for landscape level studies (Moser et al. 2010). The latest cycle of FIA inventory (2006-2011) was extracted by FIA personnel to maintain FIA plot spatial discrepancy.

A total of 20 variables were used in a Random Forest (RF) algorithm to estimate AFB and determine variable importance within the MRC (Table 4-1). Random Forest is a non-parametric, ensemble regression tree method that is an extension of the family of Classification and Regression Techniques (CART) that grows based on random subset

sampling from a pool of variables (Bou langer et al. 2013). Numerous nodes and trees are constructed using boot-strapped samples of the subset data where the final prediction is a weighted average across all regression trees (Breiman 2001; Cutler et al. 2007; Prasad et al. 2004). The number of variables and iterations used in the model is decided by the user (*mtry* and *ntree* parameters in the RF code). Four (*mtry*) variables from the pool of 20 were used for random selection at each node to grow a total of 1,500 (*ntree*) regression trees. The coefficient of determination ( $R^2$ ) was used as the statistical measure for model fit and the mean absolute error (MAE) was used to assess the size of the prediction error per model (Miehle et al. 2006). Ecological subsection boundaries were used to summarize, compare, and validate AFB estimates with observed FIA AFB data (Table 4-2).

Random Forest has the capability of listing variable importance that is based on a permutation process between the original (training: variables used for estimating) and out-of-bag (OOB: variables left out of the random selection) data. When regression trees are generated, a prediction error is generated for both training and OOB data. The differences between prediction errors are calculated, normalized by standard deviation of the differences, and then averaged across all trees to generate a mean square error value for each variable. The importance value indicates how much a variable increases the MSE when randomly permuted, which means that when a variable is less important, it will contribute less to the model resulting in very small changes in MSE (Prasad et al. 2004). Variable importance output from RF can be used to determine which variables best represent the response variable and may be used to narrow down the number of variables sufficient to make parsimonious predictions.

#### 4.3.2.2 Landscape Variables

The 20 variables were chosen based on representative characteristics of the region. Five remote sensing variables were derived from a 30-meter resolution Landsat Thematic Mapper (TM) satellite image captured on June 3, 2003 (prior to failure of the scan line corrector, which creates gaps in reflectance information) and cloud cover was 0% (Table 2-1). The image was acquired from the United States Geological Survey's (USGS) Earth Resources Observation and Science Center (EROS). Normalized Difference Vegetation Index (NDVI) was calculated based on the combination of red and near-infrared bands according to the following equation:

$$NDVI = \left( \frac{NIR - R}{NIR + R} \right)$$

eq. 1

where NDVI is the ratio between the difference (numerator) and summation (denominator) of near-infrared (*NIR*) and red (*R*) spectral bands. The Normalized Difference Water Index (NDWI) was also used because it is a good indicator of healthy vegetation and is based on the following equation:

$$NDWI = \left( \frac{G - NIR}{G + NIR} \right)$$

eq. 2

where NDWI is the ratio between the difference (numerator) and summation (denominator) of near-infrared (*NIR*) and Green (*G*) spectral bands, which are within the water absorption band centered at 1200 nm (Gao 1996). A tasseled-cap transformation of the Landsat reflectance values was performed to extract 3 additional variables:

brightness, greenness, and wetness (TCB, TCG, and TCW, respectively), where recently disturbed forestland exhibit high values of TCB, low TCG, and low TCW in relation to an undisturbed forested area (Healey et al. 2005). All remote sensing transformations and index calculations were performed in ERDAS IMAGINE® 2011.

We used annual mean temperature and precipitation seasonality (i.e. seasonality as measured by the coefficient of variation in precipitation) (BIO\_01 and BIO\_15, respectively) as climate variables to estimate AFB. These two climate variables were acquired from the WorldClim database, which provides freely available global climate data specific for spatial modeling ([www.worldclim.com](http://www.worldclim.com), accessed May 2011). The topography within the study region is dominated by fluvial activity creating heterogeneity in relief and soil characteristics between stream valleys and adjacent major ridges. Six topographic variables were derived from a 30-meter resolution Digital Elevation Model (DEM), which is a cartographic dataset of elevation values. The DEM was obtained from Missouri Spatial Data Information Services where, in addition to elevation, the following metrics were derived in a Geographic Information System (GIS): aspect, slope, topographic wetness, ruggedness, and topographic position. Seven soil variables representative of the top 20 cm of the surface (sand, silt, clay, coarse fragment, pH, hydraulic conductivity, and organic matter) were obtained from the United States Department of Agriculture Natural Resource Conservation Service's Soil Data Viewer v6.0 (NRCS 2011).

A GIS database containing the landscape information of interest, AFB potential (BP), forest transition (FT), and Euclidean Distances to major river corridors (ED) was prepared for application of a Fuzzy Overlay analysis. Initially, the NLCD 1992/2001

Retrofit Land Cover Change Product was used to determine the geographic locations where forest cover had changed. This NLCD product provided land cover change information at the Anderson Level I classification scale, which is designed for regional applications. Euclidean distances were determined for pixel-areas defining percent changes between historic and current AFB values and major river corridors (i.e. the Mississippi and Missouri Rivers).

### **4.3.3 GIS Processing**

#### **4.3.3.1 Fuzzy Membership**

Fuzzy logic models have long been used for multi-criteria decision-making and spatial modeling for application in site selections or suitability assessments (Lewis et al. 2014). Fuzzy systems are constructed using a set of fuzzy functions based on relationships between the inputs and the output of a system (Ocalir et al. 2010). Prior to the overlay analysis, a preprocessing step was conducted to transform the input criteria to fit a 0 to 1 scale based on a specified function that establishes a membership for each value. The 4 variables used were: 1) Biomass quantity (BQ), 2) Biomass percent change (BP), 3) Forest transition (FT), and 4) Euclidean distance (ED) (Figure 4-2). The 0 to 1 scale indicates the strength of membership to a specific set where locations that represent low membership are assigned 0 whereas locations that have high membership receive a value of 1 (i.e. a greater possibility of representing that set). There are 8 functions to choose from (FuzzyGaussian, FuzzyLarge, FuzzyLinear, FuzzyMSLarge, FuzzyMSSmall, FuzzyNear, and FuzzySmall) that define continuous functions that each captures a different type of transformation, which depends on the desired effect.

For this investigation, a negative relationship was assigned to the variable BP because a negative percent change between historic and current biomass was defined as

having a greater potential for current AFB to reach historic conditions. Therefore, the FuzzyMSSmall function was used to reclassify the values. The FuzzyMSSmall function defines the membership of the data using a function that is based on the mean and standard deviation, assigning smaller values closer to 1. A negative relationship was also assumed for ED because greater distances from the Missouri and Mississippi River indicate minimal potential for using AFB at those distances. The FuzzyMSSmall function was also used for ED membership transformation. A positive relationship was assumed for both FT and BQ because negative values indicated a net loss of forest cover or low amounts of AFB, respectively. The FuzzyMSLarge function was used for membership transformation. The Fuzzy Large function defines the membership using a function based on the mean and standard deviation of the actual input raster values, where larger values from the raw data attain a membership closer to 1.

#### **4.3.3.1 Kernel Density**

A kernel density estimate (KDE) was calculated only for the forest transition parameter. This kernel density function was applied after the fuzzy membership (0- 1 scale) was processed. The KDE was applied to produce a smoothed surface that represented a gradient of potential forest change. The goal of creating a smoothed surface from forest transition was to attempt to account for potential urban sprawl and create a very conservative mask to select only forest pixels that would be certainly free (for now) from transition to urban pixels. To do this, we selected pixels that had values generated from the KDE that were greater than or equal to 0.90. This threshold was chosen to have a conservative estimate of forest pixels possibly available for use and resulted in a spatial mask that was used for constraining the calculation of total AFB by “leaving out” only the pixels regarded as available.

A KDE produces smooth surfaces by using a quadratic function where the greatest value is attributed to the center of a defined geographic window or area, and decreases based on a user-specified search window distance. We set the cell resolution for the KDE operation to 1,500 m with a search radius of 10,000 m, which equates to an area of approximately 314 km<sup>2</sup>. These thresholds were chosen to account for a very conservative estimate of forest transition to urban pixels. Further, the resolution and search radius that we used in this study was an attempt to fulfill the goal of finding a balance between observing the spatial distributions of forest area throughout the study region and achieving a relatively smooth surface that represents conservative distributions of forest change. If we choose a smaller cell resolution and search radius, the KDE surface would be negligible because only a few pixels would have changed, and we thought that this would be an underestimation of the actual potential of urban sprawl. As a result, the intention of using a larger radius was to smooth the spatial variation and illustrate a general trend over the study area that was tangible and usable, rather than highlighting more localized effects of each data point that would occur at singular pixels.

#### **4.3.3.1 Fuzzy Overlay**

To investigate the relationships between all four factors (BQ, BP, FT, and ED) with respect to suitability for increased AFB production, a final suitability map was generated using a fuzzy overlay application. Five specific fuzzy logic operators are available in GIS and are considered to operate similar to Boolean logic (OR, AND, PRODUCT, SUM, and GAMMA) (Boroushaki and Malczewski 2010). For instance, the fuzzy OR function is a logical union where maximum values control the output data. The fuzzy AND function represents a logical union where minimum values control the output data. The PRODUCT is a decreasing function while the SUM is an increasing function. The

GAMMA function is product operation between PRODUCT and SUM functions raised to the power of gamma. The Gamma overlay function was used for this analysis because 1) It has the potential to be a good compromise between the increasing effects of SUM and the decreasing effects of PRODUCT, and 2) It has the best possibility to fit a forest productivity map generated from SSURGO information. The GAMMA function uses the following equation:

$$\text{Fuzzy GAMMA value} = (\text{Fuzzy SUM})^\gamma * (\text{Fuzzy PRODUCT})^{1-\gamma}$$

eq. 3

where the GAMMA function multiplies the *fuzzy sum* by the *fuzzy product* to the power of *gamma* ( $\gamma$ ); 0.7 was used as the gamma value in this investigation as a median value to try to achieve a compromise between the increasing effects of the SUM function and the decreasing effects of PRODUCT function, and fit an existing spatial map of forest generated by SSURGO data. Gamma values of 1 are closer to the fuzzy sum whereas gamma values approaching 0 are closer to PRODUCT. Values in between allow for a combination of evidence between two extremes and possibly different than that from other fuzzy functions. The GAMMA function is considered a judicious choice that offers a flexible compromise between the increasing tendencies of SUM and decreasing tendencies of PRODUCT (Pradhan 2011).

## 4.4 RESULTS

### 4.4.1 Potential Biomass

Historic AFB was estimated using 10,857 inventory points where each location had  $\geq 2$  trees sampled. Total historic AFB within the MRC was estimated at  $8.1 \times 10^8$  Mg with an average biomass per ecological subsection of  $234 \text{ Mg ha}^{-1}$  and a standard deviation of 26



Mg ha<sup>-1</sup>. The least squares regression between observed and modeled historical AFB had an adjusted R<sup>2</sup> = 0.094. Current AFB was estimated using a total of 1,100 inventory plots. The total AFB for the current model was estimated at 3.4 × 10<sup>8</sup> Mg with an average biomass value per ecological subsection of 98 Mg ha<sup>-1</sup> and a standard deviation of 3 Mg ha<sup>-1</sup>. The least squares regression between observed and estimated biomass had an adjusted R<sup>2</sup> = 0.022. The average AFB for the current model estimated by RF was not statistically different from FIA estimates according to a Student T-test (p=.45, α = 0.05).

Potential AFB was calculated per ecological subsection as the percent change between estimated historic and current data (Figure 4-3). The range in percent changes across the entire site was -83 – 31%. In all cases, the summarized mean historic AFB levels were greater than current AFB levels which resulted in negative percentages. The top 3 subsections where there was the greatest, negative percent change was found in the St. Francois Knobs and Basin (-66%), Mississippi River Hills (-66%), and the Missouri River Alluvial Plain (-64%). The smallest average percent change was found in the Central Plateau at -46%. The greatest differences in biomass (> 65 Mg/ha) and percent changes (> 60%) were illustrated along the Missouri River Alluvial Plain and Mississippi River Hills subsections. Minimal differences (< 45 Mg/ha) and percent changes (< 50%) occurred within the Central Plateau subsection, along the southern portion of the study area that is part of the Ozark region, and also along the Deep Loess Hills subsection, found in the upper northeast corner of Missouri.

