COMPARING THE USE OF SYNTHETIC APERTURE RADAR

AND LANDSAT THEMATIC MAPPER TO PREDICT

SLASH PINE (*PINUS ELLIOTTII* Englm.) PLANTATION ATTRIBUTES

by

JAMES MATTHEW LANDRETH

(Under the Direction of Richard F. Daniels)

ABSTRACT

Aerial photographs and, more recently, Landsat Thematic Mapper (LTM) have long been a source for spatial information in forest management. Recently, Synthetic Aperture Radar (SAR) has been introduced as another source upon which to base forest management decisions. This study explores use of LTM and SAR data to predict Slash Pine basal area and yield. Forest attribute estimation models were developed using two model-building techniques – stepwise multiple-linear regression and a more biologically motivated approach.

SAR data were found to explain more variability in both basal area and yield than LTM. Combining SAR and LTM significantly improved the amount of variation explained in basal area and yield in comparison to models developed from either SAR or LTM independently. The biologically motivated models explained more variation than a standard stepwise regression procedure due to model forms that were consistent with biological behavior. These more biologically motivated models also exhibit better extrapolative behavior.

INDEX WORDS: Forest inventory, SAR, LTM, Yield equations

COMPARING THE USE OF SYNTHETIC APERTURE RADAR AND LANDSAT THEMATIC MAPPER TO PREDEICT SLASH PINE (*PINUS ELLIOTTII*) PLANTATION ATTRIBUTES

by

JAMES MATTHEW LANDRETH

B.S., Florida State University, 1998

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2002

COMPARING THE USE OF SYNTHETIC APERTURE RADAR AND LANDSAT THEMATIC MAPPER TO PREDEICT SLASH PINE (*PINUS ELLIOTTII*) PLANTATION ATTRIBUTES

by

JAMES MATTHEW LANDRETH

Approved:

Major Professor: Richard F. Daniels

Committee: Michael L. Clutter
Bruce E. Borders

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2002

DEDICATION

To my family and friends that have supported me through this endeavor, for without which, this would not have been possible.

ACKNOWLEDGMENTS

I am grateful to Dr. Mike Clutter for his unwavering support for me and for the completion of this project. Without his continual advice, direction and leadership, this project would likely not have been a success. I wish to thank Dr. Dick Daniels and Dr. Bruce Borders similarly for their patience and dedication to the completion of this study, and for their open door policy and willingness to be there for those "quick questions". I am indebted to Mr. Bob Izlar, for without meeting him, it is likely that I would have never made it to the University of Georgia and had the honor of meeting a man who is a true mentor. Lastly, thanks and appreciation is required to Tripp Lowe for his invaluable role in the development of this project, in acting at times as a 'sound-mind', 'kindred spirit', and 'devil's advocate'.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I.

INTRODUCTION

DEVELOPMENT OF REMOTE SENSING FOR FORESTRY APPLICATIONS

The forests of Georgia are a financially and socially important resource for the State. Forestry and forest products contributes approximately $20 billion a year to Georgia's economy, providing nearly 177,000 jobs directly dependent on commercial forestry and within related commercial forest product manufacturing (Forest Facts, 2001). Currently, approximately sixty-six percent of the State's 36.8 million acres is forested. Nearly seventy-two percent is owned by non-industrial private landowners and seven percent falls under public ownership, leaving approximately twenty-one percent to be owned and managed by forest products companies (Forest Facts, 2001). Regardless of the owner's management objectives, forest stewardship is dependent upon having reliable and timely forest inventory information. Forest managers are under increasing demands to implement forest management plans that adequately supply forest products, but also satisfy ecosystem management objectives including conservation of wildlife, water resources, and rare ecosystems. This increasing demand requires frequent measurement of forest parameters, and as a result, there are significant advantages for improving both the precision and efficiency of data collection. To supplement costly and time intensive field surveys, forest product industries are investigating the ability to extract high quality, efficient forest information through the use of remote sensing.

A description of the forest – a forest inventory – is a necessity for rational forest management. Forest management today is increasingly defined by values, cultures, communities, and politics but the forest inventory continues to represent the critical component (Carter *et al.* 1997). Conventionally, forest inventory data have been collected primarily by means of field surveys, which is both expensive and time-consuming. Construction and maintenance of a forest inventory system has frequently depended upon remote sensing - usually in the form of aerial photography. The modern forest inventory system has evolved into a complex, and multifaceted digital database along with associated analytical tasks all contained within a geographical information system (GIS). A forest inventory and GIS system contains information on forest structure, composition, and development; and it is based on the concepts of first delineating, then treating homogeneous areas as forest stands (Franklin 2001). For over fifty years, forest stands have been mapped through aerial photographic interpretation. Advancing technologies of global positioning systems (GPS) and imaging software has transferred the delineation process from a user defined interpretation of an aerial photograph to an iterative, on-screen process. Forest managers are now involved in creative technological advances to explore what a forest inventory should contain, and concomitantly, how the information could be acquired. Remote sensing data are emerging as important components of forest inventory information systems.

The primary applications of remote sensing have largely focused on identifying and mapping forest conditions at a specific time. Although these qualitative products represent a powerful observation tool to better understand and visualize the spatial distributions of forest resources, they fail to provide any quantitative measure to be

integrated with ancillary data for input to empirical forest planning models. However, with the evolving development and integration of remotely sensed data sources and the availability of analytical processing algorithms, geographic information systems (GIS) are becoming common place in forest management information systems. The merging of digital remotely sensed imagery with forest inventory ancillary data has appropriately been labeled geospatial " data integration" and provides an effective tool for forest management decision making.

Extensive research has investigated plot level estimates of forest measurements from remotely sensed data. However, forest management decisions are based upon stand level conditions. Plot level inventory sampling is a valid technique for traditional forest inventory systems, but integrating plot level based estimation models from remotely sensed data with this ancillary data may not be the most suitable technique for accurate and efficient estimates. Stand level registration models are presumed to minimize variability that might exist both within the stand as well as within the remotely sensed data. Consequently, the purpose of this research is to integrate measured stand level inventory data and two sources of digital remotely sensed data, Synthetic Aperture Radar (SAR) and Landsat Thematic Mapper (LTM), to examine the potential for estimation of forest inventory information at the forest stand level.

OBJECTIVES

Recently, the application of newly developed remote sensing SAR and LTM data has been explored for a variety of forest management purposes. Few researchers have reported results from these efforts to estimate stand attributes. The intent of this study is

to investigate techniques to best develop whole-stand volume and basal area prediction equations for slash pine plantations by applying multi-frequency and multi-polarized data collected from Synthetic Aperture Radar (SAR) in conjunction with LTM data. Prediction equations will be developed from inventory data collected from managed slash pine (*Pinus elliottii*) plantations in Georgia's southeast Coastal Plain. Specific objectives of this research are:

1. Develop a mathematical model using aggregated mean SAR backscatter data to predict volume and basal area per acre at the stand level. While developing this model the following components will be evaluated:
    A. Identifying SAR bands that provide the most predictive power for estimating volume and basal area.
    B. Evaluate the use of the variance associated with aggregating pixels for an individual band as predictor variables for estimating volume and basal area.

2. Develop a mathematical model using aggregated mean LTM spectral reflectance data from both a summer, leaf-on and a winter, leaf-off scene to predict volume and basal area per acre at the stand level. The following components will be evaluated:
    A. Identifying LTM bands that provide the most predictive power for estimating volume and basal area per acre.
    B. Evaluate the use of the variance associated with aggregating pixels for an individual band as predictor variables for estimating volume and basal area.

3. Develop a mathematical model from combining SAR mean backscatter data and LTM mean spectral reflectance data to predict volume and basal area per acre at the stand level. The following components will be evaluated:
    A. Identifying SAR and LTM variables that provide the most predictive power for estimating volume and basal area per acre.
    B. Evaluate the use of the variance associated with aggregating

pixels for an individual band as predictor variables for estimating volume and basal area per acre.

4. Develop independent set whole-stand models for volume and basal area from stand inventory data provided from Plum Creek Timber Company's inventory database, to be used to compare the developed models from Model 1 and Model 2 above.

5. Apply two model development approaches, and contrast their performance to achieve the objectives stated above:
   A. Stepwise Multiple-Linear Regression
   B. A Biologically Motivated Multiple-Linear Regression Model

JUSTIFICATION

Remote sensing technology has emerged as a support tool for data collection and analytical methods for use in forest management planning. Digital remote sensing has developed from the implementation and technological advances surrounding aerial photography and photo-interpretation. With respect to forest management, information extracted from aerial photography is generally well understood and heavily integrated within forest inventory systems. In contrast, forest managers, perhaps due to an underlying assumption of the inherent knowledge required to interpret and integrate this new source of information, have been hesitant to apply information extracted from digital remote sensing. But as with remote sensing technology, forest management and forest managers are changing. The forest products industry is investigating opportunities to improve the quality of information available to forest managers through the use of

satellite data (Whiffen *et al.* 2000). Spatial data including SAR and LTM have been examined to facilitate this call for new technology for forest management.

In forest inventory systems, managers depend upon accurate and efficient forest estimates, primarily to predict and project forest growth. Remotely sensed data retrieved over various forest types provide unique spectral responses that are specifically determined by the forest's composition and characteristics. To investigate the potential of new remote sensing techniques, the relationship between SAR backscatter and LTM spectral responses with slash pine (*Pinus elliottii* Englem.) was examined to assess SAR and LTM data predictive capability for estimating volume and basal area per acre. Fitted prediction models generated using various techniques may determine whether pure statistical relationships or a more biologically model provides more appropriate predictive models.

BACKGROUND OF REMOTE SENSING FOR FORESTRY APPLICATIONS

Remote sensing has been defined as, "…the science of detection, recognition, and evaluation of an object, region, or phenomenon through analysis of informative data acquired by means of distant sensing and recording devices (Franklin 2001)." The scientific principle behind remote sensing is a method to measure structural and dielectric properties that can not be measured by touch, but rather can be measured in terms of the amount of radiation they absorb, emit, or reflect at various electromagnetic wavelengths.

Forest managers have a long-standing history of using remote sensing techniques, such as aerial photography to assist in quantifying and characterizing terrestrial ecosystems at the landscape level. As early as the 1920's, foresters implemented aerial

photography interpretation to create cover type maps and perform land area estimations(Baker 1970). By the mid-1960's, the ability to acquire beneficial remotely sensed data from timberland resources under commercial management was an invaluable asset to the major forest products companies in the South. It has been estimated that in 1970, approximately 82% of the commercial timberland holdings in the South were in some manner managed using aerial photography (Baker 1970). These management techniques involved timber procurement, stand mapping, reforestation planning and road and property line delineation.

The transition of commercial forest products industry to the use of satellite based remote sensing techniques to manage forestlands did not evolve until the early 1970's when NASA, in conjunction with USGS, launched the first satellite specifically designed to provide repetitive global coverage of the Earth's surface. Aerial photography and optical remote sensing systems have been widely accepted as viable tools for forest management decision-making, and as a result these systems have been commonly applied within the forest products industry. Despite of their wide acceptance, these remote sensing techniques have shortcomings with respect to ecological applications. Both aerial photography and optical imaging systems are "passive" sensing systems that are dependent on solar radiation to act as a light source to create an image. Both techniques use lenses and sensors to capture reflected light in to create an image of the terrestrial surface. An image created by either aerial photography or an optical imaging system is an interpretation of how the solar light interacts, in terms of absorption or reflection from the surface's features. The amount of light absorbed and/or reflected is highly sensitive to biophysical attributes, primarily the moisture content in the topsoil and the percent of

chlorophyll in the ground feature. Further operational limitations have been addressed for optical imaging systems specifically for forestry by Le Toan *et al.*(1992) who state that despite encouraging results, the limitations of optical data include the following points: 1) inadequacy to retrieve forest parameters from mixed, heterogeneous stands; 2) sensitivity of reflectance data to forest parameters is not significant after total canopy closure; and 3) in cases of incomplete crown closure, understory vegetation presents a disturbing effect both with the ability to retrieve and the validity of the retrieved reflectance data as it may appear as data noise that may need to be accounted for when determining relationships.

Landsat Thematic Mapper (LTM) program, presently under supervision of NASA and the Department of Defense, was developed to evaluate the use of space-borne remote sensing systems to monitor the Earth's surface. The Landsat-5 Thematic Mapper satellite was launched on March 1, 1984. LTM-5 has onboard a 7-channel scanning radiometer calibrated to record reflected energy from the Earth's surface at seven distinct fractions of the electromagnetic spectrum. Optical remote sensing systems, like the LTM-5 system, make use of visible, near-infrared, and short-wave infrared wavelengths of the electromagnetic spectrum to obtain an image of the feature of interest on the Earth's surface (Figure 1). Recording the intensity of the radiation reflected by the ground's features creates images. Biological entities absorb and reflect energy at distinct frequencies. The level of radiation absorbed and reflected varies across the electromagnetic spectrum. Since unlike materials absorb and reflect varying percentages of radiation along differing wavelengths of the electromagnetic spectrum, using spectral

responses from ground features enables users to differentiate between various cover types, as well as physical attributes from these ground features.