Variable importance for both the historic and current model estimates derived by RF (Figure 4-4) ranked climate (Annual Mean Temperature and Precipitation Seasonality) and elevation as the top 3 predictors, respectively. The mean square error

(MSE) values for the historic model ranged between 0 – 90%, whereas the current model exhibited a lower range in MSE (0 – 35%). Subsequent variables in the current model primarily consisted of soil factors, whereas in the historic model, the remaining top-ranked variables consisted of more topographic characteristics as opposed to soil.

#### **4.4.2 Available Biomass**

We calculated total and average AFB for the available regions specified by the KDE mask (Figure 4-5). The total and average available AFB resulting after the constraints was quantified at  $7.2 \times 10^6$  Mg and  $50 \text{ Mg ha}^{-1}$ , respectively, per ecological subsection (Figure 4-7 and Table 4-2). The areas covered by the mask were considered available and used as the specific area for the suitability analysis. Total available AFB per ecological subsection resulted in an overall decrease in AFB compared to current estimates where the greatest departures were found in the Outer Ozark Border (OOB), Osage River (ORH), and Meramec River Hills (MRH) subsections. With respect to forest transitions, there was an average, negative percent change in forest area (-1.5%), where the Mississippi and Missouri River Alluvial Plains had the greatest percent changes in forest area (-18 and -5%, respectively) (Figure 4-6).

#### **4.4.3 Suitability**

Input variables were standardized using a fuzzy membership function that transformed the values to fit between 0 and 1: biomass quantity (0-0.8), Euclidean distance (0.3-1), forest transition (0.03-1), and percent change (0.09-1) (Figure 4-8). The most suitable areas for AFB production were determined to be within the Loess Hills, Missouri and Mississippi River Alluvial Plains where the suitability index was greater than 0.9 (Figures 4-9, 10, and 11).

### **4.5 DISCUSSION**

#### **4.5.1 Aboveground Forest Biomass Potential**

This study illustrated a plausible, methodological approach to determining AFB suitability within the Missouri River corridor. The results of AFB suitability were based on summarized values that were constrained by spatial restraints that consisted of biomass quantities (historic and current estimated by Random Forest), percent changes (between historic and current AFB estimates), forest transition, and Euclidean distances (from forested pixels to Missouri or Mississippi Rivers).

The RF procedure for predicting AFB resulted in poor estimates. This may be due to the combination of resolutions at which we attempted to acquire AFB predictions (30 meter resolution) and the scale of the predictor variables. There may be a disjunction between the scales where estimating AFB at 30-m resolution was an ineffective strategy using FIA data. A relationship between observed and predicted AFB estimates was formulated after summarizing AFB estimations at the ecological subsection scale. An improved relationship was formulated most likely because the scales of FIA data and predictor variables were only adequate at this larger landscape-scale. Therefore, the ecological subsection scale may be the most appropriate for estimating AFB using observed, FIA information.

Stratifying the study region by ecological subsection provided additional insight into the geographic characteristics of AFB potential as a function of two points in time (i.e. 1800s and 2011). This temporal scale is limited because it is only two points in time; however, establishing a baseline measure of AFB and compared it with current estimates provided information on how the recovery rate of forests across the landscape. Additionally, the ecological subsection boundaries delineate critical areas where landscape scale analyses are most useful for forest planning decisions, especially when

multiple management objectives exist, such as along the Missouri River corridor (Galindo-Leal and Bunnell 1995; Lars et al. 2012; Bardhan and Jose 2012)

Results indicated that the greater percent changes occur along the Missouri and Mississippi River Alluvial Plains, as well as in a subsection near the Ozark Region (St. Francois Knobs and Basin). The St. Francois region is characterized by igneous knob outcrops and sedimentary basins. The two alluvial plains, however, consist of a much different morphology where a greater difference between historic and current vegetation exists. This may be due to changes in disturbance regimes that have altered AFB densities such as harvest, floods, windthrow, or agriculture. Agriculture continues to persist in this region and is the dominate land-use type (Hanberry, Kabrick, et al. 2012; Nigh and Schroeder 2002).

#### **4.5.2 Aboveground Forest Biomass Constraints**

The constraints used in this study were a function of limitation in the available number of forested pixels used in the analysis. We used the constraints to represent areas most probable for use in analysis and quantifying conservative estimates of AFB (i.e. where biomass may be used for production, as opposed to simply quantifying total biomass). The constraints were considered simple and easily accessible via GIS processing, and we believe they effectively captured particular physical aspects of the landscape that could potentially limit the use of particular forested areas. For instance, forest transitions from urbanization and agricultural expansion were taken into consideration because Missouri has many rapidly sprawling urban areas (Sunde et al. 2014). The most prevalent forest transitions were found in the Ozark, western (Kansas City), and eastern (St. Louis) regions of the study area.

The Ozark region, in particular, was found to have the greatest densities of AFB and there was still a considerable amount of AFB technically available even with the constraints applied. The final constraint, Euclidean (straight-line) distance from the Missouri and Mississippi Rivers, was used to identify areas most conducive to transporting biomass to processing facilities along the rivers. The ‘straight-line’ measure was used simply to gain a broad perspective on the regional variability in AFB locations with respect to river boundaries and did not consider structure and type of road networks. A separate road network analysis or GIS-based Manhattan process (distance measures following only axis-aligned directions, which is more representative of actual road networks) was out of the scope of this investigation; albeit would be important for local-scale cost analyses or bioenergy facility placement.

#### **4.5.2 Aboveground Forest Biomass Suitability**

Evaluation of AFB production suitability was conducted according to estimates of available AFB from a RF model and a series of constraints that exclude particular areas of AFB from exploitation. The final map output exemplified the Mississippi and Missouri River floodplains and associated Missouri River Alluvial Plains as the most suitable locations for AFB production. The modeling procedure was conducted in an effort to fit the results to an existing map of forest productivity derived from SSRUGO information (Figure 4-12). The forest productivity map shows the floodplains of the Mississippi and Missouri Rivers as having good productivity, which is a product of fertile soils that are critical for plant growth. The results of the suitability map did not align perfectly with the forest productivity map from SSURGO (FIGURE 4-12); however, based on the attributes used in the criteria analysis, the Mississippi and the upper portion of the Missouri River were considered most suitable. This portion of the results is similar to those from the

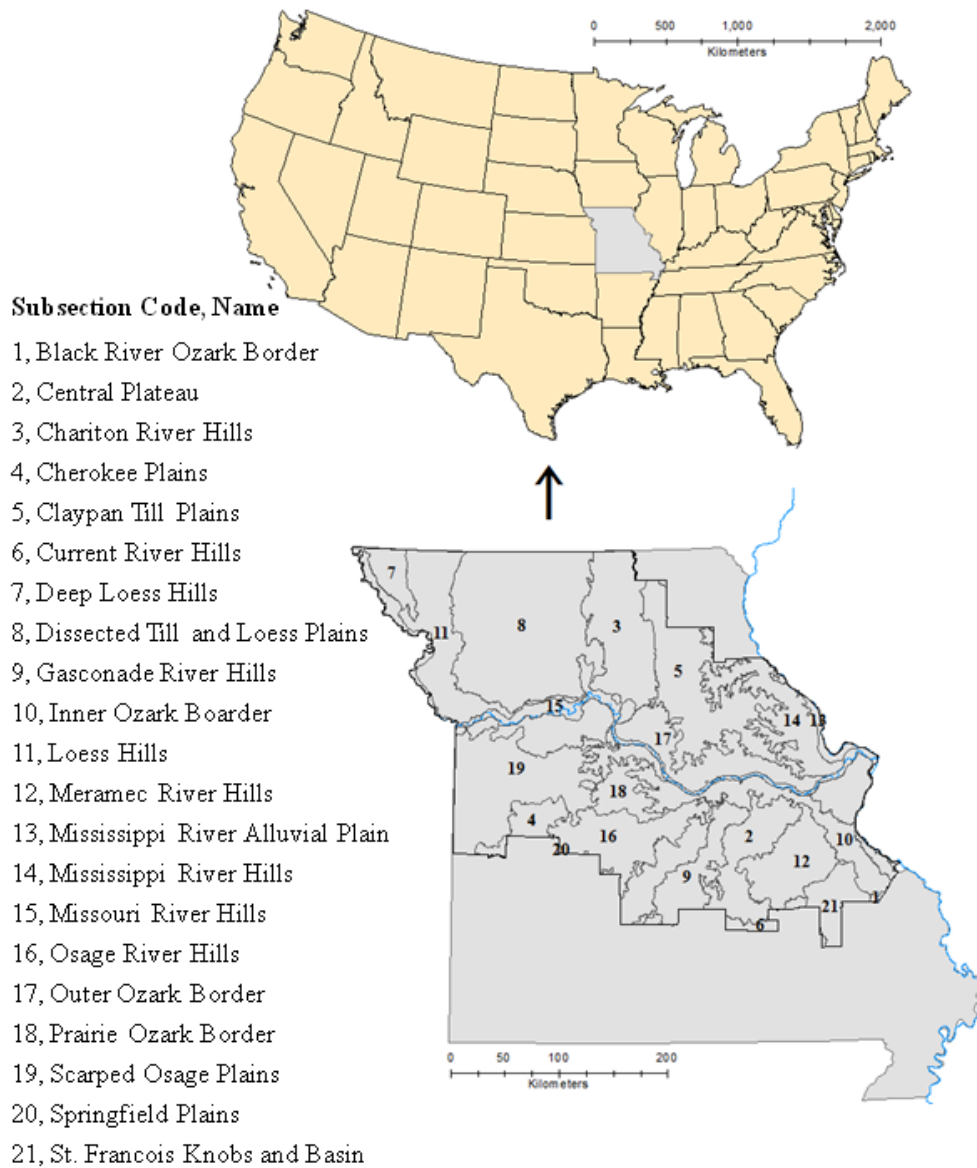
SSURGO forest productivity map. An increase in forestland area has been observed specifically along the Mississippi River where forest expansion has occurred, reverting agriculture and barren lands to forest, except in the most productive agricultural zones (Yeo and Huang 2013).

The fuzzy logic process was an important GIS application that was capable of facilitating a coarse-scale decision process using multi-criteria. The applicability of this process is minimal due to the poor estimates of the AFB with the RF model. This was due to the inability of estimating AFB at fine (30 meter) resolution with coarse-scaled, FIA information. Thus, the given results should be used with great caution. Future research should focus on improving models that estimate AFB. This may be accomplished by concentrating on the types and number of variables used for estimating AFB (we were only allowed a maximum of 20). Other successful studies have used  $\geq 30$  variables (Prasad et al. 2004; Falkowski et al. 2009).

This investigation attempted to use a fuzzy logic approach using multiple criteria to understand spatial dynamics of AFB. Because of the poor AFB estimates, the suitability model cannot be used to ascertain any information regarding AFB; however, the methodology remains appropriate for constructing a multiple criteria framework, but will require a more accurate estimate of AFB to improve this model. Multi-criteria analyses are widely used to determine the influence of particular variables in group decision processes (Borouhaki and Malczewski 2010) such as land susceptibility(Pradhan 2011), plant growth (Lewis et al. 2014) facility placements, or land-use assessments (Lars et al. 2012). To make this approach feasible and applicable, the uncertainties in estimating the focal parameter, AFB, requires further investigation.