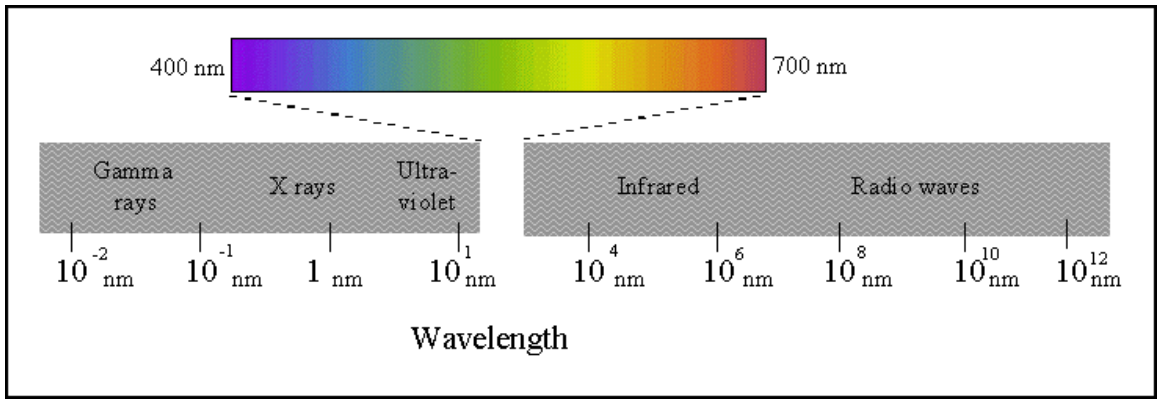


Figure 1. The Electromagnetic Spectrum (Lillesand and Kieffer 1994)

LTM-5's seven channels were calibrated to maximize the discrimination and monitoring of different types of natural resources (Figure 2). Channel 1, more commonly referred to as the "blue band" records reflected energy in the 0.45-0.52 micrometers(μm) range. Designed for bathyometric applications, Channel 1 is also used to distinguish between vegetation covers and soil covers. Channel 2, the "green band", is measured in the 0.52-0.60 micrometers(μm) range and is sensitive to peak green reflectance from healthy vegetation cover which can aide in the assessment of vegetation vigor. Channel 3, the "red" band, is measured in the 0.63-0.69 micrometers(μm) range is sensitive to chlorophyll light absorption by vegetation, and as a result, is beneficial for differentiating between various vegetated covers. Channels 4-7 record reflected energy from portions of the spectrum that are invisible to the naked eye, 0.76-0.90, 1.55-1.75, 10.4-12.5, and 2.08-2.35 micrometers(μm) respectively. Channel 4 is the near-infrared band and is most sensitive to moisture. It is useful for detecting and identifying healthy vegetation cover

as well as distinguishing bodies of water. Channel 5 is a mid-infrared band indicative of vegetation and soil moisture content. Channel 7 also is a mid-infrared band and is most suited for vegetation and soil moisture distinction. Band 6 records data from the thermal infrared, but is not applied for this research due to its larger spatial resolution (Lillesand and Kieffer 1994)

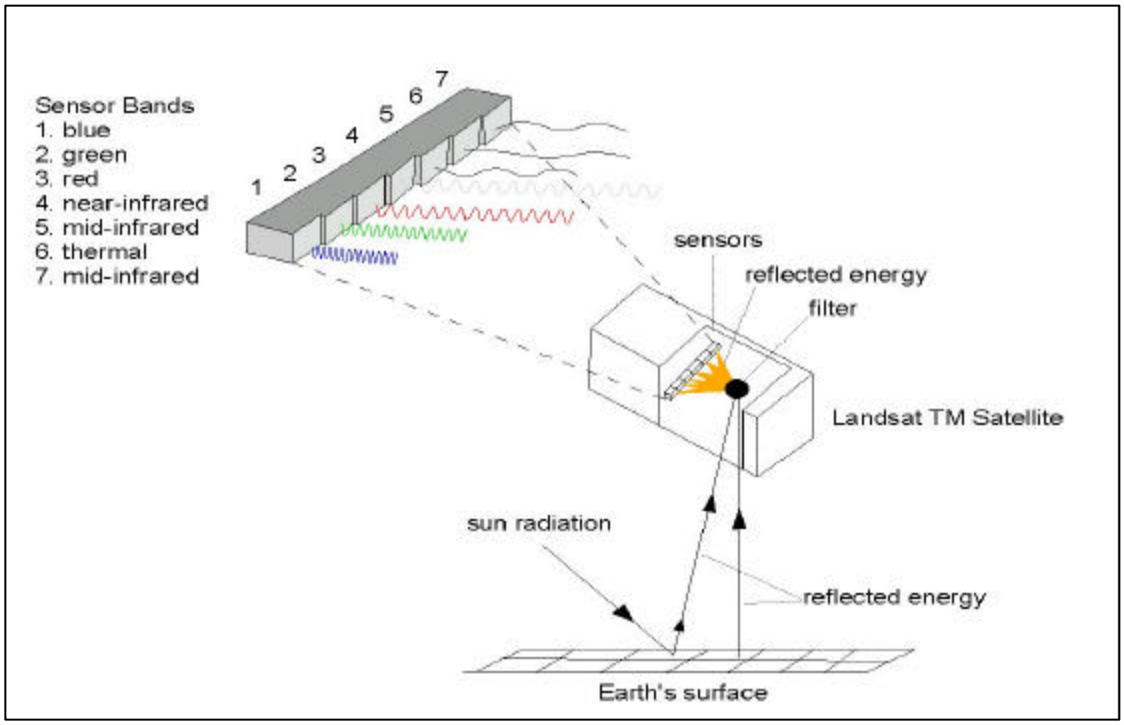


Figure 2. LTM-5 description and visualization

Remotely sensed imaging Radio Detection and Ranging system (RADAR) was developed as a means of using electromagnetic radio waves to detect the presence of ground features, as well as, determine their range or position (Lillesand and Kieffer, 1994). Side-looking airborne radar (SLAR) was first developed during World War II primarily for military reconnaissance. The initial benefit of using radar systems was that radar functions independently of solar radiation, which enables image collection without concern for time of day or atmospheric conditions. The ability to collect remotely sensed

data unaffected by atmospheric conditions is particularly attractive for ecological applications, and particularly forestry due to the range of atmospheric conditions and geographic locations of the majority of worldwide forestlands.

Use of remotely sensed imaging radar provides a number of distinct advantages for applications related to forest management (Chipman *et al.* 2000). Radar differs from optical remote sensing by the manner in which the two sensor systems generate and collect images. In contrast to the inherently passive approach of aerial photography and optical imaging systems that rely solely on reflected solar energy to generate an image, synthetic aperture radar (SAR) systems are an active imaging system that produce their own illumination. This characteristic mitigates the limitations of atmospheric moisture content and canopy interference experienced by photographic and optical imaging systems. SAR's illumination is generated through the transmission of short, powerful pulses of microwave energy. Microwaves are electromagnetic (EM) waves in approximately the 1-1000 GHz region of the EM spectrum (Figure 1). Radar antennas transmit pulses at various frequencies delineated by discrete bands according to the frequency at which the pulse is transmitted. Bands are labeled by a single letter respectively from shortest to longest wavelengths K, X, C, L, S, and P (Table 1). SAR systems in application for ecological purposes primarily use the C, L, and P bands with the corresponding frequencies; C-band (8,000- 4,000 MHz), P-band (2,000 - 1,000 MHz), and L-band (1,000 - 300 MHz).

Radar antennas and receivers measure the duration of time that is required before the emitted microwave pulse is reflected off a distant surface feature and is returned to the radar platform. Reflected pulses are also measured in terms of their amplitude, or

intensity, versus the strength of their original transmission. Microwave pulse's returned echo is recorded by the radar's antenna and is referred to as radar backscatter. Backscatter is measured for each discrete band transmitted by the antenna. Reflection from continuous pulse results in a unique spectral representation of ground surface features, and this spectral response is known as the radar's signature. By measuring the time delay between the transmission of a pulse and the reception of the backscatter from different ground features as the antenna platform moves forward, backscatter signals are able to provide an attribute backscatter coefficient for the ground features. Attribute coefficients are measured in decibels (dB) and are later processed accordingly to their specific frequency to construct a multi-dimensional image that is a spectral interpretation of the landcover's features.

Table 1. Radar Band Designations (Lillesand and Kieffer 1994)

Band Designation	Wavelength (cm)	Frequency v = ? [MHz (10^6 cycles $secs^{-1}$)]
K_a	0.75 – 1.10	40,000 – 26,500
K	1.10 – 1.67	26,500 – 18,000
K_u	1.67 – 2.40	18,000 – 12,500
X	2.40 – 3.75	12,500 – 8,000
C	3.75 – 7.50	8,000 – 4,000
S	7.50 – 15.0	4,000 – 2,000
L	15.0 – 30.0	2,000 – 1,000
P	30.0 – 100	1,000 – 300

Backscatter from various layers of a forest depends on how strong radar signals interact with each layer. Strength of interaction is determined by the intensity of the radar parameter being employed. Backscatter retrieved from surface features are affected primarily by two separate components of the feature as detailed in Figure 2: 1)

geometrical factors related to structural attributes of the surface and any overlaying vegetation-cover sensitive to the sensor properties of wavelength and polarization, or orientation, and 2) dielectric properties mainly controlled by the relative moisture content of the underlying vegetation and surface. Both sets of factors may be time-variant (Dobson *et al.* 1995). Because of the resonance of water in the microwave region of the EM, the dielectric properties control the magnitude and relative phase of backscatter as a function of wavelength. Structure attributes also exert control over frequency, polarization, and relative phase of radar backscatter. Within a given resolution cell, the structural attributes of importance are known to be the tree's architecture (size, shape, and orientation of woody limbs); leaf size, shape and orientation, and surface roughness (Dobson *et al.* 1995). More simply stated, it can be asserted with confidence that the expression of surface-related scattering in total forest backscatter depends on forest dynamics, forest floor properties, radar wavelength, polarization and incidence angle. As a result, backscatter from identical ground features may have significant variations if different wavelengths, polarization, observation times, or observation angles are applied.

Discrete radar bands each have varying sensitivity to identical components of a ground feature; as a result, certain bands experience higher incidences of scattering. This sensitivity can be examined when looking at a forest and its components (Figure 3). Crown layers play a crucial role in radar-vegetation interaction because it is the first point of contact and likewise, first point of interaction. Radar cross-section at the shorter C-band (3.75 – 7.5 cm wavelengths) is most sensitive to the foliage and crown layer of the canopy and interacts within this crown layer, thus making this frequency very capable for crown biophysical parameter retrieval. Mid-range wavelengths, like the L-band (15 - 30

cm) are insensitive to small foliage and branches and effectively pass through the majority of the crown layer and interact with the lower canopy and branches. In crown layers of considerable density, penetration depth of the L-band can be significantly reduced resulting in an inability to interact with the upper-stem and branch networks. In this cases, the L-band may be as, if not more, suitable to retrieve crown parameters than the shorter C-band which would likely get completely absorbed or refracted due to the density of the crown layer. Very long wavelengths such as the P-band (30 - 100 cm), are likely to penetrate the crown and be sensitive to the primary components of woody biomass and possibly even the earth's surface (Stellingwerf and Hussin 1997).