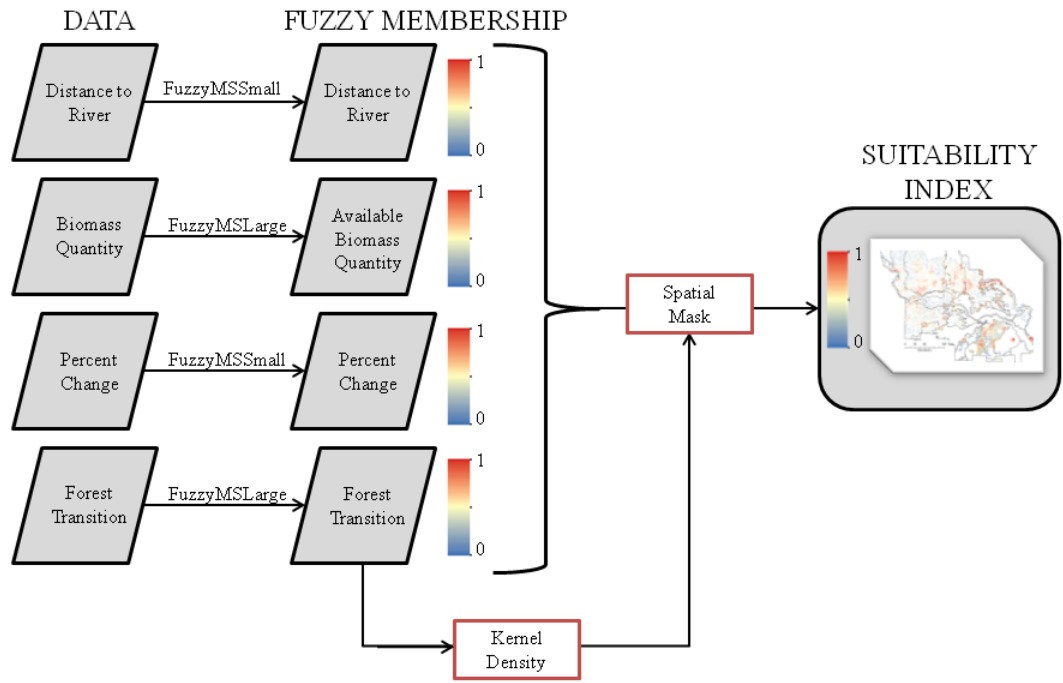
## **4.6 CONCLUSION**

A GIS-based analysis was performed to assess both the theoretical and available AFB potential at a large landscape scale. The results illustrated that the greatest changes between theoretical and available AFB potential exist along the Missouri River floodplain, which also corresponds with production suitability generated from the Fuzzy Logic Overlay analysis. The AFB suitability maps presented in figures 4 and 5 shows the spatial distributions of AFB by ecological subsection and are approximately consistent in comparison with forest productivity maps derived from SSURGO. In this study, we sought to develop a multi-criteria indicator method to illustrate AFB production suitability, which could be used to help support future decisions about sustainable forest production at a regional scale.

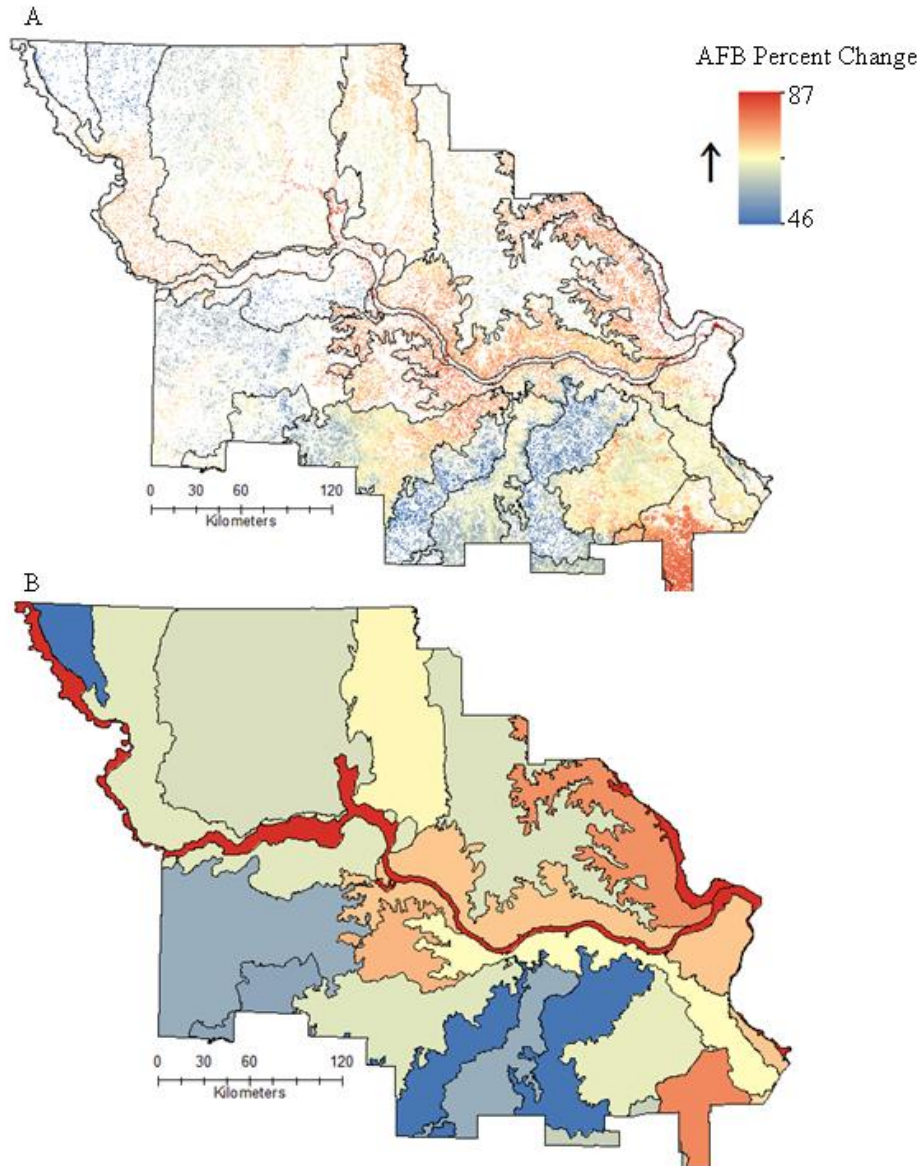


**Figure 4 - 1: The Missouri River corridor is approximately 106,000 km<sup>2</sup> covers 21 counties and is divided into 21 ecological subsections.**

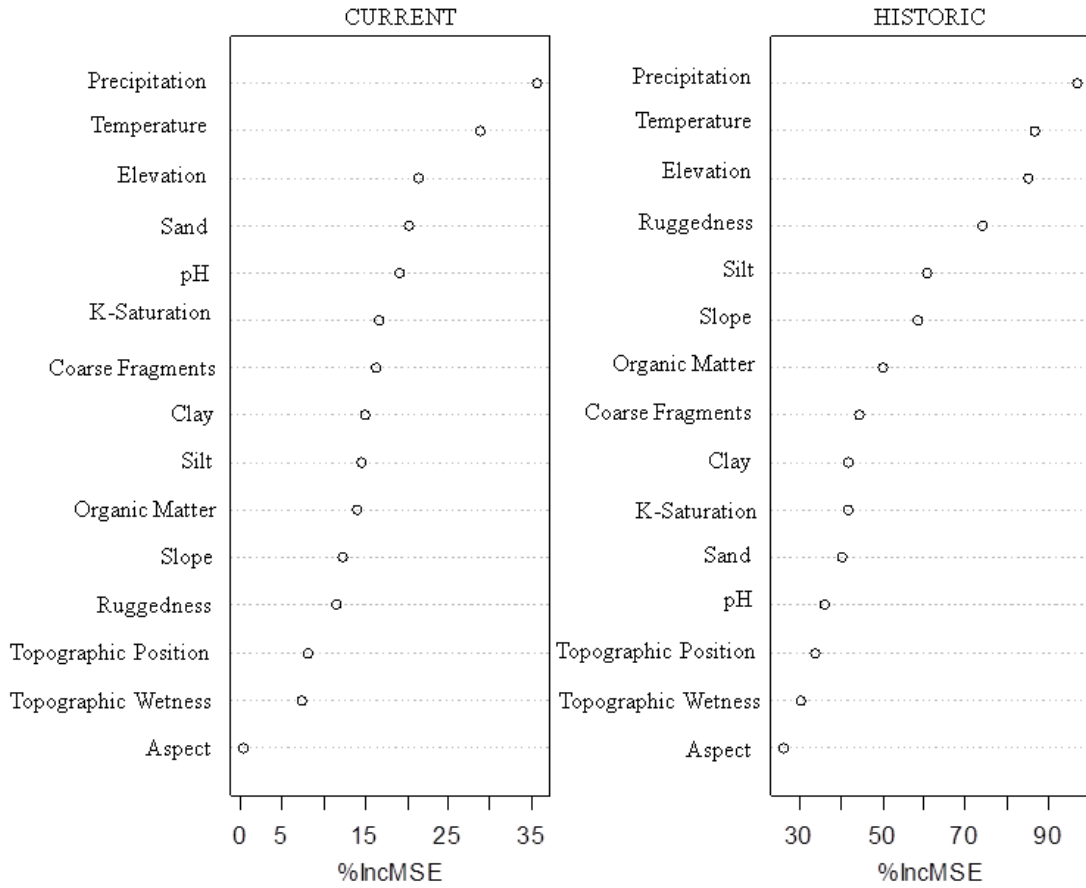




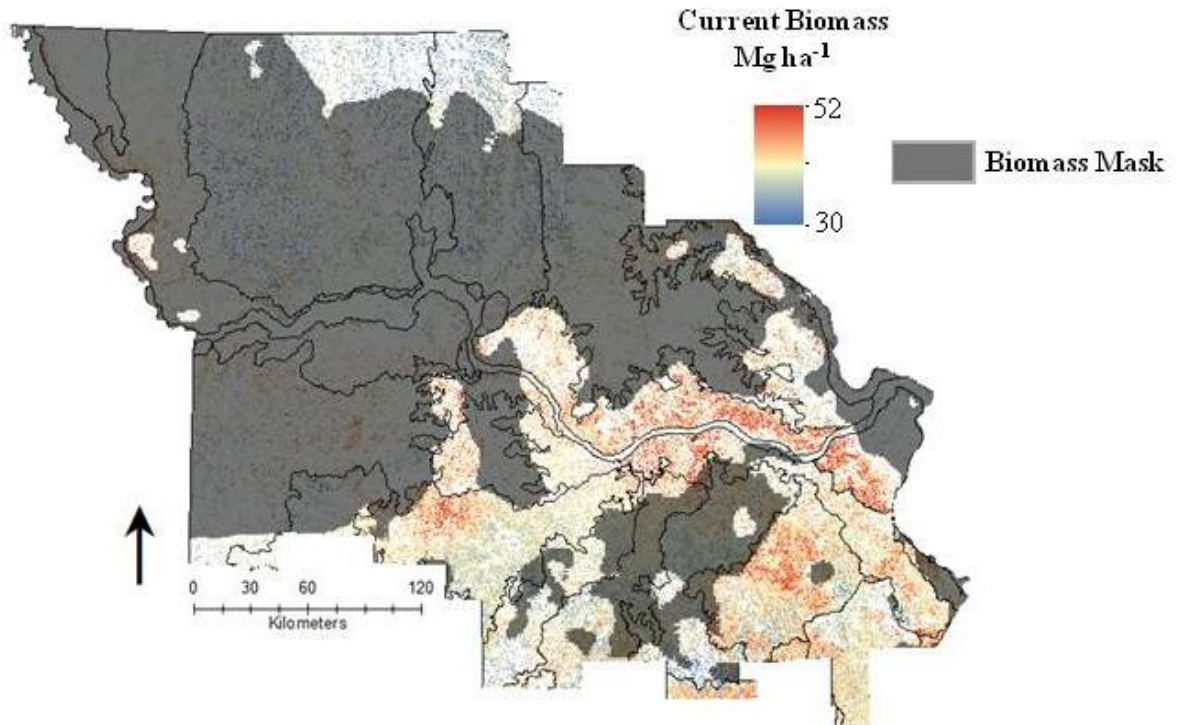
**Figure 4 - 2: Flow chart illustrating processes to obtain suitability index.**