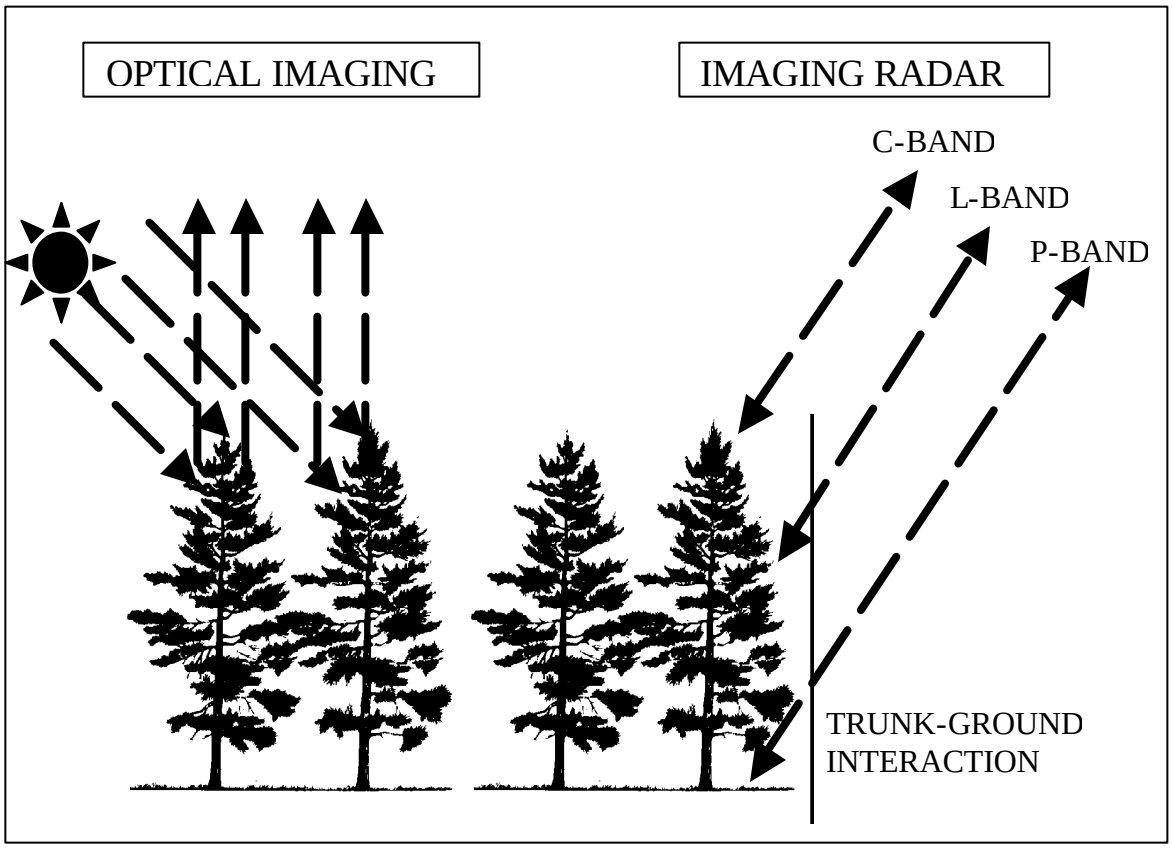


Figure 3. Optical Imaging versus Imaging Radar (Franklin 2001).

Polarization of the radar signal refers to the orientation of both the transmission and reception of electromagnetic waves. Examining different polarizations can often be useful when distinguishing between different types of ground features. Vegetative landcovers generate different signatures when the various polarizations are examined.

Polarization directly affects how the radar signals interact with ground features, and ultimately influences the resulting backscatter. Most radar imaging antennae are designed to transmit microwave radiation that is polarized both horizontally (H) and vertically (V) (Figure 5). Transmission of the pulse can either be for an individual polarization, or thanks to recent technological advances, the antenna has the ability to simultaneously transmit multi-polarized pulses. Generally, radar antennae also have the capability to receive either horizontally or vertically polarized backscatter energy, again for an individual polarization or for a multi-polarized reflectance. The respective letter of polarization denotes the orientation of the transmitted and reflected radar pulse. For example, a polarization labeled -HH denotes a "like-polarization" with a horizontal transmission of the radar pulse and horizontal reception of the backscatter. A polarization labeled -HV denotes a "cross-polarization" with horizontal transmission and vertical reception of the radar pulse. In all, there are four combinations of polarizations for radar -HH, -VV, -HV, and -VH.

Understanding the relationship between frequency and polarization is essential for understanding how radar generates data. It has been established that while the radar's antenna platform passes over the ground scene, the antenna has the ability to simultaneously transmit multi-frequency bands. The antenna also has the ability to

continuously transmit and receive multi-polarized energy for each of these multi-frequencies as it passes over the ground scene.

These two distinct characteristics allow the antenna to collect data from the three distinct band frequencies (C-, L- and P-bands) while simultaneously collecting multi-polarized radiation (-HH, -VV, -HV, and -VH). This allows a total of twelve combinations of band frequency and polarization to be simultaneously collected as the radar's antenna platform passes over the ground scene.

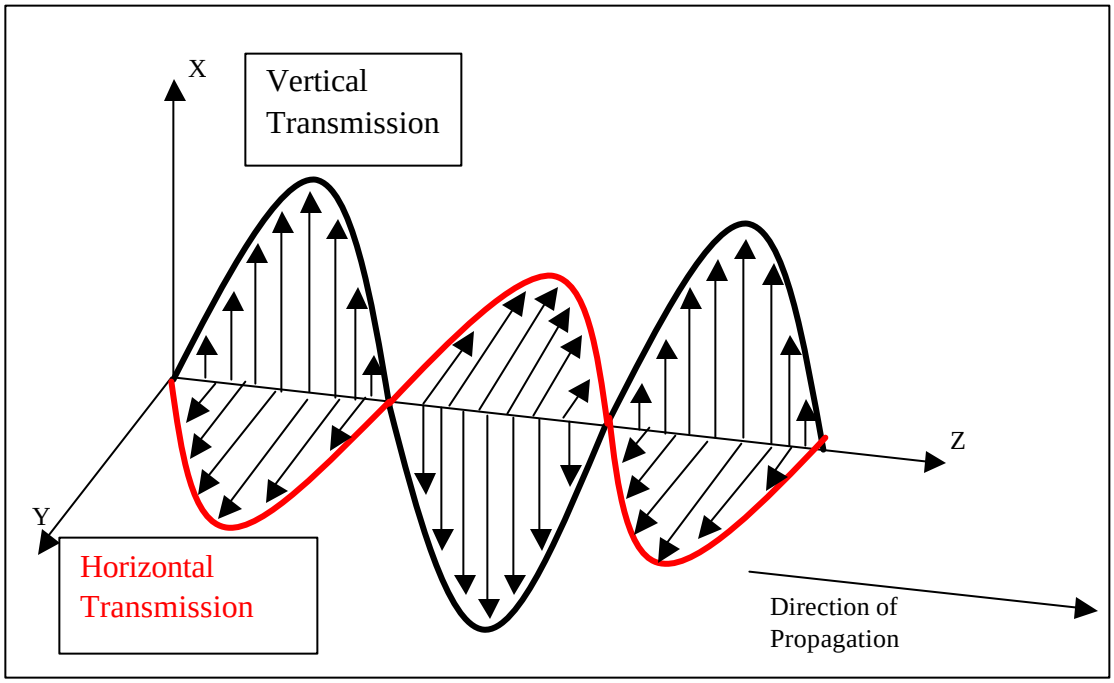


Figure 4. Orientation of Polarization of Radar Electromagnetic Waves

However, the SAR scene for this project was collected using only the C-, L-, and P-bands, and only the –HH, -VV, and –HV polarizations were acquired per frequency. As with SAR frequencies, the utility of different polarizations is sensitive to various components among forest structure and area capable of retrieving different backscatter.

The roles of SAR polarization similarly depend on the structural components of the ground target. A particular polarization interacts more strongly with features containing specific structural characteristics. For example, dually horizontal polarization (-HH) interacts more with features that have a horizontal structural component. Similarly, dually vertically polarized channels (-VV) interact more strongly with vertically oriented structural components such as foliage and branch networks. Within the context of biophysical attribute estimation, sensitivity of different polarizations to distinct structural components can be very useful for modeling purposes, particularly with estimations related to biological growth within the forest. With increasing age, crown layers develop continuously changing the structural complexity of the foliage and branch layers. More complex structural components depolarize the incoming radar signal. (Izzawati 1998). Use of cross-polarized signals (-HV) allows for more of the original signal to be reflected back to the sensor since it is not as susceptible as like-polarized signals to specific structural orientations. Structural change with the stem related to growth includes increases in height and stem diameter. Such increases strongly influence the strength of the trunk-ground interaction for dually horizontal polarization (-HH) and less so dually vertically polarized (-VV) signals. Again, use of a cross-polarized signal better allows the signal to overcome the absorption of similarly polarized signals by structural components of similar orientation (Izzawati 1998). For example, the cross-polarized -HV and -VH are found to be more sensitive to total biomass and hence yield higher correlation and predictive ability. Meanwhile, the like-polarized –HH and –VV are more sensitive to crown layer attributes and hence tend to saturate at lower levels of total biomass (Dobson *et al.* 1995).

CHAPTER II.

LITERATURE REVIEW

BACKGROUND OF SAR AND LTM APPLICATIONS

The past decade has witnessed significant growth in research focused on developing approaches for using synthetic aperture radar (SAR) to study ecological processes (Kasischke *et al.* 1997). Purposes of these research efforts have generally been twofold – classification and biophysical estimation. Spatial and statistical output from a classification procedure comprises one of the major information products on forest condition available by remote sensing (Franklin 2001). Classification produces information on features that are primarily categorical classes imposed on the image data. Continuous variable estimations produce information on features that vary continuously over the landscape depicted in the imagery.

For species classification, aerial photography has been the primary instrument for developing type mapping processes through photographic interpretation followed by validating field surveys. Digital classification has been used extensively for type mapping processes, but has been less frequently used when the objective is to develop classification variable classes as these generally resemble more the estimation of a continuous variable then a discrete categorization. Extensive research examining SAR's ability to perform accurate landcover classifications have been applied to categorize a SAR scene's landcover into four general structural classes: open water surfaces, short vegetation, tall vegetation, and urban features (Van Zyl and Burnette 1992, Rignot and

Chelleppa 1992, Wong and Posner 1993, Pierce *et al.* 1998, Ranson and Sun 1993, Rignot et al 1994, Dobson *et al.* 1995, and Dobson *et al.* 1996). Unlike classification algorithms, which are generally derived by statistical analysis of the unique spectral responses, empirical models are derived from defining a relationship incorporated from the physically based radiative transfer in forests (Franklin 2001). Broad-scale estimation of biophysical properties of forests represents the second central research objective of these research efforts. SAR's ability to develop predictive models from such spectral relationships that are capable of retrieving accurate and reliable estimations of forest stand attributes has gained increasing significance and attention (Hussin *et al.* 1991, Le Toan *et al.* 1992, Lang *et al.* 1993, Dobson *et al.* 1995, Dobson *et al.* 1996, Kasischke *et al.* 1995, Fransson and Israelson 1999, and Vona 2000). More recently there has also been limited research which has examined the synergy of SAR and optical remote sensing to obtain more accurate and precise classification and attribute estimation (Leckie 1990, Rignot *et al.* 1994, Pohl and Van Genderson 1998, and Kuplich *et al.* 2000). Collectively, research efforts have been directed towards developing a better understanding of the dependence of microwave backscatter on variations in forest structure, as well as, individual forest attributes particularly total stand volume, basal area, and height. Findings from these studies have provided wide and varying conclusions, and may in fact be incomparable as to the most appropriate methodology to best model the correlations. Inconsistent findings from these past studies attempting to analyze the correlation between measured radar backscatter and forest attributes such as basal area and volume have been attributed related to the limited number of radar frequencies and polarizations used or available and, in certain instances, the numbers of

stand characteristics considered. Kasischke *et al.* (1995) addressed these and other limitations and recognized that where significant correlations have been found in these studies, the correlations that were found tended to vary from study to study. He stated the limitations are in fact: 1) the availability of band frequencies and polarizations used or available at the time of the study to examine the correlation between SAR and stand parameters; 2) the dynamics and composition of the forests which were used in the research were strongly homogeneous, generally dominated by a particular species type, basal area, or within a given age-range; and 3) the range of forest components examined for a correlation with backscatter have been limited.

The principle hypothesis for using SAR systems to monitor terrestrial ecosystems is that a variety of surface characteristics unobtainable from LTM data may be uniquely detectable by SAR and possibly lead to improvement in the capability not only to classify landcover, but also to retrieve attribute estimates (Dobson *et al.* 1991, Dobson *et al.* 1992, Hussin *et al.* 1991, Dobson *et al.* 1995, Kasischke *et al.* 1995, Stellingwerf and Hussin 1997, and Fransson and Israelsson 1999). It has been established that SAR is quite capable of identifying major structural differences in landcover, such as forest versus clearcuts, or marshes versus bottomland hardwoods which provides useful forest composition information for forest inventory and management purposes (Dobson *et al.*, 1995). But more importantly, there is great potential for SAR's proficiency for identifying and measuring differences in aboveground biomass, especially following a disturbance such as clearcutting or forest fires (EOS SAR Facility Instrument Team 1991). SAR's potential monitoring capability for classifying and estimating forest attributes is supported on the basis that temperate pine forest systems, similar to those

predominately found in the southeastern United States, experience low biomass levels during the establishment phase of their successional growth periods (Kasischke *et al.* 1994). This successional growth generally occurs following recently established stands, an intensive plantation management thinning or fertilization, or following a disturbance such as a fire or commercial timber harvest. The ability to identify, map, and estimate these initial biomass growth levels could be a significant tool in projecting future growth rates and volume estimates within the stand.

To enhance the understanding of SAR's capability for forest discrimination and predictability, recent research efforts have attempted to overcome earlier limitations of SAR research. These new efforts have included the use of multi-temporal, multi-frequency, multi-polarization, and the combination of SAR with complementary optically remote sensed data for forest classification and attribute estimation.