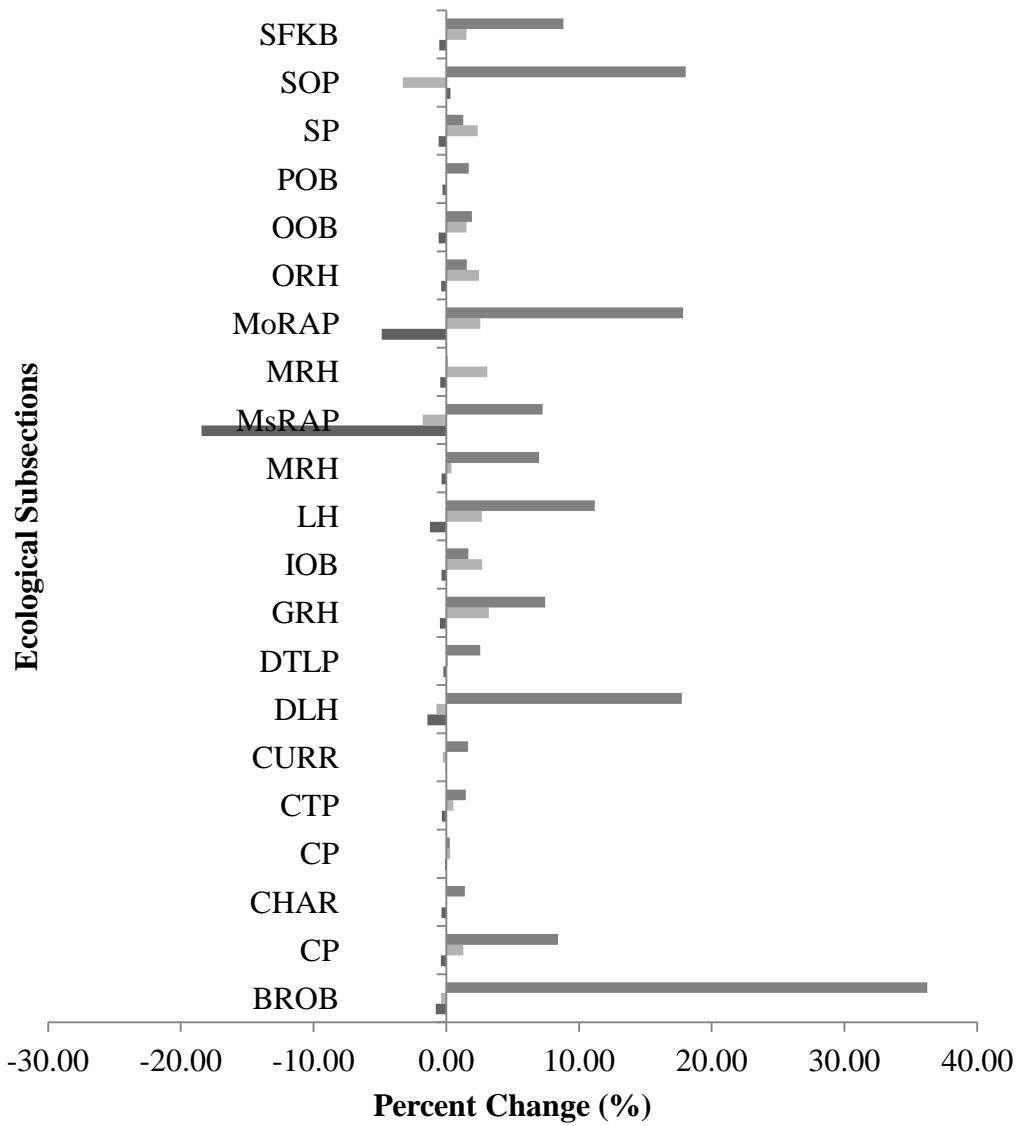
**Figure 4 - 3: (A) Percent changes between historic and current aboveground forest biomass (AFB) estimated using a Random Forest procedure at the pixel scale. (B) Average AFB summarized by ecological subsection.**



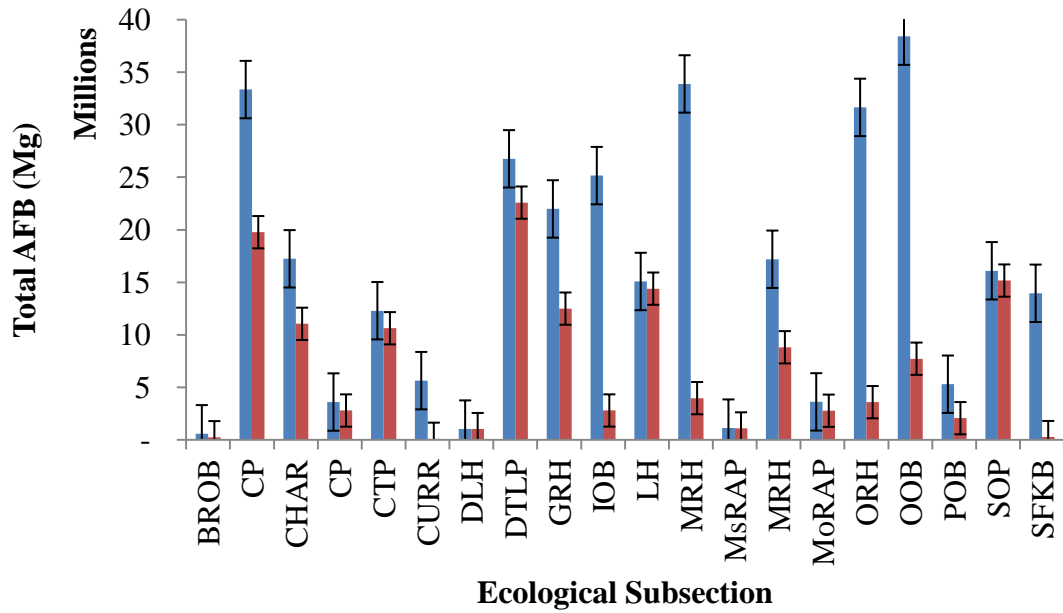
**Figure 4 - 4: Variable importance plot for estimates of Current and Historic aboveground forest biomass (AFB) generated from a Random Forest procedure.**



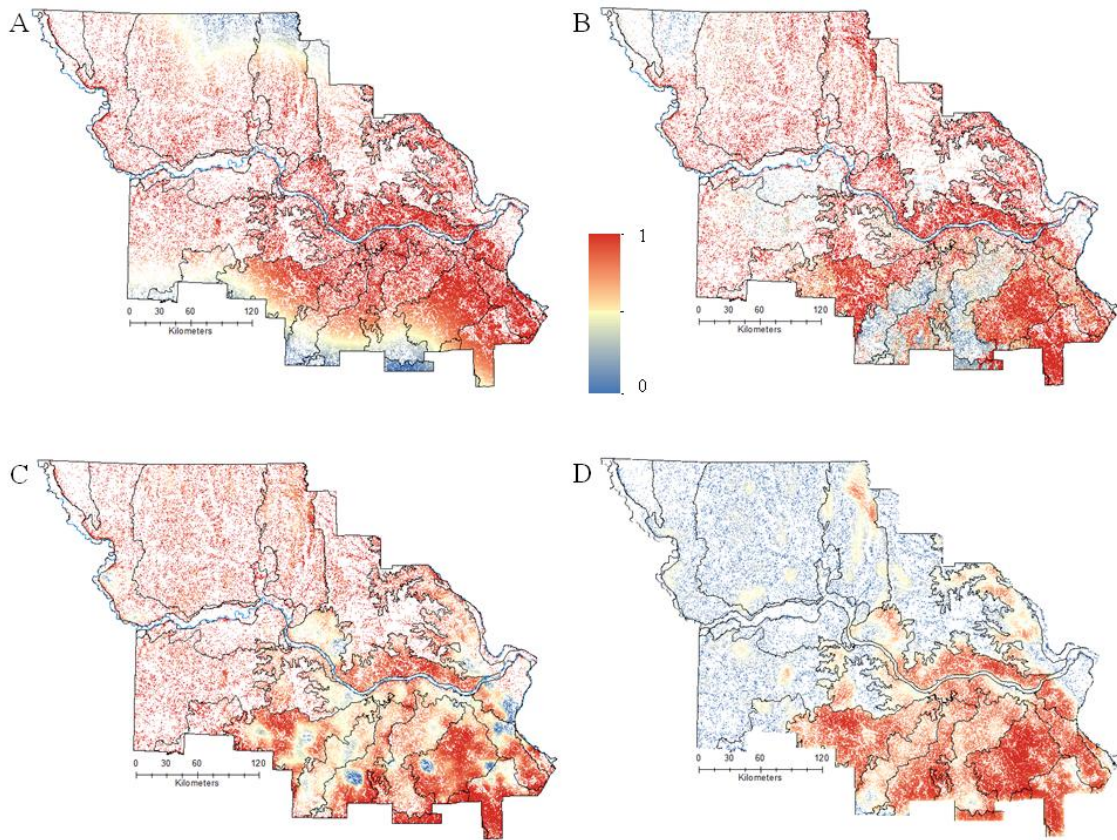
**Figure 4 - 5: The biomass mask is used to constrain total aboveground forest biomass (AFB) estimated by the RandomForest model and identify areas of available AFB. The mask was generated based on a kernel density estimate of forest transition values that were ranked from 0 – 1 (i.e. after fuzzy membership algorithm was applied to values of percent forest transition. Locations that were  $\geq 0.95$  based on the fuzzy membership analysis were masked to represent available areas. All other areas not covered by the mask were considered not available and left out of the suitability analysis.**



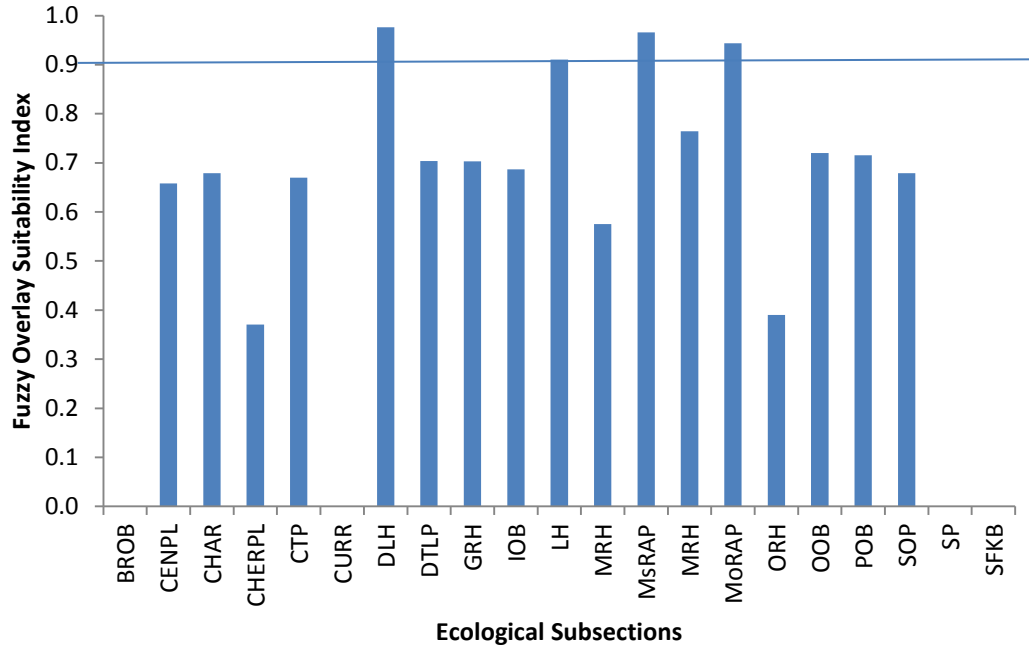
**Figure 4 - 6: Percent changes in Forest (top-bar), Agriculture (middle bar), and 'Other' (bottom bar) between 2001 and 2006.**



**Figure 4 - 7: Total current (Unconstrained – Blue) and total available (Constrained – Red) aboveground forest biomass summarized by each ecological subsection with standard errors.**

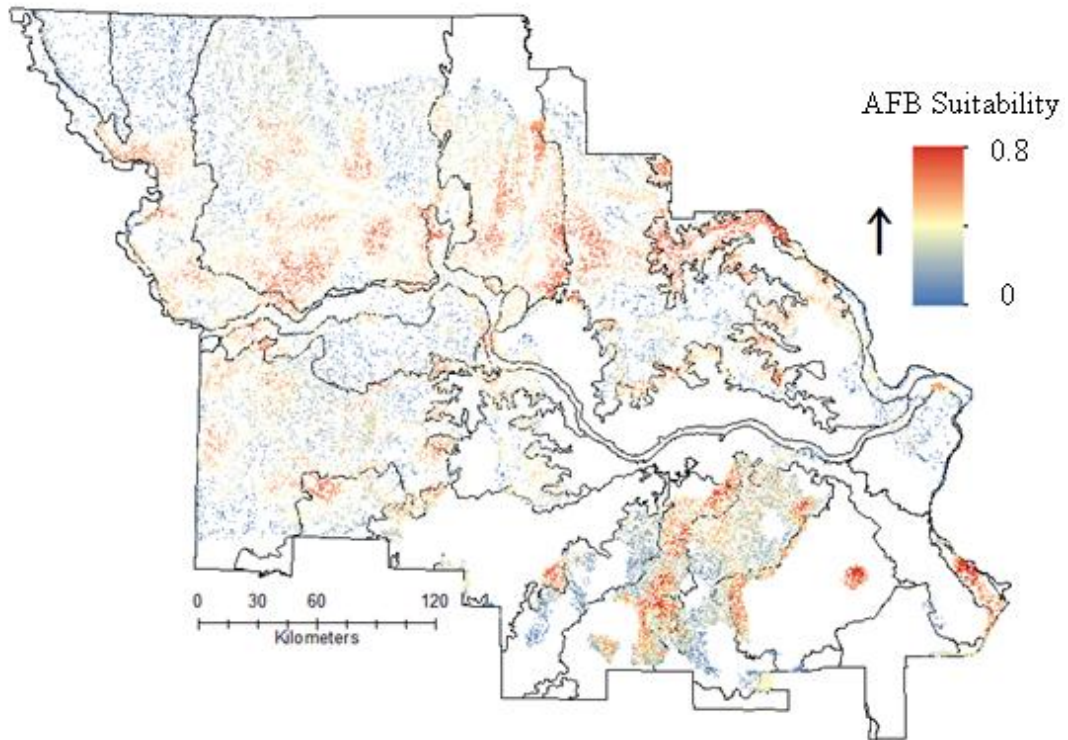


**Figure 4 - 8: Fuzzy membership for (A) Euclidean distance, (B) percent change, (C) forest transition, and (D) current AFB quantity covering all forested pixels ( $30\text{m}^2$ ).**