Use of multi-temporal data in predictive models has significantly improved the accuracy and precision of the model based predictions (Fransson and Israelsson 1999 and Ranson and Sun 1993). The idea behind temporal data is that although remotely sensed bands may have the capability to penetrate and interact with differing components, they are still greatly influenced by the percent of crown closure or amount of foliage existing on a tree at a particular time. So a temporal approach allows for creation of a "leaf-on – leaf-off" image acquisition approach that accounts for seasonal fluctuations of crown/foliage conditions that species types may exhibit. As a result, use of multi-temporal data also might allow better modeling of the understory and/or any within-stand heterogeneity influence that may exist. Multi-temporal data is also being examined on an annual basis to attempt to model growth rates to establish more accurate

projection models. Fransson and Israelsson (1999) directly address the use of multi-temporal data and conclude that temporal variation could be modeled as an additive effect that is constant throughout the range of the stem volume. This could be attributed to a number of causes, for instance varying temperature, soil moisture, and vegetation moisture conditions.

Use of multi-frequency SAR is a new approach in research efforts, and has been the result of improved technology in SAR data acquisition. Application of multi-frequency SAR fosters the concept that SAR's discrete frequencies are sensitive to specific characteristics of the forest attributes and that a multiple-frequency C-, L-, and P-band model may have higher predictive ability to account for the variation captured by band sensitivities. An assumption behind this premise is the amount of backscattering from the crown, trunk, and ground layers are directly dependent on how strong the interaction is between the radar's pulse and each of these forest layers. Sensitivity of scattering from the shortest wavelength (C- band) mainly comes from the crown layer, and it has no interaction with lower layers of the stand composition. Meanwhile, longer wavelengths (L- and P- bands) have been shown to possess higher sensitivity to lower layers of forest biomass (Wang *et al.* 1995, Luckman *et al.* 1997, and Vona 2000). This sensitivity is due to the longer wavelength's ability to penetrate the crown layer, which enables radar pulses to interact with the trunk, branch system, as well as the forest's surface. The hypothesis that a full C-, L-, and P-band model would be most successful for capturing the forest's variability was disproved by Rignot *et al.* (1994) who found the model with a combination of the C- and L-bands to have the highest landcover classification accuracy. Individually between C-, L-, and P-bands, classification

accuracies were 74%, 63%, and 60% respectively. Classification efforts for this research were only concerned with delineating between cover types of grasslands, coniferous and deciduous species, and water. Both C- and L-bands are more sensitive to physical and dielectric properties that would differentiate the spectral separation from these limited landcover types. But these findings leave speculation as to what might happen in a more homogenous species forest where the objective is not to classify forest cover type, but rather to retrieve forest attribute estimates.

The potential for employing multi-polarized SAR data assumes that a fully polarized SAR may provide a higher percent of correlation with the targeted ground feature since polarization orientation does have an influence on backscatter. Obtaining cross-polarized backscatter which was hypothesized to provide better forest textural properties than like-polarization, Rignot *et al.* (1994) found that use of cross-polarization was better suited to classify between species specific characteristics. Differing polarization's ability to obtain information from various crown layers from pine species with different structure was successfully shown by Beaudoin *et al.* (1994) who examined the affect of structural components on backscatter from like-polarized signals (-HH and –VV). Pine species were selected because of their changing structural composition within various layers of the crown. The upper most layer of crown layer contains predominately vertical branch orientation. Within this layer; dually vertical polarized (-VV) backscatter was much higher than dually horizontal polarized (-HH) signals. But as structural composition becomes more complex at lower layers of the crown, the two like-polarized signals' backscatter intensity became very similar. Beaudoin *et al.* (1994)

also concluded that as crown layers developed complexity, backscatter from cross-polarized (-HV) signals increased.

Combining imaging SAR data and optical imaging LTM data for forest resource applications is becoming of practical importance (Leckie 1990, Rignot *et al.* 1994, Pohl and Genderson 1998, and Kuplich *et al.* 2000). The rationale for combining the two imaging systems is that optical data is sensitive to physical properties that allow for vegetative differentiation, while SAR is more suited for providing data to extract physical attribute estimations. In efforts where this process has been explored, it was shown that SAR data in conjunction with optically remote sensed information improved the level of accuracy for forest type discrimination between softwoods and hardwoods. Leckie (1990) used the addition of SAR to a Landsat model, and recorded an increased model precision of approximately seven percent versus that from just Landsat data. Kuplich *et al.* (2000) also examined the addition of SAR data to a LTM classification model. Despite relatively low levels of model accuracy, overall correlation between data generated from the combination of SAR and Landsat bands and targets previously undetectable, suggests combining the two sources was a successful approach. In instances where ground targets were delineated, combining SAR and Landsat data increased the ability to discriminate target classes. Such increases were attributed to the unique spectral characteristics of optical and radar band properties, which are: 1) SAR's ability to penetrate crown closure, 2) Landsat's sensitivity to soil moisture to distinguish between pasture and barren soil, and 3) SAR's ability to identify planted forests as a result of their distinct backscatter behavior caused by their homogeneous composition in contrast to the backscatter of natural forests (Kuplich *et al.* 2000). Present research has

not examined the combination of SAR and LTM data for predicting forest stand level characteristics of height, basal area, and volume. However, early success from combining the two data sources provides justification for such an approach to be investigated.

PREDICTIVE ABILITY

Spatial and statistical output from classification procedures comprises one of the major information products on forest conditions available from remote sensing, but equally important, a second set of forestry information is obtained by estimating continuous variables located within the forest (Franklin 2001). Continuous variable estimation produces information of forest attributes that are likely to range in value or condition continuously over the landscape depicted in the imagery. Unlike classification efforts, which are systematically driven by statistical interpretation of spectral response patterns, empirical models are based more on underlying relationships due to radiative transfers of forest structure and dynamics (Franklin 2001).

Hussin *et al.* (1989) examined correlations between SAR L- and P-bands with southern pine plantations in northern Florida. The authors found significant correlations between both P- and L-bands and forest volume, height, basal area, and trees-per-acre within early aged 4 –17 year old slash plantations, but plantations of ages 18-48 showed very weak correlations. Hussin *et al.* (1991) using the identical dataset from Hussin *et al.* (1989), used SAR L- and P-bands to fit a system of equations using a generalized least squares technique that predicted height, basal area, and total volume. The authors presented prediction models with R^2 of 0.98, 0.83, and 0.98 respectively for height, basal

area and volume with an average bias of less than 1%. Despite their overwhelming success, the authors' lack of discussion regarding their sources of inventory data, model development criteria, statistical analysis, and the concluding predictive model does not lend confidence in their results. With the percent of errors reported, their findings clearly suggest that their models are more capable of estimating stand attributes than a well applied ground based inventory.

Wang *et al.* (1998) using airborne SAR data to model stand attributes found that the L-band was more sensitive to variations in larger components of the tree's crown and stem, whereas the C-band is more sensitive to variations in small branches and foliage in the crown layer. A stepwise, multiple regression technique with multi-channel SAR data and possible SAR band ratios for the independent variables was used to generate prediction models. Simple, linear stepwise regression accounted for varying sensitivity between C- and L-bands by considering which SAR channels best model these variations in the SAR backscatter signature. Use of the ratio from cross-polarized frequencies L_{HV}/C_{HV} was an attempt to normalize the influence of canopy scattering and the attenuation in the canopy layer on the L_{HV} signature, which is most sensitive overall. Correlations using L_{HV}/C_{HV} ratios were slightly lower than correlations with each frequency independently, but found these ratios resulted in significantly lower RMSE when used to estimate biomass than using by either L_{HV} or C_{HV} alone.

In an attempt to better model variation in overall stand structure Dobson *et al.* (1995), developed a system of simultaneous equations to estimate total volume as a function of height and basal area prediction models. This approach recognized that although overall sensitivity to variation in total biomass was adequately modeled at low

levels of biomass (<15kgm^{-2}) in pine stands, higher levels of biomass were dominated by scattering from the crown layer with both C- and L-bands. Because stand attributes are highly correlated, a two-step system was also developed to predict total biomass.

Fransson and Israelsson (1999) developed a simple linear model to examine the relationship between SAR L-band backscatter from boreal conifer forests in northern Sweden and forest stem volume. Evaluating this model using 22 validation stands showed a correlation coefficient of 0.78 between SAR estimated stem volume and forest inventory data. The model indicated that stem volume from these stands could be reasonably estimated from SAR data.

A variety of techniques have been proposed to estimate above ground stand parameters using radar data, and the most success has been achieved lower frequency, L- and P-bands (Dobson *et al.* 1992, Kasischke *et al.* 1995, Le Toan *et al.* 1992, Ranson and Sun 1995, and Rignot *et al.* 1994). And although there are several techniques proposed to develop algorithms that best estimate forest stand attributes from SAR, there is still no certainty about an optimal technique. Harrell *et al.* (1997) compared leading techniques in an attempt to conclude which method is best suited for estimating stand attributes in southern United States pine forests. Four different approaches were evaluated:

1) Simple Radar Cross Section (Simple RCS) Model
   - Total volume was estimated by use of a single-step method using stepwise, multiple-linear regression with various radar channels as the independent variables.
2) Polarization Ratio Model (Model adapted from Ranson and Sun 1994)
   - Total volume was estimated, using a one-step method using a simple linear regression model with the ratio of the low-frequency and the high-frequency SAR data as the independent variable.

3) Dobson Method (Model adapted from Dobson *et al.* 1995)
   - A system of three simultaneous equations in which;
   a) multi-channel SAR data are used to estimate crown biomass, basal area, and height
   b) trunk biomass is then estimated from predicted basal area and height, using allometric equations developed from the ground-truthed data
   c) crown and trunk biomass then provided for total biomass

4) Kasischke Method (model adapted from Kasischke *et al.* 1995)
   - A two equation system method whereby;
   a) The total branch biomass was estimated by using multi-channels of SAR data
   b) The remaining biomass components are estimated on the basis of their allometric relations to branch biomass.

Varying techniques were each developed with the specific intent to model the differential scattering and attenuation of multi-frequency and multi-polarization SAR data within the different layers of the stand's canopy.

Harrell *et al.* (1997) research was not designed to present a specific quantitative comparison between presented models' level of predictability; rather it addressed success of the various published modeling approaches. Findings from this comparative research effort were consistent with both earlier studies from data collected over the Duke Forest Supersite located in Raleigh-Durham, North Carolina and earlier efforts to estimate total stand volume in these forests (Dobson *et al.* 1992, Dobson *et al.* 1995, Kasischke *et al.* 1994, Kasischke *et al.* 1995, Ranson and Sun 1995, and Wang *et al.* 1994, and Wang *et al.* 1995). Results indicated that multi-channel SAR data improve total stand volume estimation. Specifically, those instances that developed estimations from a system of

multi-step equations which modeled specific components of volume independently – such as the Dobson *et al.* (1995) and the Kasischke (1995) models had the highest accuracy of predictability. However, a simple RCS model which modeled total volume from a selection of multi-channel radar data had very comparable rates of error, and actually out-performed all others when the all stands were included in the estimation process.

In general, continuous variable estimation from remotely sensed data are based upon weak relationships between spectral responses and stand structure as captured by the imagery of differences in illumination of forest canopies. The earliest idea behind variable estimate assessment with remote sensing was that enough correlation would exist between what is detected by the sensor and stand structure, such that attribute estimations could be predicted from sensor data without necessarily separating out or even understanding the various influences on the signal (Poso *et al.* 1984). Much research has been directed at developing a better understanding of the dependence of remotely sensed spectral reflectance on variations in forest structure, with the ultimate goal to develop methods which use either optical and/or imaging radar to estimate biophysical attributes from forest ecosystems. The primary focus of early research efforts have been to identify correlations between SAR and LTM data with forest characteristics such as stand age, height, basal area, and volume. Limited effort has attempted to develop predictive models from these correlations. In efforts where prediction equations were developed from these correlations, SAR's calibration was either from a sampled-plot measurement scale or generated specific Areas-of-Interest (AOIs). Both techniques attempt to register spectral response data with stand inventory

data on a per pixel basis and later *"blow-up"* the model to a stand level basis (Kasischke *et al.*1995, Le Toan *et al.* 1992, Vona 2001, Phelps 2001, and Sader 1987). In other instances, selection criteria for participating stands based on *a priori* knowledge was so narrow that often times selection resulted in a small sample with little variation among the selected sites used to develop the predictive model. This selection criterion brings to question the range of data used to fit and validate these models (Kasischke *et al.* 1994, Kasischke *et al.* 1995, and Wang *et al.* 1998). If too few are used to develop the relationship, a tight fit could be found, but the results could only be applied locally for that site since as deviation from the calibration site increases the relationships weaken. (Iverson *et al.* 1990). If to many are used, there is a greater possibility of introducing large-scale uncontrolled variation such as climate soils, silvicultural techniques, or natural understory. These influences of in-stand heterogeneity cause variability in spectral responses. However, since installing measurement plots is a timely and costly investment, especially for remote sensing experiment research, as is often times the case where fewer plots were used. Results from this strategy generally overestimate of the power of the few prediction equations that have been developed (Franklin 2001).