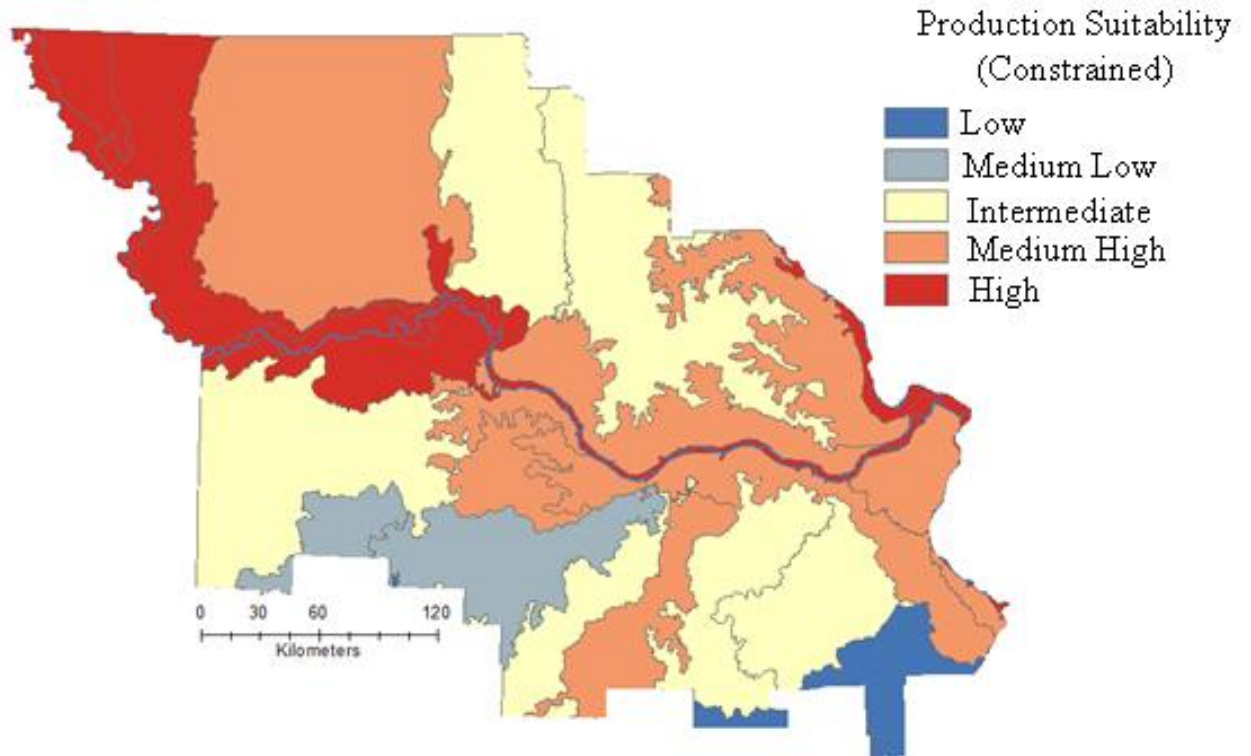


**Figure 4 - 9: Suitability index per ecological subsection where the horizontal bar indicates the value used (0.9) to categorized subsections with the greatest suitability (Loess Hills, Mississippi, and Missouri River Hills).**





**Figure 4 - 10: Available aboveground forest biomass suitability index based on Fuzzy Logic processes in ArcGIS.**



**Figure 4 - 11: Constrained (available) aboveground forest biomass suitability categories stratified by ecological subsection.**

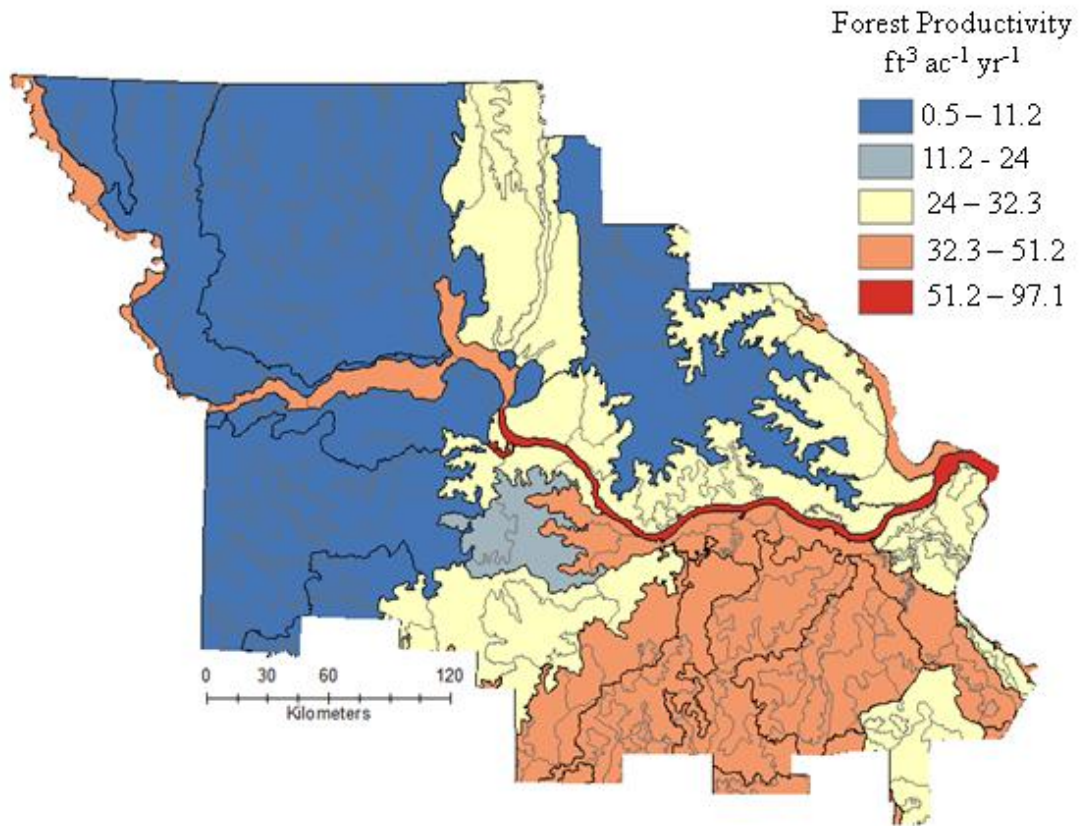


Figure 4-12: Forest productivity derived from SSURGO database. The forest productivity map was used to fit the suitability model to identify areas most suitability for increased AFB production.

**Table 4 - 1: Description of environmental variables used to estimate aboveground forest biomass in a Random Forest modeling procedure.**

<b>Environmental Group</b>	<b>Variable Code</b>	<b>Description</b>
<b>Climate</b>	BIO_01	Mean annual temperature (Celsius)
	BIO_15	Covariation annual precipitation (inches)
<b>Soil</b>	SAND	Sand content (%)
	SILT	Silt content (%)
	CLAY	Clay content (%)
	OM	Organic matter content (%)
	CF	Coarse fragments (%)
	PH	The sum of NH <sub>4</sub> OAc extractable bases (pH 7.0)
	KSAT	Hydraulic Conductivity
<b>Topography</b>	ELEV	Elevation (ft)
	ASP	Aspect (Degree)
	SLOPE	Slope (%)
	TPI	Topographic Position Index
	TWI	Topographic Wetness Index
	VRM	Vector Ruggedness Index
<b>Remote Sensing</b>	NDVI	Normalize Difference Vegetation Index
	NDWI	Normalized Difference Wetness Index
	B*	Brightness
	G*	Greenness
	W*	Wetness

*\*Derived from tasseled-cap transformation.*

**Table 4 - 2: Summary of average aboveground forest biomass estimates per ecological subsection between Potential and Available AFB values.**

SUBSECTION NAME	AVAILABLE		CURRENT		PERCENT CHANGE
	TOTAL (Mg)	AVERAGE (Mg ha <sup>-1</sup> )	TOTAL (Mg)	AVERAGE (Mg ha <sup>-1</sup> )	
Black River Ozark Border	2.5E+05	41.5	5.8E+05	96.6	-57.0
Central Plateau	2.0E+07	54.2	3.3E+07	91.4	-40.7
Chariton River Hills	1.1E+07	56.6	1.7E+07	88.2	-35.9
Cherokee Plains	2.8E+06	70.9	3.6E+06	91.4	-22.4
Claypan Till Plains	1.1E+07	78.8	1.2E+07	91.2	-13.6
Current River Hills	1.0E+05	1.7	5.6E+06	95.4	-98.2
Deep Loess Hills	1.0E+06	97.4	1.0E+06	97.4	0.0
Dissected Till and Loess Plains	2.3E+07	74.7	2.7E+07	88.5	-15.6
Gasconade River Hills	1.2E+07	52.7	2.2E+07	92.8	-43.2
Inner Ozark Border	2.8E+06	10.6	2.5E+07	95.5	-88.9
Loess Hills	1.4E+07	90.0	1.5E+07	94.2	-4.5
Meramec River Hills	4.0E+06	11.1	3.4E+07	95.0	-88.3
Mississippi River Alluvial Plain	1.1E+06	86.3	1.1E+06	89.1	-3.1
Mississippi River Hills	8.8E+06	47.9	1.7E+07	93.5	-48.7
Missouri River Alluvial Plain	2.8E+06	69.5	3.6E+06	90.6	-23.3
Osage River Hills	3.6E+06	10.6	3.2E+07	93.6	-88.7
Outer Ozark Border	7.7E+06	19.4	3.8E+07	96.4	-79.9
Prairie Ozark Border	2.1E+06	36.9	5.3E+06	94.7	-61.1
Scarped Osage Plains	1.5E+07	88.1	1.6E+07	93.5	-5.8
St. Francois Knobs and Basins	2.6E+05	1.7	1.4E+07	93.1	-98.1

**Table 4 - 3: Area and total aboveground forest biomass summarized by ecological subsection.**

<b>Subsection Name</b>	<b>Area (ha)</b>	<b>Current (Mg)</b>	<b>Available (Mg)</b>
Black River Ozark Border (BROB)	$7.59 \times 10^3$	$2.86 \times 10^6$	$1.27 \times 10^6$
Central Plateau (CP)	$7.69 \times 10^5$	$1.66 \times 10^8$	$1.02 \times 10^8$
Chariton River Hills (CRH)	$7.01 \times 10^5$	$8.56 \times 10^7$	$6.05 \times 10^7$
Cherokee Plains (CP)	$1.75 \times 10^5$	$1.79 \times 10^7$	$1.46 \times 10^7$
Claypan Till Plains (CTP)	$9.18 \times 10^5$	$6.10 \times 10^7$	$5.73 \times 10^7$
Current River Hills (CRH)	$6.73 \times 10^4$	$2.80 \times 10^7$	$5.20 \times 10^5$
Deep Loess Hills (DLH)	$1.57 \times 10^5$	$5.07 \times 10^6$	$5.09 \times 10^6$
Dissected Till and Loess Plains (DTLP)	$1.68 \times 10^6$	$1.33 \times 10^8$	$1.21 \times 10^8$
Gasconade River Hills (GRH)	$3.57 \times 10^5$	$1.09 \times 10^8$	$6.39 \times 10^7$
Inner Ozark Border (IOB)	$4.94 \times 10^5$	$1.25 \times 10^8$	$1.46 \times 10^7$
Loess Hills (LH)	$9.85 \times 10^5$	$7.49 \times 10^7$	$7.28 \times 10^7$
Meramec River Hills (MRH)	$4.56 \times 10^5$	$1.68 \times 10^8$	$2.02 \times 10^7$
Mississippi River Alluvial Plain (MsRAP)	$8.14 \times 10^4$	$5.55 \times 10^6$	$5.85 \times 10^6$
Mississippi River Hills (MsRH)	$5.00 \times 10^5$	$8.54 \times 10^7$	$4.65 \times 10^7$
Missouri River Alluvial Plain (MoRAP)	$4.32 \times 10^5$	$1.79 \times 10^7$	$1.46 \times 10^7$
Osage River Hills (ORH)	$5.29 \times 10^5$	$1.57 \times 10^8$	$1.88 \times 10^7$
Outer Ozark Border (OOB)	$8.71 \times 10^5$	$1.91 \times 10^8$	$4.14 \times 10^7$