CHAPTER III.

METHODS AND MATERIALS

The research site for this project is defined by the SAR scene, and is centered on 82°00'00" West longitude, 31°75'00" North latitude (NAD27) which encompasses Jesup, Georgia and forestlands surrounding this Lower Coastal Plain Region of Georgia. Natural and planted loblolly (*Pinus taeda* L.) and slash *(Pinus elliottii* Englem.*)* pine dominate the research area. Minimal populations of hardwood species including red maple (*Acer rubrum*), pond cypress (*Taxodium distichum* var. *nutans*), and sweet gum (*Liquidambar styraciflua*) are scattered throughout the sites. Plum Creek Timber Company (TTC) owns and manages a large percent of timberland on this particular study site, as well as the majority of surrounding timberland.

SAR DATA ACQUISITION

SAR data were collected on August 31, 1998 through a federally funded, collaborative research project between the Jet Propulsion Laboratory (JPL), Envisense Corporation, and The Timber Company. The Jet Propulsion Laboratory utilized an airborne SAR system (AIRSAR) attached to a DC-8 flying at an altitude of 10,000 m. The radar swath (east-west width of the imaged scene) was approximately 5 miles wide with an azimuth (north-south direction) distance of 20 miles, covering approximately 44,000 acres. SAR data was collected at multi-frequencies, C-, L-, and P-bands, with multi-polarizations of -HH, -VV, and -HV per frequency. This collection process

provided nine possible combinations of frequency and polarization to be used for this analysis (Table 2).

Table 2. SAR Band Notation

SAR Data Description	Band Frequency and Polarization
SAR BAND 1	C_{HH}
SAR BAND 2	C_{VV}
SAR BAND 3	C_{HV}
SAR BAND 4	L_{HH}
SAR BAND 5	L_{VV}
SAR BAND 6	L_{HV}
SAR BAND 7	P_{HH}
SAR BAND 8	P_{VV}
SAR BAND 9	P_{HV}

To transfer backscatter data to an image-processing format, a signal processing procedure was performed by JPL in Pasadena, California which presented the data in the form of a Mueller matrix. Envisense Corporation performed all remaining radiometric and geometric processing. Radiometric processing involved conversion from a 32-bit real power to an 8-bit dB data. Conversion resulted in a significant data reduction. For computational and display purposes, the conversion drastically decreased required time for on-screen processing and analysis. To achieve a stable radar signal the 8-bit data were filtered using an algorithm developed by Envisense and University of Michigan's Department of Electrical Engineering and Computer Science. The Enhanced Edge Preserving Optimized Speckle Filter (EEPOS) significantly reduces the fading effect by preserving linear and point features to a high degree.

Air-SAR data was converted from the complex Mueller matrix to a 32-bit real image channel. The image was also converted from a slant range to ground range using

algorithms in PCI Geomatica imaging processing software. Orthorectification was performed to register the image by selecting unique and clearly visible ground control points between the SAR ground range image and the USGS 1:24,000 Digital Raster Graphic (DRG) in UTM projection using North American Datum 1927 (NAD27) meters.

Radiometric correction was performed to convert the received power into a backscatter coefficient that accounted for both range loss and local area illumination. Spatial filtering and radiometric segmentation were applied to ensure high spatial and radiometric resolution would be retained while minimizing multiplicative 'noise' in the data. SAR data were finally resampled to a 6.25 x 6.25 meter pixel spatial resolution.

Upon developing the design for the research, it was decided that the projection of the data would be UTM Zone 17N NAD (1983). This criterion was determined on the basis of the projections of the available coverages and LTM data. Knowing that the two LTM scenes were previously geo-referenced in this projection and the shear magnitude required to reproject two Landsat images, it was a logical decision to select UTM Zone 17N NAD (1983) as the standard projection for the project. To comply with this criterion, SAR imagery was reprojected to UTM Zone 17N NAD (1983) using an ERDAS Imagine Projection utility.

LTM DATA ACQUISITION

Landsat Thematic Mapper (LTM) data was acquired multi-temporally, with the summer scene being recorded on May 12, 1998 and the winter scene recorded on December 19, 1998. Temporal scenes are accordingly referred to as "leaf-on" and "leaf-off" respectively due to foliage's seasonal fluctuation. Both scenes recorded data for

Row 17, Path 38 along the satellites orbital path, which encompasses the research site in Georgia's Coastal Plain region. The benefit of acquiring multi-temporal data is that variation in above ground biomass due to seasonal variation may be accounted for and possibly better modeled through recording and measuring these levels of variations. Also, multi-temporal data allows more in-stand heterogeneity from hardwood encroachments or ground cover shrubbery can be identified between scenes and better accounted for.

LTM-5 recorded data at seven distinct bands (portions) of the electromagnetic spectrum. The resulting product is a 7-layer image with each layer containing reflectance information from different portions of the electromagnetic spectrum. Layers 1 – 5 and 7 have a 30-meter ground resolution, while layer 6 has a 120-meter ground resolution. For this research LTM Band-6 was not used due to its resolution characteristics. Band- 1, 2, 3, 4, 5, and 7 were incorporated for study. Information recorded from each of the seven channels was converted to 8-bit data, containing 256 shades of grey, and transferred to one of 16 worldwide ground receiving stations. The resulting product was a seven-layer image consisting of spectral reflectance data from differing portions of the electromagnetic spectrum. LTM data layers were recorded at a lower spatial resolution of 30-meter pixels than that of the SAR's 6.25-meter resolution.

FOREST INVENTORY DATA

The Timber Company (TTC) performed forest inventory sampling continuously during 1998. Inventory sampling data were collected in accordance with TTC standard inventory sampling design, which consists of a 10-factor, point-sampling technique with

an allowable error of ten percent to determine an appropriate number of sample plots. In instances where the shape of the stand was not conducive to this design, a minimum of 20 points were sampled within each stand. Spacing for inventory cruises was a 3 x 3-chain grid. A total tally of trees within the 10-factor radius, along with a dominant and co-dominant height was collected per plot to calculate estimates of trees per acre, basal area, average height, and total volume to then calculate stand level inventory data. Also accumulated from the TTC inventory dataset was establishment age and site index for the individual stands.

A digital stand boundary coverage was provided by TTC that consisted of their regions 19 and 22 which contained stands within both the SAR and the LTM scenes. Stand digital coverage was in an ArcView.shapefile format and provided a unique stand ID for each stand in the region. Projection of this digital coverage was UTM Zone 17N NAD 83 feet. Using an ArcView Projection utility, the coverage was reprojected to UTM Zone 17N NAD 83 meters to be consistent with the projection of SAR and LTM scenes. Consistency with data projection ensures proper overlay and spatial registration of stand boundaries to the SAR and LTM scenes (Liellesand and Kieffer 1994). Once the stand boundary coverage and SAR and LTM scenes were reprojected to a consistent projection, stand level inventory data was spatially joined to the existing digital stand boundary shapefile coverage by using the stand ID field as the linking attribute. These processes attached stand level inventory data to the stand boundary shapefile coverage so it consisted of all data required to perform the analysis.

Specific criteria were used to select feasible stands for analysis. Criteria were: 1) stands must be completely within the SAR scene; 2) stands must be planted pine

plantations; 3) stands must have been established by time of SAR acquisition; 4) stands must be a minimum of 5 acres; and 5) stands must not have been thinned or harvested prior to SAR and LTM acquisition or between the time of SAR and LTM acquisition and the latest inventory cruise. Query selection resulted in providing a possible 203 stands which satisfied these preliminary criteria.

However from these possible 203 stands, a second query process removed that were identified as having a stand basal area of zero. Removal of these stands was determined to be appropriate since Plum Creek's inventory sampling design strategy does not perform inventory sampling on stands prior to age eight. There was concern young stands (ages 4-8) and newly replanted stands might provide erroneous data by suggesting a stand at age 7 (which may have a considerable amount of basal area and possibly merchantable volume) would have the same spectral response as would a newly planted stand. In a remote sensing perspective, these stands due to their different stand dynamics and attributes' composition would likely have differing spectral response patterns. Removing stands identified as having a basal area of zero also had the effect of removing stands with total volume of zero from the dataset resulting in a possible 152 stands.

Visual examination of these 152 stands revealed seven stands either violated the criterion of appropriate number of pixels to obtain reliable spectral information or geographic location difficulties that supported their removal from the study, and they were subsequently removed. Statistical analysis determined that loblolly and slash pine species were significantly different from each other and loblolly were removed from the candidate stands. The selection process concluded with 109 slash pine stands that

satisfied all criteria. Table 3 presents the distribution and ranges of the stands' inventory data for these 109 stands.

Table 3. Stand Inventory Data for 109 Slash Pine Stands.

INVENTORY	MINIMUM	MAXIMUM	AVERAGE
Stand Acres	6.00	394.00	66.18
Stand Age	10.00	36.00	18.73
Trees Per Acre	104.26	735.00	370.50
Stand Height	43.20	86.48	53.56
Basal Area ft^2/acre	34.08	152.20	87.50
Volume tons/acre	15.22	146.40	59.30

DIGITAL PROCESSING

Developing attribute estimation models by registering remotely sensed data with measured plot level data may have certain violations of error that are not problematic at the plot level where the model calibration occurs, but when the models are attempted to be "blown up" to the stand level may become troublesome. Developing an inventory cruise requires that a predetermined decision must be made to determine how close of an estimate to the true parameter the sample estimate is desired (Shiver and Borders 1996). Specification of this *"allowable error"* requires calculation of an appropriate sample size to satisfy this level of accuracy. By defining this level of accuracy, a constant measure of error is independently associated with each sampling unit which when expanded to the stand level, is within the bounds specified by the allowable error. Practices of sampling at a plot level then "blowing up" data to a more desirable and manageable stand level is

common practice in forest inventory. However, registering remotely sensed data at the plot level and then extrapolating to stand level may not be well suited for developing attribute estimation models. With inventory sampling, each plot accounts for a "*constant*" level of error. Registering remotely sensed data to these plots violates this assumption of error. Error associated with plot level registration are not random, independent error. Plot level error with remotely sensed data that has been registered when aggregated are increasingly greater than those assumed and allowed for by the sample size. Further, expanding these errors to stand level estimates only further expands these errors. Individual plot level estimation may be very good from this technique, but collectively may be much more heavily influenced by the level of error than does the sampling technique.

This research applied a "stand level based" analysis. Rationale for this approach is that it is common for forests to be managed at the stand level, despite being measured at the plot level (Clutter *et al.* 1983). It already has been hypothesized that registration performed at the plot level is likely to parallel this technique and be extrapolated to the stand level. With such an approach, within stand heterogeneity or mis-registration of spectral responses with measured data is much more influential on accurate attribute estimation. Registering the remotely sensed data by aggregating spectral response at the stand level attempts to minimize the influence the error due to heterogeneity or possibility of mis-registration.

An aggregated mean backscatter coefficient and variance value was calculated for each stand. This follows the objective to examine SAR and LTM predictive ability at the stand level rather than the more commonly applied plot level. A customized ArcView

script was written that used the stand boundary coverage as a mask and aggregated the values of the underlying pixels from the SAR and LTM scenes independently to calculate the stand's zonal statistics. Prior to such a process, SAR and LTM scenes first had to be converted to a raster GRID format image, with each of the specific bands becoming an independent GRID retaining their SAR and LTM band coefficients.

Previous studies both within remote sensing literature and forest inventory systems have addressed error introduced due to edge-effects (Shivers and Borders 1996). In terms of forest inventory mapping and estimation with remote sensing techniques, this error could be the result of a poorly digitized stand boundary coverage or merely stand structural difference along stand boundaries. To examine this effect, stand boundaries were overlaid on the imagery. In some instances, stand boundaries appeared to match up directly with visible features from the imagery; while in other areas there was a clear, unsystematic shift, generally within a five-pixel shift. This shift was addressed with three separate techniques. The initial process was to manually edit the digital stand coverage to correct for misalignment of stand boundaries with the imagery. This very time-consuming and only marginally successful process, required editing the individual nodes of the digital coverage in order to align stand boundaries and the imagery. A second, and more effective, technique to minimize the amount of error introduced by edge-effects, buffered candidate stands inwards a distance of 30-feet or approximately one-half of a chain. Finally, a third technique to help minimize remaining error was created within the ArcView's script itself. Traditionally, when a polygon shapefile is used to overlay a raster image, any pixel that the polygon intersects is included in the aggregation. This means any SAR or LTM pixel that the stand boundary intersects will be used in

calculating a mean backscatter response for that particular stand. As a result, a stand boundary not "cleaned-up" by the buffering process or the presence of interior heterogeneity will carry significant weight in the averaging process due to this inclusion of incorrect pixel values. To reduce this introduced error caused by the stand-pixel overlay dilemma, the ArcView script was written to calculate mean backscatter at a 0.5 meter pixel size rather than SAR's 6.25 and LTM's 30 meter pixel size. The calculation processes internally resample underlying SAR and LTM raster pixels that intersect a stand boundary to a 0.5 meter pixel and calculates a mean backscatter coefficient per band using these much finer spatial resolution raster GRID pixels. The intent of this application is comparable to a weighted regression technique where it will weigh those pixels that are spectrally consistent with each other and encompass the true spectral representation of the stand's attribute heavier than any edge-effect or in-stand heterogeneity (Draper and Smith 1983). Their influence is minimized in the mean calculation by reducing the number of misclassified pixels and their influence on the mean backscatter calculation. The end result becomes a digital shapefile coverage that has stand level inventory data as well as a mean backscatter coefficient and associated variance from each SAR and LTM band for each unique stand. Upon analysis of the first attempt at this technique, it was evident that there were unrecorded SAR backscatter coefficients in the original SAR data. Errors were apparent when the zonal statistics were calculated that included identifying the minimum and maximum backscatter coefficient values for each particular band for each individual stand. In a few instances, minimum values were being recorded as zeros. Upon visual interpretation of these zero valued SAR pixels, it was determined that these values were erroneous due to their location

within the stand and surrounding pixel values. To correct for these zero values, a 3 x 3-pixel filter was passed across the SAR image using ERDAS Imagine imaging software. This 3 x 3 pixel filter calculated the mean of the nine pixels and assigned that value to the center pixel. Using both the original SAR scene and the newly averaged image, a new corrected SAR image was generated by developing a conditional statement in ERDAS Imagine's Model Builder that stated that if the original SAR image recorded a pixel value of zero for a particular SAR band coefficient then replace only that specific pixel with the new averaged image's pixel coefficient, otherwise maintain the original SAR coefficient. This process was successful in removing erroneous zeros from the SAR's band coefficients, and the zonal attribute script was rerun to calculate a stand level mean backscatter value and variance for each SAR and LTM band.

CHAPTER IV.

MODEL DEVELOPMENT

WHOLE-STAND MODEL

An underlying objective of this research is to compare two separate model development approaches to identify SAR and LTM bands that provide the most predictive power for estimating volume and basal area. Since forest dynamics and forest structure play an integral role on SAR and LTM spectral responses, biological stand models were appropriately developed from Pienaar growth and yield equations to generate a set whole-stand models representative of the Plum Creek inventory data. Development of such models assumes that volume is some function of age, basal area (BA), height (HD), and measure of stand density (TPA); and subsequently, basal area is a function of age, height, and stand density (Clutter *et al.* 1983). To identify these corresponding assumptions for predicting volume and basal area in model development format, the following was identified:

VOLUME = *f* [Age, Dominant Height, TPA, Basal Area]

BASAL AREA = f [Age, Dominant Height, TPA]

Using these whole-stand model assumptions, natural logarithmic transformations were performed on both the independent and dependent variables. Natural logarithmic transformations are consistent with growth and yield models that define volume and basal area as an exponential relationship. Furthermore, natural logarithmic transformations of the predicted yield of basal area and volume as the dependent variables is an appropriate

way to mathematically express the interaction of the independent variables effect on the dependent variable (Avery and Burkhart 1994). Performing this transformation for the whole stand basal area model became:

$$\text{LN } (BA) = b_0 + b_1 \text{Age}^{-1} + b_2 \text{LN(HD)} + b_3 \text{LN(TPA)}$$

F-value	Pr > F	MSE	R^2	Parameter	Estimate	Std. Error	Pr > |t|
193.4	<.0001	0.01188	0.8468	b_0	-3.00088	0.60461	<.0001
				b_1	-8.74191	1.44888	<.0001
				b_2	0.96199	0.11631	<.0001
				b_3	0.71691	0.03338	<.0001

However, upon fitting these original models, it was apparent that the model form did not adequately fit volume data. As a result the model was adapted to convey that the interaction of age may weigh heavily on whole-stand basal area. Allowing the stand level variables to be expressed with an inverse of stand age allows for the asymptotic effect of yield and basal area with increased age. The adapted stand level model for volume becomes:

$$\text{Ln } (VOL) = b_0 + b_1 \text{Age}^{-1} + b_2 \ln(\text{BA}) + b_3 \ln(\text{TPA}) + b_4 \ln(\text{BA})\text{Age}^{-1} +$$

$$b_5 \ln(\text{HD})\text{Age}^{-1} + b_6 \ln(\text{TPA})\text{Age}^{-1}$$

F-value	Pr > F	MSE	R^2	Parameter	Estimate	Std. Error	Pr > |t|
3225.16	<.0001	0.00183	0.9948	b_0	-8.2822	0.6794	<.0001
				b_1	77.6679	13.3051	<.0001
				b_2	2.0145	0.0972	<.0001
				b_3	0.7562	0.0569	<.0001
				b_4	20.8092	0.5582	<.0001
				b_5	-19.6230	1.9162	<.0001
				b_6	-16.0502	1.2014	<.0001

The newly modified whole-stand model for volume and the original model form for basal area were the best models for the Plum Creek inventory data. Residual plot analysis of these models suggested little heteroscedasticity and assumptions of

independent error terms and constant variance for linear regression analysis were not violated. However, when examining the residual plots from estimation of volume, it is obvious that two stands are negatively impacting the regression line, and are potential outliers. Upon visual inspection of the stands on both the SAR and LTM data there appeared to be no significant cause to remove the stands due to possible misregistration. Removal of the two stands without due justification is also not suggested. The two stands remained in the data, with the understanding that these will likely be a cause of higher variability at lower volumes for the fitted models. Despite these two stands, diagnostic statistics support this finding that these models sufficiently fit the data. Whole-stand models resulted in very high R^2 and low MSE of 0.9944 with 0.002 and 0.8424 with 0.0119, respectively for the natural logarithm of total volume and basal area (Appendix 1 and 2). Successful development of whole-stand models validate that TTC inventory data is reliable and capable of being modeled. Fitted whole-stand models also serve as a reference for model predictive capability for those models derived from SAR and LTM data.

RELATIONSHIP BETWEEN SAR AND LTM DATA WITH STAND PARAMETERS

Possible relationships between remotely sensed data and slash pine stand parameters were first examined graphically by plotting SAR and LTM bands against transformed stand volume and basal area. Results from plotting SAR and LTM bands against transformed stand data illustrate varied relationships between SAR and LTM bands and stand data (Figures -14).

SAR backscatter plots from C-, L-, and P-bands revealed that the shortest wavelength C-band and the longest wavelength P-band exhibited the strongest linear relationship with both volume and basal area. And among these bands it is evident that the cross-polarized C_{HV} and P_{HV} bands show the strongest relationships. Results from these scatter plots are consistent with previous literature that longer wavelength P-band is best suited to model stand structural characteristics such as volume and basal area because its electromagnetic characteristics to penetrate the crown layers and interact directly with stand's stems and branch layers.

Plots also support literature suggesting that cross-polarization is better suited to penetrate and reflect with a greater degree of intensity to allow a better representative predictor variable. Meanwhile, L_{HH}, L_{VV}, and L_{HV} have the weakest relationships, exhibiting minimally detectable linear relationships. Findings vary from previous literature that suggests L-band frequencies should exhibit stronger relationships than C-band channels, which is not the case here. This may be due to the aggregate stand level approach. Since L-band channels interact with the canopy as well as the underlying crown layer its variability may be greater due to detection of heterogeneity than shorter C-band and the longer P-band channels.

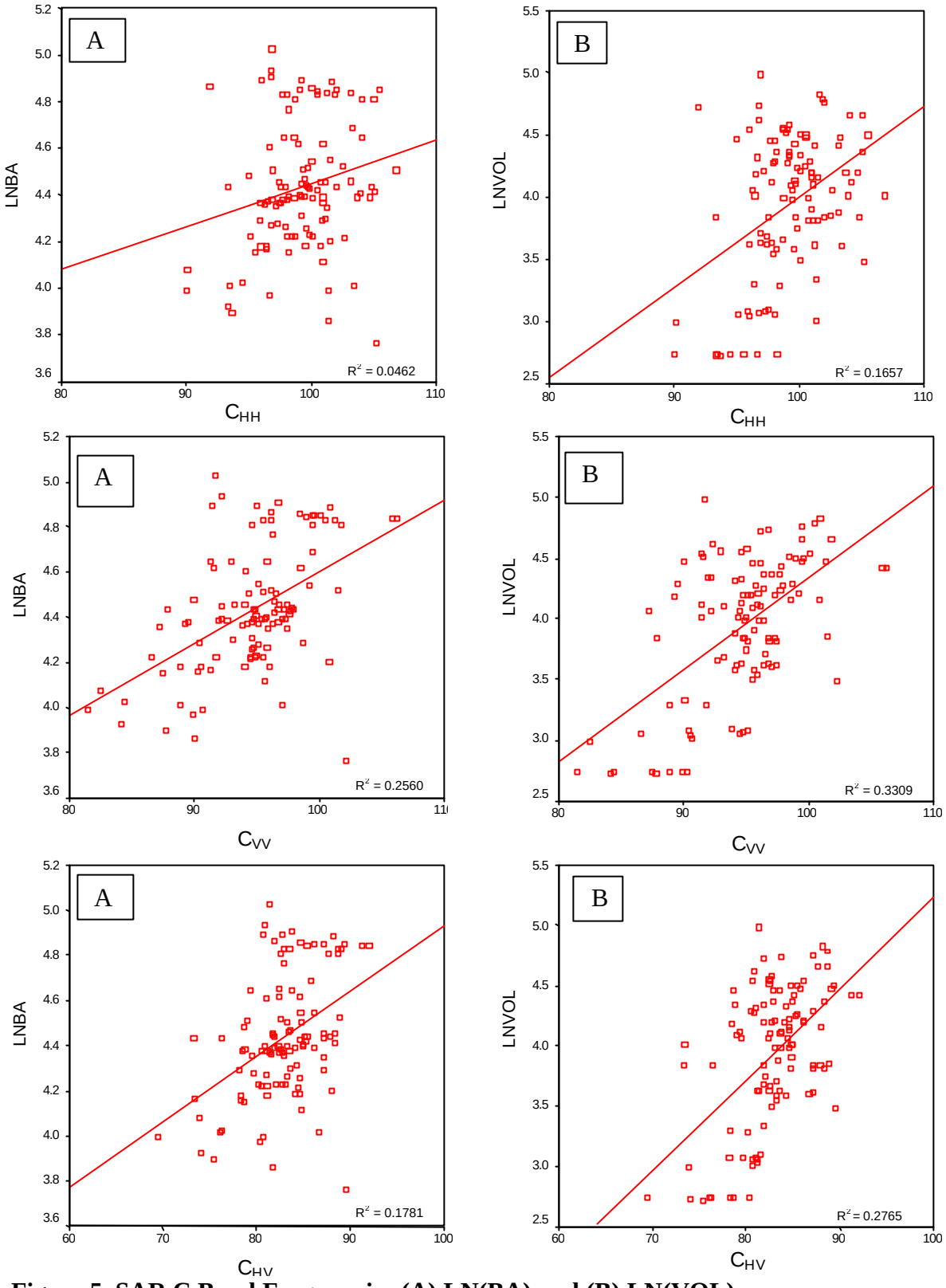


Figure 5. SAR C-Band Frequencies (A) LN(BA) and (B) LN(VOL)