Prairie Ozark Border (POB)	$2.27 \times 10^5$	$2.63 \times 10^7$	$1.06 \times 10^7$
Scarped Osage Plains (SOP)	$9.36 \times 10^5$	$7.99 \times 10^7$	$7.72 \times 10^7$
St. Francois Knobs and Basins (SFKB)	$1.89 \times 10^3$	$6.93 \times 10^7$	$1.30 \times 10^6$

## CHAPTER 5: CONCLUSION

*Shove nature around a little bit and find out if anything happens. And if it does, you might be on the edge of an important breakthrough, and then you sit down, devise experiments, and take notes.*

*- E.O. Wilson*

### 5.1 OVERVIEW

The objectives of this research were to demonstrate a Random Forest (RF) methodology that could be used to help guide future research for estimating aboveground forest biomass (AFB) and potentially help managers design effective plans by understanding how much is available and where it is located. Moreover, the work aims to identify spatially explicit areas that may have the best potential for increasing AFB. The outcome of this dissertation research was a coarse-scale, systems approach for estimating AFB across the Missouri River corridor. A hybrid modeling structure was created using a RF procedure with soil, topographic, climate, and remote sensing variables to estimate AFB. We applied the RF model to both current and historic (circa 1850) AFB values at a 30m resolution and then summarized all forested pixel values by ecological subsections. We then attempted to create a suitability map using our estimated biomass quantities, the percent changes in AFB between historic and current, and an additional two variables that simulated physical constraints on the landscape (forest transition and distance to rivers suitable for biomass transport). Based on these variables, a fuzzy logic, overlay analysis was conducted to produce a suitability map that illustrated spatial locations well suited for AFB production.



The RF procedure resulted in poor estimates of both current and historic AFB at the 30m resolution. A better statistical relationship (greater  $R^2$  value in the correlation between observed and predicted) was observed when all estimated AFB values were summarized by ecological subsection. This may suggest that this RF procedure may not be able to predict AFB at such a fine scale either due to the type of variables used (i.e. they could not explain much of the variation between observed AFB values) or that the number of variables used (maximum of 20) for estimating AFB was not appropriate. Another possibility is that the observed AFB data from FIA was too coarse to be able to estimate at a 30m resolution. Thus, we conclude that the RF model performance created inconclusive results at the pixel resolution (30m) and that AFB estimates summarized at the ecological subsection scale may be more representative and appropriate for using the observed data derived from the USFS FIA.

The RF procedure for estimating AFB and stratification by ecological subsection may be an effective procedure to pursue future AFB estimations using FIA data; however, our results suggest that the ecological subsection may be a more appropriate scale (compared to the pixel scale) to accomplish a broad-scale analysis of AFB dynamics. Additionally, the RF model output provided measures of variable importance that may be important to either narrow down prospective variables for future modeling attempts or to help understand what primary factors have the greatest influence on AFB. For this current modeling attempt, climate was determined to be the primary determinant of AFB, but only for a model that explained very little of the total variation in AFB.

Despite poor statistical estimation in model results, the value of this dissertation research lies in expanding our understanding of what it may take to effectively estimate a

dynamic resource, such as AFB. A small amount of variability was explained using the RF approach with a limited number of variables. Therefore, adding more and different variables to the RF procedure may have the potential to produce better estimates of AFB and be capable of lending more insight into the spatial availability of AFB. Additionally, altering or adding more variables to the RF models that better reflect site-specific landscape characteristic (e.g. tree species physiology, different time frames, or spatial scales) may result in better estimates. We conclude here that the RF procedure can be an effective methodology, but stress that this approach requires better information (observed and predictor variables) to train the model. Ultimately, this research can only be used as an example of a methodology that was not successful and suggest alternatives to improve this methodology.

A post-hoc test was performed to clarify if the RF forest model was running appropriately and gain a better understanding of the role that disturbances and stand age may be playing in current AFB estimates. Three cycles of FIA data (Cycles 5, 6, and 7) that span 15 years (N = 12498) was used to train a new RF model. The predictor variables were taken directly from the FIA dataset in contrast to our models which assigned landscape variables to FIA plot locations. The FIA variables were forest disturbance (DISTRB), site index (SICOND), stand age (STAGE), slope, and aspect. The new RF model also resulted in poor estimates of AFB ( $R^2 = 0.14$ ), although superior to our initial results (Figure 5-1). The top two variables in the model were stand age and site index with the disturbance parameter at the very bottom of the list (Figure 5-2). This post-hoc test suggests that the RF model is working appropriately due to an increased  $R^2$  value resulting from more scale-appropriate variables, and that the major limitations to the

methodology are the scale, quantity, or quality of input variables. We suggest that remote sensing indices, which are known to be good indicators of AFB, be used in future studies to improve the RF model for estimating current AFB.

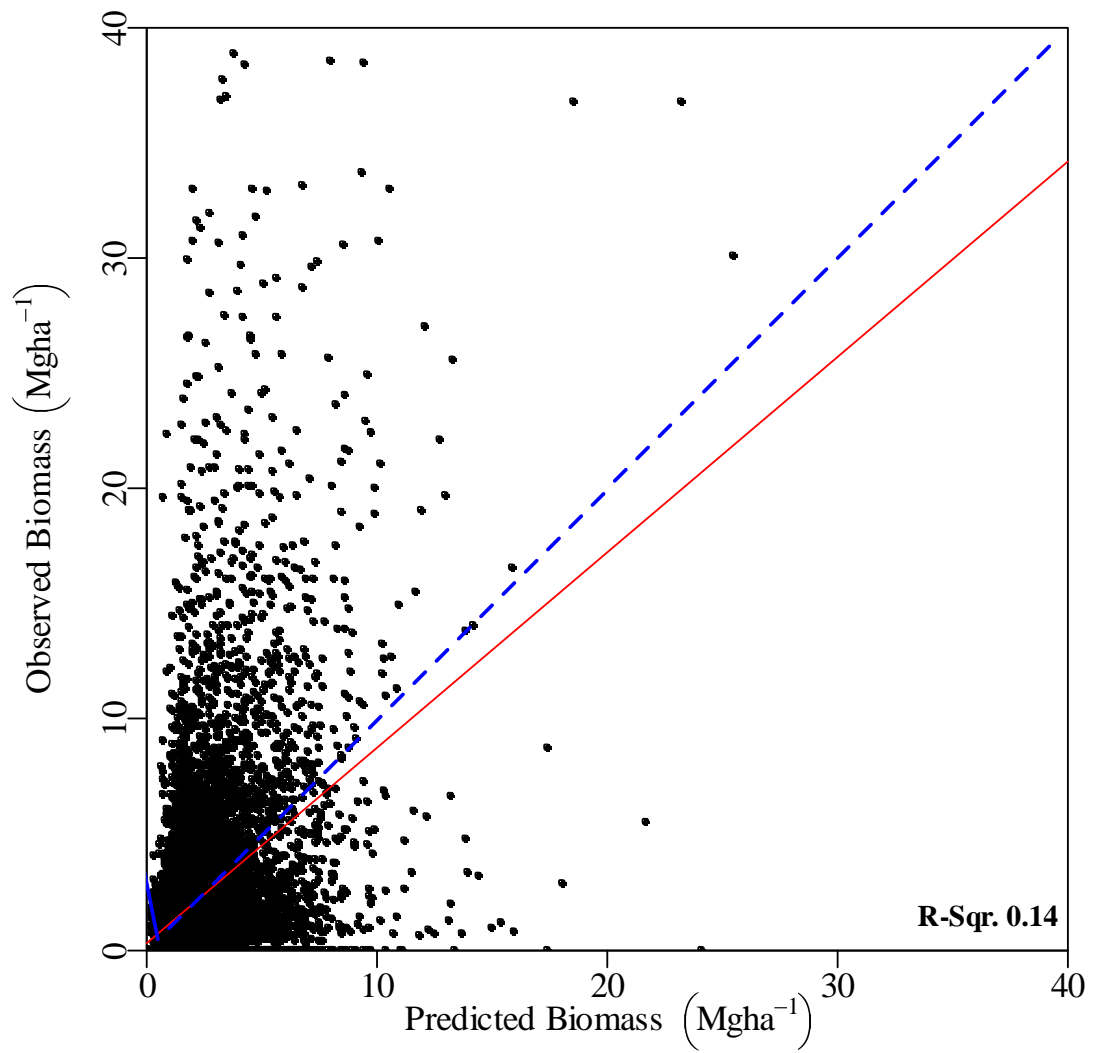
## **5.2 LIMITATIONS**

The restrictions on access to spatial locations of FIA plot data were a major constraint in this research. Without direct access to FIA plot location data, we were limited to a maximum of 20 predictor variables in this study and could not include additional variables such as other Bioclim climate variables (e.g. Max Temperature of Warmest Month, Temperature Annual Range, Precipitation of wettest Month, etc.). While we understand the need for protecting the spatial coordinates and ownership of individual FIA plot locations, those restrictions limit the types of analyses that are possible in a study such as this one.

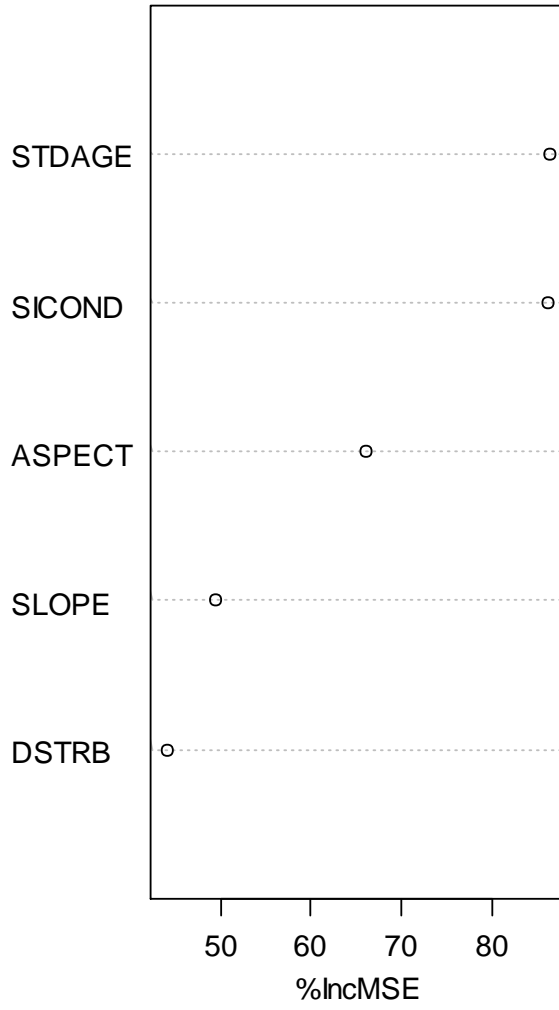
Another limitation was the scale at which the FIA information was collected. We attempted to use coarse-scaled information (e.g. one FIA plot per roughly 2500 ha (USDA FIA 2013)) to estimate AFB at a fine (30m) resolution, which proved to be unsuccessful. This methodology was necessary to test the utility of: 1) the RF model, and 2) FIA data. A better relationship (increased  $R^2$  value) was observed after summarizing all forested pixel values of estimated AFB by ecological subsection. This suggests that the variables represented broad-scale drivers of AFB, such as climate and soil characteristics. Further, interactions among local and broad-scale processes were not considered, but could be important in explaining patterns of AFB distributions.