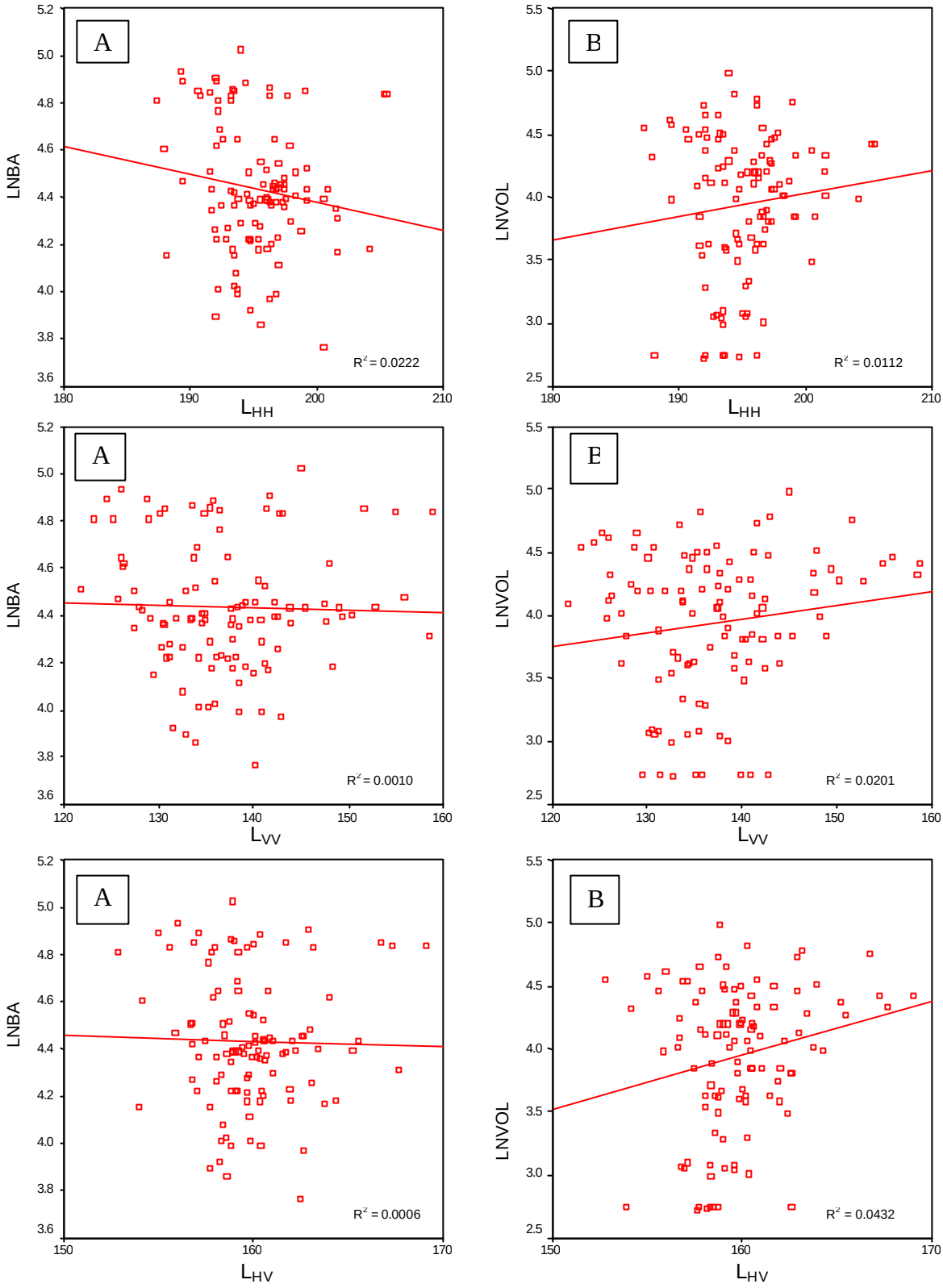


Figure 6. SAR L-Band Frequencies Scatter Plots (A) LN(BA) and (B) LN(VOL)

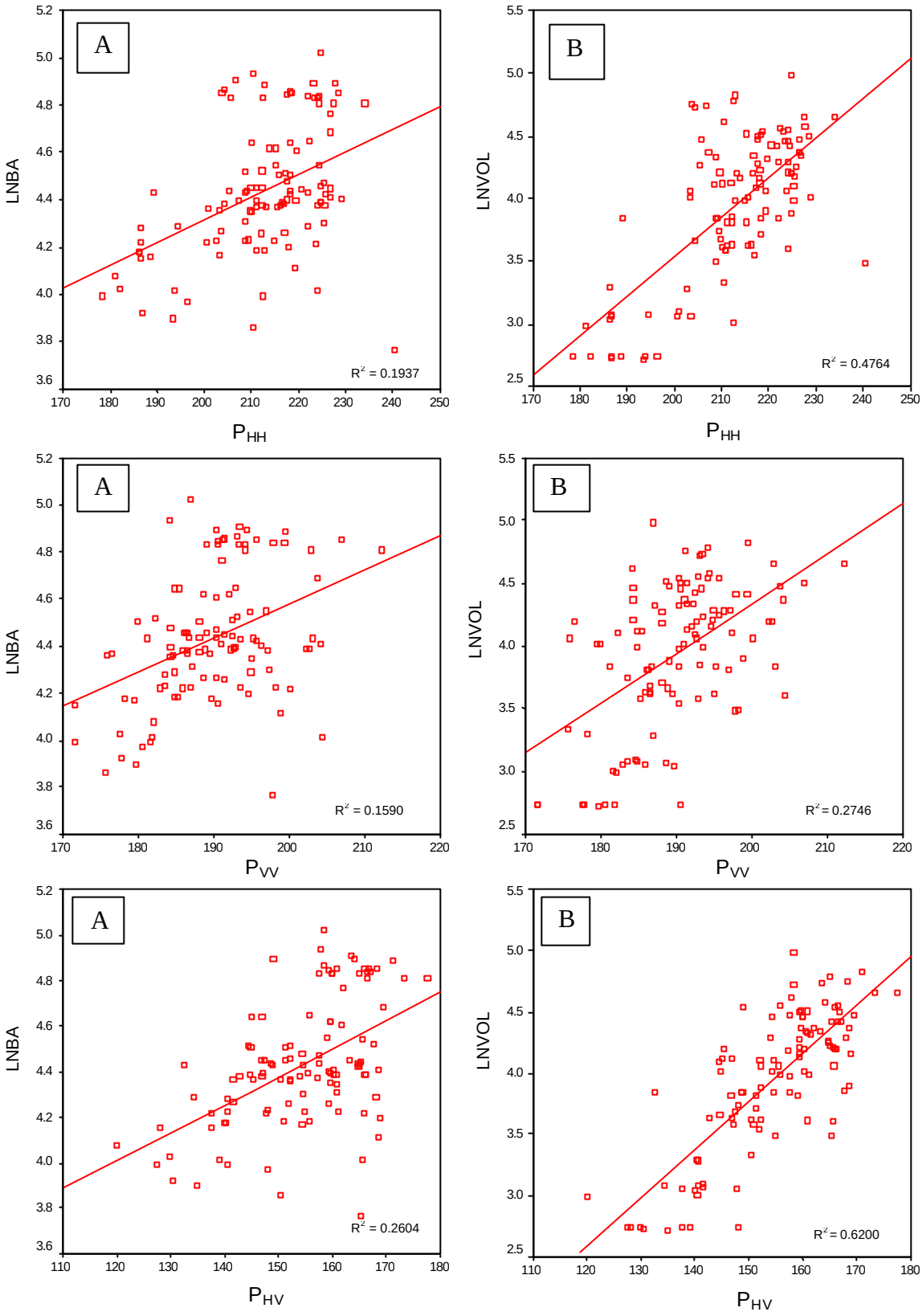


Figure 7. SAR P-Band Frequencies Scatter Plots (A) LN(BA) and (B) LN(VOL)

Examination of SAR bands individually is beneficial for potentially identifying bands that have the strongest simple linear relationships. However, the ability of SAR data is also to not only transmit at various frequencies but is to transmit at various polarizations. Graphically examining how the various polarizations interact with basal area and volume similarly may aid in identifying the most potential predictor variables (Figure 8). In comparing the overlaying scatter plots, it is clear that the each polarization from the C- and P-band have a positive relationship with both basal area and volume. Each discrete polarization also exhibits *"good"* separability from the other unique polarizations for that specific frequency. This indicates that multi-polarized frequency does capture different backscatter intensity. However, each unique L-band polarization has very little relationship with basal area or volume, yet they do have a very distinct separation from other L-band polarizations. The inability to capture an increasing tendency in basal area and volume suggests that C_{HH}, C_{VV}, and C_{VH} do not have a high correlation with the either basal area or volume.

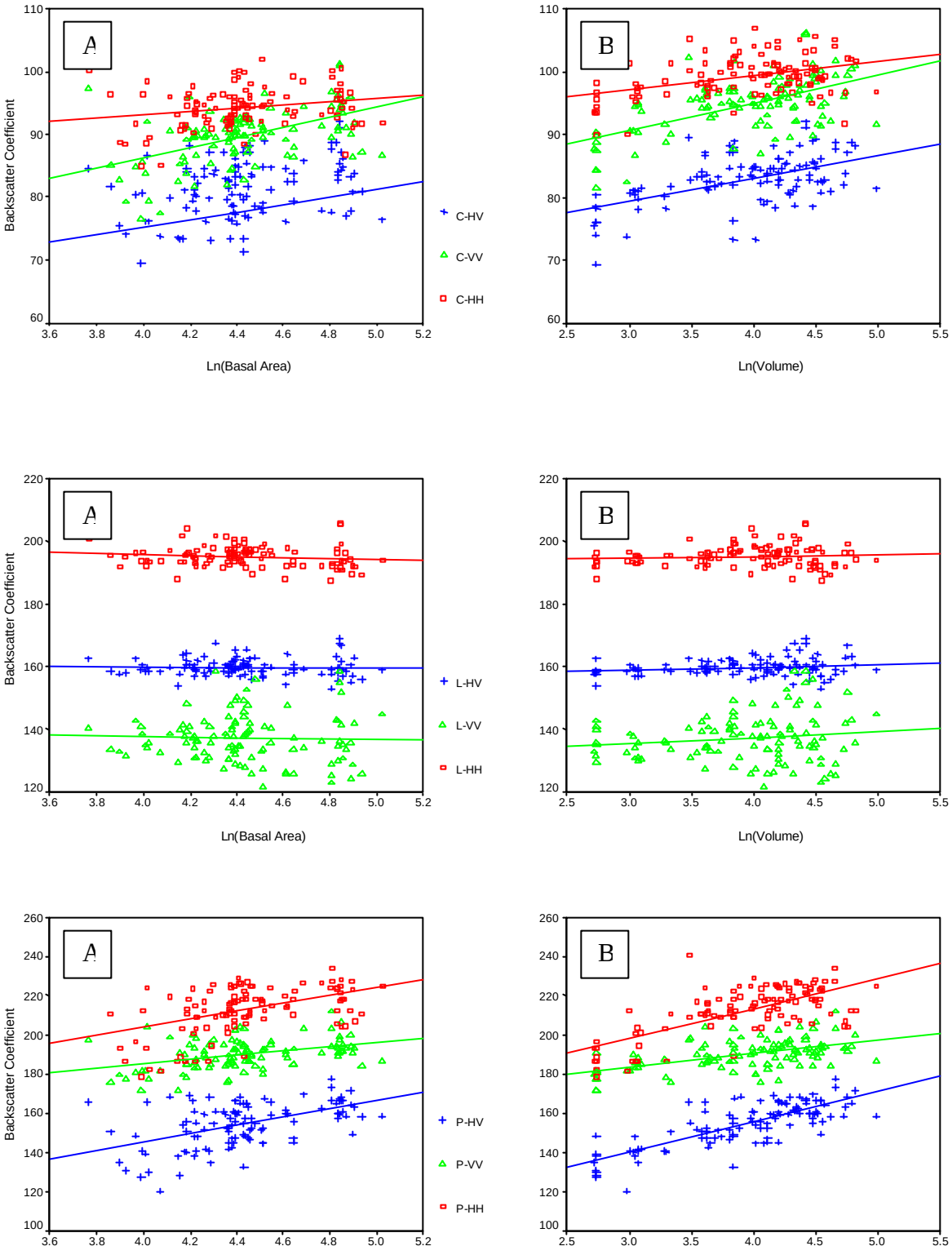


Figure 8 . C-, L-, and P-band polarizations vs. (A) ln(Basal Area) and (B)ln(Volume)

Relationships between LTM bands and slash pine stand parameters were also explored graphically (Figure 9-14). Both LTM -On and –Off exhibit negative linear relationships with basal area and volume. This is generally expected for optical imaging sensors that acquire images by measuring absorption and reflectance from a vegetative landscape. A simple rule-of-thumb is the more vegetative cover, the more solar radiation absorbed and consequently, the less radiation reflected. Coniferous forests exhibit this rule. As the crown closes and becomes more structurally complex, shadowing within the canopy increases, and the transmission of light is reduced because the vertical development of the canopy causes the reflectance to decrease. Conversely, as the crown closes basal area and volume continue to increase even though there are fewer individual stems supporting these basal area and volume increases.

LTM sensors may not capture this continuous increase because it is most sensitive to the degree of crown closure that has already peaked earlier in the stand's development. The scatter plots for the data are consistent with this simple principle. Scatter plots also illustrate that LTM Leaf-On Band -2, -5 and -7 and Leaf-Off Band -3, -5, and -7 show strongest relationships with the natural logarithm of basal area. While, LTM Leaf-On Band -1, -2, -5 and -7 and Leaf-Off Band -3, -5 and -7 best exhibit the strongest linear relationships with the natural logarithm of volume. Overall, Band -2, -3, -5, and -7 present the strongest linear relationships and provide the greatest opportunity as predictive variables.

Collectively, plots exhibit the possibility that certain interactions between the remotely sensed bands may provide a stronger linear relationship and account for more variability within the stands. Interaction terms from both SAR and LTM bands as

significant predictor variables is supported by previous literature. For this research, all possible combinations of band ratios and products seemed appropriate. Although, a host of alternative band transformations are possible, these simple trans formations seemed most warranted and suitable.