## **5.3 FUTURE DIRECTIONS**

The models represent only a single scale and time period at which to identify the nature of AFB dynamics across the Missouri River corridor. It remains important for future research to unravel potential phenomena based on the interactions between AFB and ecosystem processes at various scales. Models are only extensions of our ability to comprehend scale-dependent details of the dynamic nature of AFB and there are certainly myriad phenomena occurring at scales that are so radically disparate that they need to be considered separately. For instance, intermediate-scale properties associated with spatial heterogeneity should be investigated to determine how pattern-process relationships interact from fine to broad scales such as competition between species or more human-oriented factors, such as management objectives between private and public lands. Moreover, fine-scale patterns often cannot be understood without knowledge of broader-scale processes and therefore require a systems approach to understanding the dynamics of AFB (Lewis et al. 2014). Landscape ecology deals fundamentally with how, when, and why patterns of environmental factors influence the distribution of organisms and ecological process, and vice versa. Subsequently, future directions for this research may entail modeling how anthropogenic alterations may change AFB distribution, where the changes are most likely to occur, the socioeconomic or environmental conditions that dictate why changes may occur such as process-facility placement or renewed interests in marginal lands for increasing AFB. Further, focus should be placed on which particular species may be most suitable, as opposed to an overall, broad-scale analysis.



**Figure 5 - 1: Observed versus predicted AFB from the post-hoc RF model. Red line indicates model fit and blue dashed line illustrates the 1:1 line.**



**Figure 5 - 2: Variable importance chart from the RF model.**

## REFERENCES

- Aguilar, F. X., N. Song, and S. Shifley. 2011. Review of consumption trends and public policies promoting woody biomass as an energy feedstock in the U.S. . *Biomass and Bioenergy* 35:3708-3718.
- Anaya, J. A., E. Chuvieco, and A. Palacios-Orueta. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management* 257(4):1237-1246.
- Baccini, A., M. A. Friedl, C. E. Woodcock, and R. Warbington. 2004. Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters* 31.
- Bardhan, S., and S. Jose. 2012. The potential for floodplains to sustain biomass feedstock production systems. *Biofuels* 3(5):575 - 588.
- Becker, D. R., D. Abbas, K. E. Halvorsen, P. J. Jakes, S. M. McCaffrey, and C. Moseley. 2009. Conventional Wisdoms of Woody Biomass Utilization: University of Oregon.
- Berndes, G., M. Hoogwijk, and R. Van den Broek. 2003. The contribution of biomass in the future global energy supply: a review of 17 studies. *Biomass and Bioenergy* 25:1 - 28.
- Blackard, J. A., M. V. Finco, E. H. Helmer, G. R. Holden, M. L. Hoppus, D. M. Jacobs, A. J. Lister, G. G. Moisen, M. D. Nelson, R. Riemann, B. Ruefenacht, D. Salajanu, D. L. Weyermann, K. C. Winterberger, T. J. Brandeis, R. L. Czaplowski, R. E. McRoberts, P. L. Patterson, and R. P. Tymcio. 2008. Mapping

- U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment* 112:1658-1677.
- Boisvenue, C., and S. W. Running. 2006. Impacts of climate change on natural forest productivity – evidence since the middle of the 20th century. *Global Change Biology*(12):862 - 882.
- Boroushaki, S., and J. Malczewski. 2010. Using the fuzzy majority approach for GIS-based multicriteria group decision-making. *Computers & Geosciences* 36:302-312.
- Boulanger, Y., S. Gauthier, D. R. Gray, H. L. Goff, P. Lefort, and J. Morissette. 2013. Fire regime zonation under current and future climate over eastern Canada. *Ecological Society of America*.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5 - 32.
- Brown, S., P. Schroeder, and R. Birdsey. 1997. Aboveground Biomass Distribution of US Eastern Hardwood Forests and the Use of Large Trees as Indicator of Forest Development. *Forest Ecology and Management* 96:37 - 47.
- Brzostek, E. R., D. Dragoni, H. P. Schmid, A. F. Rahman, D. Sims, C. A. Wayson, D. J. Johnson, and R. P. Phillips. 2014. Chronic water stress reduces tree growth and the carbon sink of deciduous hardwood forests. *Global Change Biology*.
- Cannell, M. G. R. 2003. Carbon Sequestration and Biomass Energy Offset: Theoretical, Potential and Achievable Capacities Globally, in Europe and the UK. *Biomass and Bioenergy* 24:97 - 116.
- Cherubini, F., and A. H. Strømman. 2011. Life cycle assessment of bioenergy systems: State of the art and future challenges. *Bioresource Technology*(102):437 - 451.



- Cutler, R. D., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random Forests for Classification in Ecology. *Ecology* 88(11):2783 - 2792.
- Delcourt, H. R., and P. A. Delcourt. 1996. Presettlement landscape heterogeneity: Evaluating grain of resolution using General Land Office Survey data. *Landscape Ecology* 11(6):363 - 381.
- DOE/EIA. 2014. May 2014: Monthly Energy Review, edited by U. S. D. o. Energy. Washington, DC: U.S. Energy Information Administration.
- EISA. 2007. Energy Independence and Security Act of 2007: A Summary of Major Provisions, edited by F. Sissine: Congressional Research Service.
- ArcGIS ArcMap™ 10.2. Environmental Systems Research Institute, Redlands, CA
- Falkowski, M. J., J. S. Evans, S. Martinuzzi, P. E. Gessler, and A. T. Hudak. 2009. Characterizing forest succession with lidar data: An evaluation for the Inland Northwest, USA. *Remote Sensing of Environment* 113:946 - 956.
- Freppaz, D., R. Minciardi, M. Robba, M. Rovatti, R. Sacile, and A. Taramasso. 2004. Optimizing forest biomass exploitation for energy supply at a regional level. *Biomass and Bioenergy* 26:15-25.
- Galindo-Leal, C., and F. L. Bunnell. 1995. Ecosystem management: Implication and opportunities of a new paradigm. *The Forestry Chronicle* 71(5):601 - 606.
- Gallaun, H., G. Zanchi, G.-J. Nabuurs, G. Hengeveld, M. Schardt, and P. J. Verkerk. 2010. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management* 260:252 - 261.

- Gao, B.-c. 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 53(3):257 - 266.
- Gasparri, N. I., M. G. Parmuchi, J. Bono, and H. Karszenbaum. 2010. Assessing multi-temporal Landsat 7 ETM<sub>p</sub> images for estimating above-ground biomass in subtropical dry forests of Argentina. *Journal of Arid Environments* 74:1262 - 1270.
- Gleason, C. J., and J. Im. 2011. A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications. *GIScience and Remote Sensing* 48(2):141 - 170.
- Goerndt, M. E., F. X. Aguilar, P. Miles, S. Shifley, N. Song, and H. Stelzer. 2012. Regional Assessment of Woody Biomass Physical Availability as an Energy Feedstock for Combined Combustion in the US Northern Region. *Journal of Forestry* 110(3):138-148.
- Gopalakrishnan, G., M. C. Negri, and S. W. Snyder. 2011. A Novel Framework to Classify Marginal Land for Sustainable Biomass Feedstock Production. *Journal of Environmental Quality* 40:1593 - 1600.
- Graham, R. L., and M. E. Walsh. 1999. A National Assessment of Promising Areas for Switchgrass, Hybrid Poplar, or Willow Energy Crop Production, edited by U. S. D. o. Energy: Oak Ridge National Laboratory.
- Grumbine, R. E. 1994. What is Ecosystem Management? *Conservation Biology* 8(1):27 - 38.

- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993 - 1009.
- Hacatoglu, K., M. James, and D. B. Layzell. 2011. Feasibility study of a Great Lakes bioenergy system. *Bioresource Technology* 102(2):1087-1094.
- Hall, J. S., B. E. Love, E. J. Garen, J. L. Slusser, K. Saltonstall, S. Mathias, M. van Breugel, D. Ibarra, E. W. Bork, D. Spaner, M. H. Wishnie, and M. S. Ashton. 2011. Tree plantations on farms: Evaluating growth and potential for success. *Forest Ecology and Management* 261:1675-1683.
- Hanberry, B. B., S. Fraver, H. S. He, J. Yang, D. C. Dey, and B. J. Palik. 2011. Spatial pattern corrections and sample sizes for forest density estimates of historical tree surveys. *Landscape Ecology* 26:59 - 68.
- Hanberry, B. B., J. M. Kabrick, H. S. He, and B. J. Palik. 2012. Historical trajectories and restoration strategies for the Mississippi River Alluvial Valley. *Forest Ecology and Management* 280:103 - 111.
- Hanberry, B. B., B. J. Palik, and H. S. He. 2012. Comparison of historical and current forest surveys for detection of homogenization and mesophication of Minnesota forests. *Landscape Ecology* 27(10):1495 - 1512.
- Havlík, P., U. A. Schneider, E. Schmid, H. Bottcher, S. Fritz, R. Skalsky', K. Aoki, S. DeCara, G. Kindermann, F. Kraxner, S. Leduc, I. McCallum, A. Mosnier, T. Sauer, and M. Obersteiner. 2011. Global land-use implications of first and second generation biofuel targets. *Energy Policy* 39:5690 - 5702.

- He, H. S., D. C. Dey, X. Fan, M. B. Hooten, J. M. Kabrick, C. K. Wikle, and Z. Fan. 2007. Mapping pre-European settlement vegetation at fine resolutions using a hierarchical Bayesian model and GIS. *Plant Ecology* 191:85 - 94.
- He, H. S., D. J. Mladenoff, V. C. Radeloff, and T. Crow. 1998. Integration of GIS Data and Classified Satellite Imagery for Regional Forest Assessment. *Ecological Applications* 8(4):1072 - 1083.
- Healey, S. P., W. B. Cohen, Y. Zhiqiang, and O. N. Krankina. 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment* 97(3):301 - 310.
- Heinimann, H. R. 2010. A concept in adaptive ecosystem management—An engineering perspective. *Forest Ecology and Management*(259):848 - 856.
- Hieisig, H., C. Kühnle, L. Paredes, and J. Wiesel. 2005. GIS based Forest Biomass Exploitation for Energy Supply in the Northern Black Forest Region. *Photogrammetrie, Fernerkundung, Geoinformation*:273-280.
- Hijmans, R. J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* (25):1965 - 1978.
- Hoogwijk, M., A. Faaij, R. van den Broek, G. Berndes, D. Gielen, and W. Turkenburg. 2003. Exploration of the ranges on the global potential of biomass for energy. *Biomass and Bioenergy* 25:119-133.
- Houghton, R. A., F. Hall, and S. J. Goetz. 2009. Importance of biomass in the global carbon cycle. *Journal of Geophysical Research* 114(G00E03).

- IPCC. 2007. Climate Change 2007: Impact, Adaptation and Vulnerability. In *Contribution of Working Group II to the Fourth Assessment Report of the IPCC*, edited by O. F. C. M.L. Parry, J.P. Palutikof, P.J. van der Linden and C.E. Hanson. Cambridge, UK: Cambridge University Press.
- Jenkins, J. C., R. A. Birdsey, and Y. Pan. 2001. Biomass and NPP Estimation for the Mid-Atlantic Region (USA) Using Plot-Level Forest Inventory Data. *Ecological Applications* 11(4):1174 - 1193.
- Jenkins, J. C., D. C. Chojnacky, L. S. Heath, and R. A. Birdsey. 2003. National-Scale Biomass Estimators for United States Tree Species. *Forest Science* 49(1).
- Johnson, F. L., and D. T. Bell. 1976. Plant Biomass and Net Primary Production along a Flood-Frequency Gradient in the Streamside Forest. *Castanea* 41(2):156 - 165.
- Jose, S., and A. R. Gillespie. 1997. Leaf area-productivity relationships among mixed-species hardwood forest communities of the Central Hardwood Region. *Forest Science* 43:56-64.
- Jose, S., and A. R. Gillespie. 1998. Compensatory mechanisms of Central Hardwood forest communities in a changing environment. *Plant Ecology* 135:125-134.
- Keith, H., B. G. Mackey, and D. B. Lindenmayer. 2009. Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. *PNAS* 106(28):11635-11640.
- Kellndorfer, J. M., W. S. Walker, E. LaPoint, K. Kirsch, J. Bishop, and G. Fiske. 2010. Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets. *Journal of Geophysical Research* 115(G2).

- Kucera, C. L., and R. E. McDermott. 1955. sugar Maple-Basswood Studies in the Forest-Prairie Transition of Central Missouri. *American Midland Naturalist* 54(2):495 - 503.
- Lambin, E. F., and P. Meyfroidt. 2010. Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences of the United States of America* 108(9):1465-3472.
- Lars, K., C. Fürst, S. Frank, and F. Makeschin. 2012. A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning. *Ecological Indicators* 21:54 -66.
- Levy, P. E., M. G. R. Cannell, and A. D. Friend. 2004. Modelling the impact of future changes in climate, CO<sub>2</sub> concentration and land use on natural ecosystems and the terrestrial carbon sink. *Global Environmental Change* 14:21 - 30.
- Lewis, S. M., G. Fitts, M. Kelly, and L. Dale. 2014. A fuzzy logic-based spatial suitability model for drought-tolerant switchgrass in the United States. *Computers and Electronics in Agriculture* 103:39-47.
- Liang, J., and M. Zhou. 2014. Large-Scale Geospatial Mapping of Forest Carbon Dynamics. *Journal of Sustainable Forestry*.
- Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. *R News* 2:18 - 22.
- Liu, J., T. Dietz, S. R. Carpenter, M. Alberti, C. Folke, E. Moran, A. N. Pell, P. Deadman, T. Kratz, J. Lubchenco, E. Ostrom, Z. Ouyang, W. Provencher, C. Redman, S. H. Schneider, and W. W. Taylor. 2007. Complexity of Coupled Human and Natural Systems. *Science* 317:1513 - 1516.

- Liu, J., V. Hull, M. Batistella, R. DeFries, T. Dietz, F. Fu, T. W. Hertel, C. R. Izaurralde, E. F. Lambin, S. Li, L. A. Martinelli, W. J. McConnell, E. F. Moran, R. Naylor, Z. Ouyang, K. R. Polenske, A. Reenberg, G. d. M. Rocha, C. Simmons, P. H. Verburg, P. M. Vitousek, F. Zhang, and C. Zhu. 2013. Framing Sustainability in a Telecoupled World. *Ecology and Society* 18(2):26.
- Lovett, A. A., G. M. Sünnerberg, G. M. Richter, A. G. Dailey, A. B. Riche, and A. Karp. 2009. Land Use Implications of Increased Biomass Production Identified by GIS-Based Suitability and Yield Mapping for Miscanthus in England. *Bioenergy Research* 2:17 - 28.
- Lu, D. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27(7).
- Lu, D., Q. Chen, G. Wang, E. Moran, M. Batistella, M. Zhang, G. V. Laurin, and D. Saah. 2012. Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. *International Journal of Forestry Research* 2012(436537):16.
- Mani, S., and N. Parthasarathy. 2007. Above-ground biomass estimation in ten tropical dry evergreen forest sites of peninsular India. *Biomass and Bioenergy* 31:248 - 290.
- MDC. 2009. Missouri Woody Biomass Harvesting Best Management Practices. , edited by M. D. O. C. (MDC).
- Miehle, P., S. J. Livesley, C. Li, P. M. Feikema, M. A. Adams, and S. K. Arndt. 2006. Quantifying uncertainty from large-scale model predictions of forest carbon dynamics. *Global Change Biology* 12:1421 - 1434.

- Moser, K. W., R. Bush, J. D. Shaw, M. H. Hansen, and M. D. Nelson. 2010. The Role of Strategic Forest Inventories in Aiding Land Management Decision Making: Examples From the U.S. Forest Inventory and Analysis Program. Paper read at Integrated management of carbon sequestration and biomass utilization opportunities in a changing climate: Proceedings of the 2009 National Silviculture Workshop, at Fort Collins, CO.
- Moser, K. W., M. H. Hansen, M. A. Hatfield, and T. A. Nigh. 2006. Evaluating Sustainability: A Method for Assessing Vegetation Change in Southern Missouri, U.S.A.: 1820 - 2003. *Landscape Ecology*.
- Nigh, T. A., and W. A. Schroeder. 2002. Atlas of Missouri Ecoregions, edited by M. D. o. Conservation. Missouri: The Conservation Commission.
- NRCS. 2011. Soil Data Viewer 6.0, edited by U. S. D. o. Agriculture.
- Nyeko, M. 2012. GIS and Multi-Criteria Decision Analysis for Land Use Resource Planning. *Journal of Geographic Information Systems* 4:341-348.
- O'Laughlin, J. 2009. The Forest—Bioenergy—Carbon Connection. Paper read at National Silviculture Workshop, at Boise, ID.
- Ocalir, E. V., O. Y. Ercoskun, and R. Tur. 2010. An integrated model of GIS and fuzzy logic (FMOTS) for location decisions of taxicab stands. *Expert Systems with Applications* 37:4892-4901.
- Pallardy, S. G., T. A. Nigh, and H. E. Garrett. 1988. Changes in Forest Composition in Central Missouri: 1968-1982. *American Midland Naturalist* 120(2):380 - 390.



- Pan, Y., R. Birdsey, O. L. Phillips, and R. B. Jackson. 2013. The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics* 44:593-622.
- Perpiña, C., J. C. Martínez-Llario, and Á. Pérez-Navarro. 2013. Multicriteria assessment in GIS environments for siting biomass plants. *Land Use Policy* 31:326-335.
- Potter, C., P. Gross, S. Klooster, M. Fladeland, and V. Genovese. 2008. Storage of carbon in U.S. forests predicted from satellite data, ecosystem modeling, and inventory summaries. *Climatic Change* 90:269 - 282.
- Powell, S. L., W. B. Cohen, S. P. Healey, R. E. Kennedy, G. G. Moisen, K. B. Pierce, and J. L. Ohmann. 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment* 114:1053 - 1068.
- Pradhan, B. 2011. Use of GIS-based fuzzy logic relations and its cross application to produce landslide susceptibility maps in three test areas in Malaysia. *Environmental Earth Sciences* 63:329-349.
- Prasad, A. M., L. R. Iverson, and A. Liaw. 2004. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9:181 - 199.
- Puhr, C. B., and D. N. M. Donoghue. 2000. Remote sensing of upland conifer plantations using Landsat TM data: a case study from Galloway, south-west Scotland. *International Journal of Remote Sensing* 21(4):633 - 646.
- R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

- Rhemtulla, J. M., D. J. Mladenoff, and M. K. Clayton. 2008. Historical forest baselines reveal potential for continued carbon sequestration. *PNAS* 106(15):6082 - 6087.
- Righelato, R., and D. Spracklen. 2007. Carbon Mitigation by Biofuels or by Saving and Restoring Forests? *Science* 317(5840):902.
- Shabani, N., and T. Sowlati. 2013. A mixed integer non-linear programming model for tactical value chain optimization of a wood biomass power plant. *Applied Energy*(104):353 - 361.
- Shifley, S. R. 2006. Sustainable Forestry in the Balance. *Journal of Forestry*:187 - 195.
- Smeets, E. M., and A. P. C. Faaij. 2007. Bioenergy potentials from forestry in 2050: An assessment of the drivers that determine the potentials. *Climatic Change*(81):353 - 390.
- Smith, W. B., P. D. Miles, C. H. Perry, and S. A. Pugh. 2007. Forest resources of the United States, edited by F. S. U.S. Department of Agriculture. Washington, DC.
- Smithwick, E. A. H., M. E. Harmon, S. M. Remillard, S. A. Acker, and J. F. Franklin. 2002. Potential Upper Bounds of Carbon Stores in Forests of the Pacific Northwest. *Ecological Applications* 12:1303 - 1317.
- Soratana, K., V. Khanna, and A. E. Landis. 2013. Re-envisioning the renewable fuel standard to minimize unintended consequences: A comparison of microalgal diesel with other biodiesels. *Applied Energy* 112:194 - 204.
- Spittlehouse, D. L., and R. B. Steward. 2003. Adaptation to climate change in forest management. *Journal of Ecosystems and Management* 4(1).

- Stanturf, J. A., C. J. Schweitzer, and E. S. Gardiner. 1998. Afforestation of Marginal Agricultural Land in the Lower Mississippi River Alluvial Valley, U.S.A. *Silva Fennica* 32(3):281 - 297.
- Stasko, T. H., R. J. Conrado, A. Wankerl, R. Labatut, R. Tasseff, J. T. Mannion, H. O. Gao, S. D. Sanborn, and G. Knott. 2011. Mapping Woody-biomass supply costs using forest inventroy and competing industry data. *Biomass and Bioenergy* 35(1):263 - 271.
- Sunde, M. G., H. S. He, B. Zhou, J. A. Hubbart, and A. Spicci. 2014. Imperviousness Change Analysis Tool (I-CAT) for simulating pixel-level urban growth. *Landscape and Urban Planning* 124:104 - 108.
- Thiffault, E., J. Barrette, D. Paré, B. D. Titus, and K. Keys. 2014. Developing and validating indicators of site suitability for forest harvesting residue removal. *Ecological Indicators* 43:1-18.
- U.S. Department of Energy. 2011. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry, edited by U. S. D. o. Energy. Oak Ridge, TN: Oak Ridge National Laboratory.
- USDA FIA. 2013. The Forest Inventory and Analysis Database: Database Description and Users Manual Version 5.1.6 for Phase 2 edited by F. S. U.S. Department of Agriculture, Forest Inventory and Analysis. Rosslyn Plaza, 1620 North Kent Street, Arlington, VA 22209.
- USDOE. 2011. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry, edited by R. D. P. a. B. J. Stokes. Oak Ridge, TN.: Oak Ridge National Laboratory.

- . 2011. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry, edited by U. S. D. o. Energy. Oak Ridge: Oak Ridge National Laboratory.
- Vizzari, M. 2011. Spatial modelling of potential landscape quality. *Applied Geography* 31:108-118.
- Voivontas, D., D. Assimacopoulos, and E. G. Koukios. 2001. Assessment of biomass potential for power production: a GIS based method. *Biomass and Bioenergy* 20:101-112.
- White, C. A. 1983. A History of the Rectangular Survey System, edited by B. o. L. Management. Washington, DC: Government Printing Office.
- Woudenberg, S. W., B. L. Conkling, B. M. O'Connell, E. B. Lapoint, J. A. Turner, and K. L. Waddell. 2010. The Forest Inventory and Analysis Database: Database description and users manual version 4.0 for Phase 2., edited by F. S. Department of Agriculture. Fort Collins: Rocky Mountain Research Station.
- Wulder, M. A., J. C. White, R. A. Fournier, J. E. Luther, and S. Magnussen. 2008. Spatially Explicit Large Area Biomass Estimation: Three Approaches Using Forest Inventory and Remotely Sensed Imagery in a GIS. *Sensors* 8:529 - 560.
- Yeo, I.-Y., and C. Huang. 2013. Revisiting the forest transition theory with historical records and geospatial data: A case study from Mississippi (USA). *Land Use Policy* 32:1 - 13.
- Zheng, D., L. S. Heath, and M. J. Ducey. 2008. Spatial distribution of forest aboveground biomass estimated from remote sensing and forest inventory data in New England, USA. *Journal of Applied Remote Sensing* 2.

- Zheng, D., L. S. Heath, M. J. Ducey, and B. Butler. 2010. Relationships Between Major Ownerships, Forest Aboveground Biomass Distributions, and Landscape Dynamics in the New England Region of USA. *Environmental Management* 45:377 - 386.
- Zheng, D., J. Rademacher, J. Chen, T. Crow, M. Bresee, J. Le Moine, and S.-R. Ryu. 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. *Remote Sensing of Environment* 93:402 - 411.
- Zhuang, D., D. Jiang, L. Liu, and Y. Huang. 2011. Assessment of bioenergy potential on marginal land in China. *Renewable and Sustainable Energy Reviews* 15:1050 - 1056.
- Zimmermann, N. E., T. C. Edwards, G. G. Moisen, T. S. Frescino, and J. A. Blackard. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44(5):1057 - 1067.
- Zolkos, S. G., S. J. Goetz, and R. Dubayah. 2013. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sensing of Environment* 128:289 - 298.

## VITA

I was born on 3 December 1980 in St. Clair Shores, Michigan to Clemens and Kathleen Bobryk. After completing my schoolwork at Lake Shore High School in 1999, I attended Wayne State University in Detroit Michigan and graduated with a Bachelors of Science in Environmental Science in 2007. During the next 3 years, I worked for an environmental consulting firm (ECT, Inc.) in Detroit, Michigan. I returned to academia in 2008 to attain a Master of Science degree in Biology from Eastern Illinois University. I continued my academic career by accepting a Ph.D. position in 2010 in the GIS/Spatial Analysis Laboratory at the University of Missouri. My wife, Shannon Marie, and I currently reside in Columbia, MO., where I am a postdoctoral research fellow in the Plant Sciences Department at the University of Missouri.