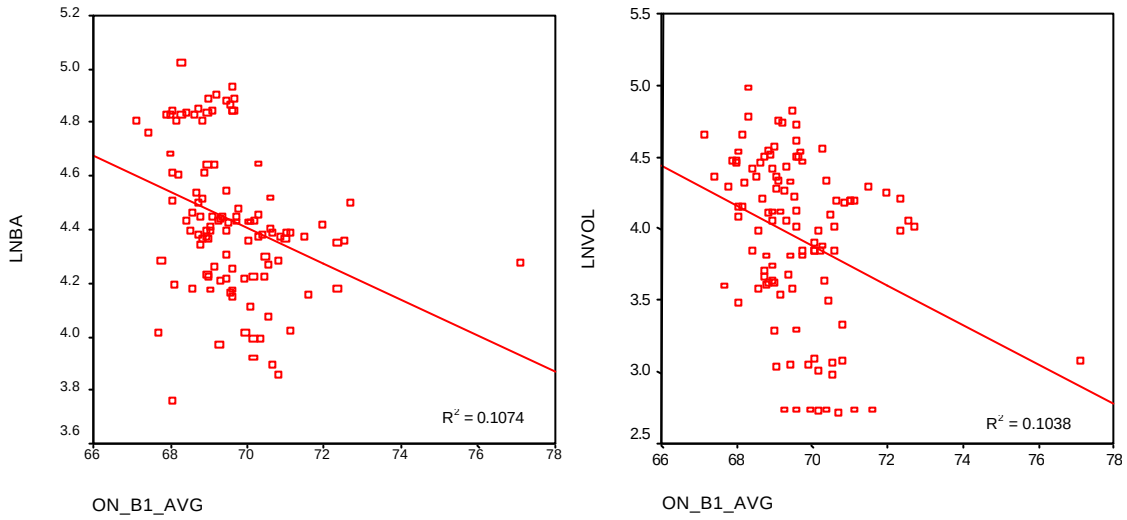


Figure 9.A. Scatter Plots for LTM-ON Band-1 with LN(BA) and LN(VOL)

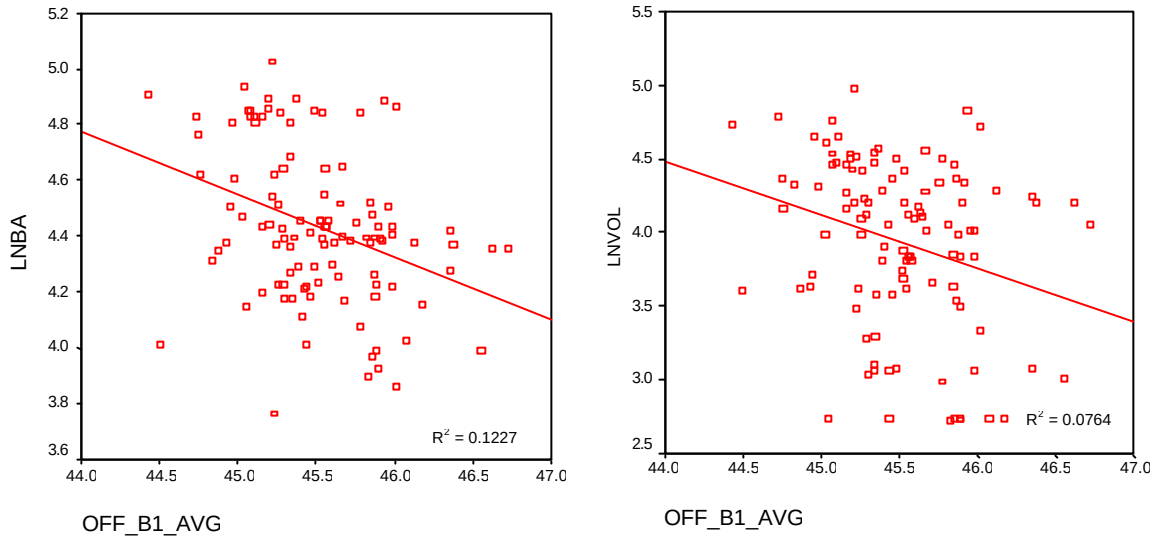


Figure 9.B. Scatter Plots for LTM-OFF Band-1 with LN(BA) and LN(VOL)

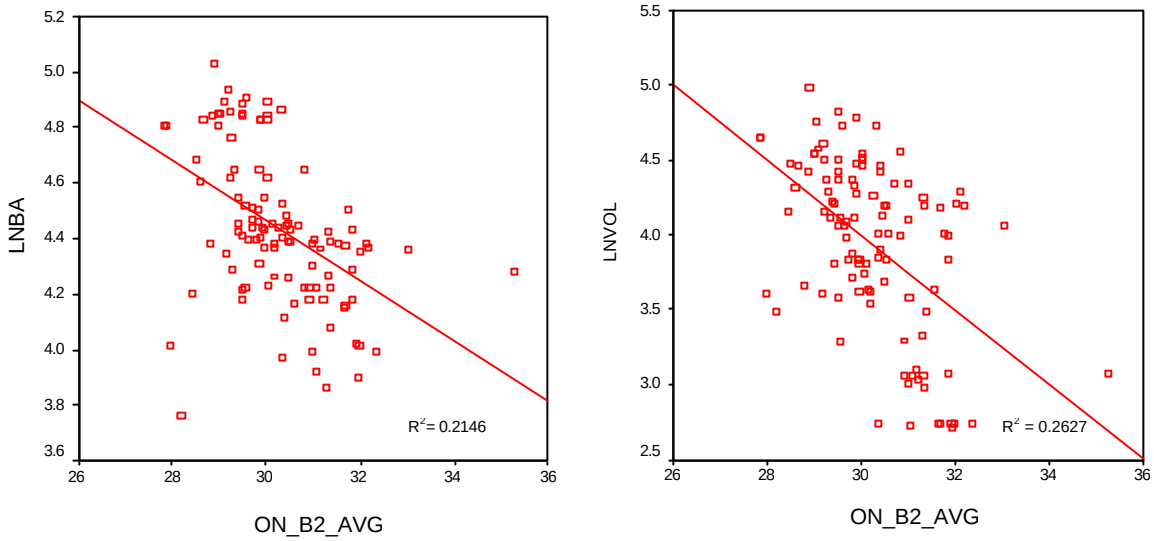


Figure 10.A. Scatter Plots for LTM-ON Band-2 with LN(BA) and LN(VOL)

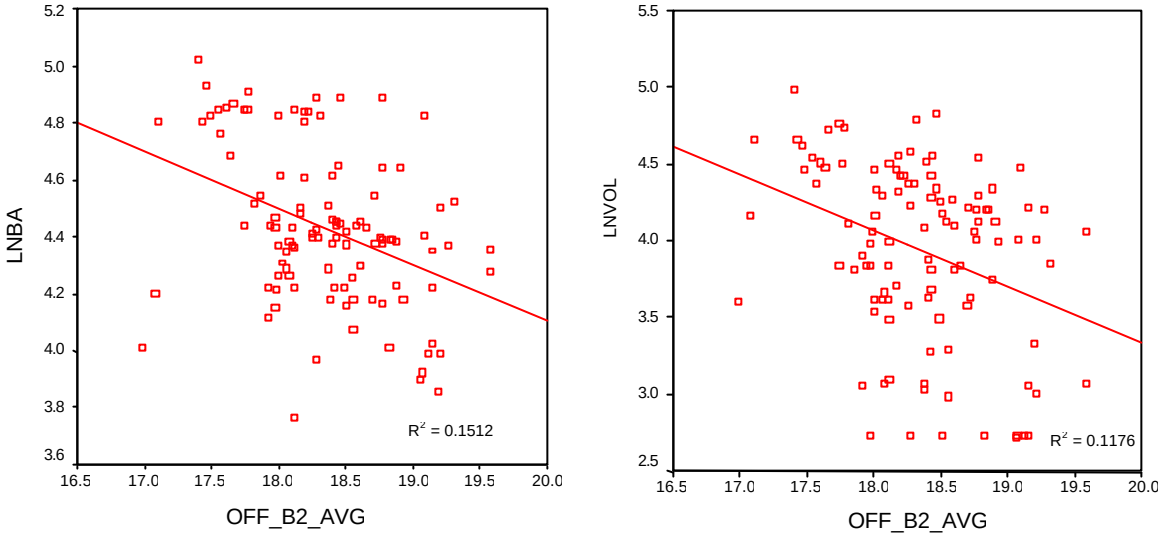


Figure 10.B. Scatter Plots for LTM-OFF Band-2 with LN(BA) and LN(VOL)

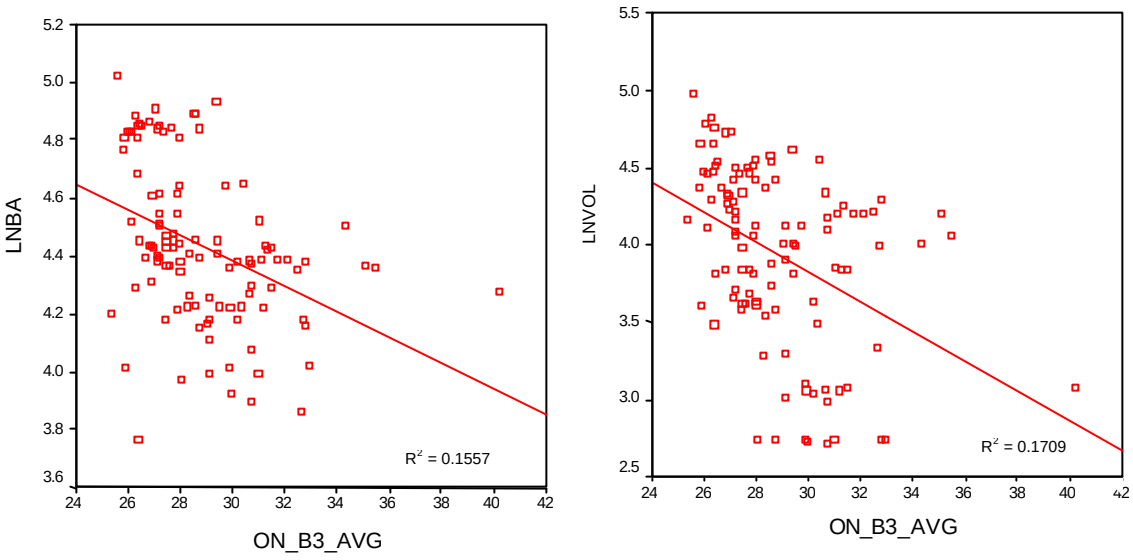


Figure 11.A. Scatter Plots for LTM-ON Band-3 with LN(BA) and LN(VOL)

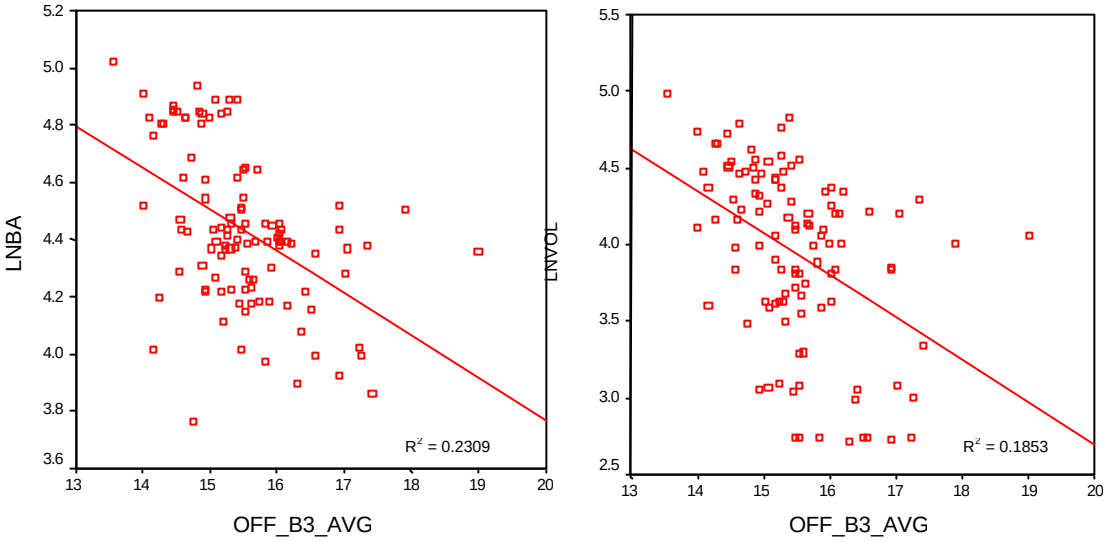


Figure 11.B. Scatter Plots for LTM-OFF Band-3 with LN(BA) and LN(VOL)

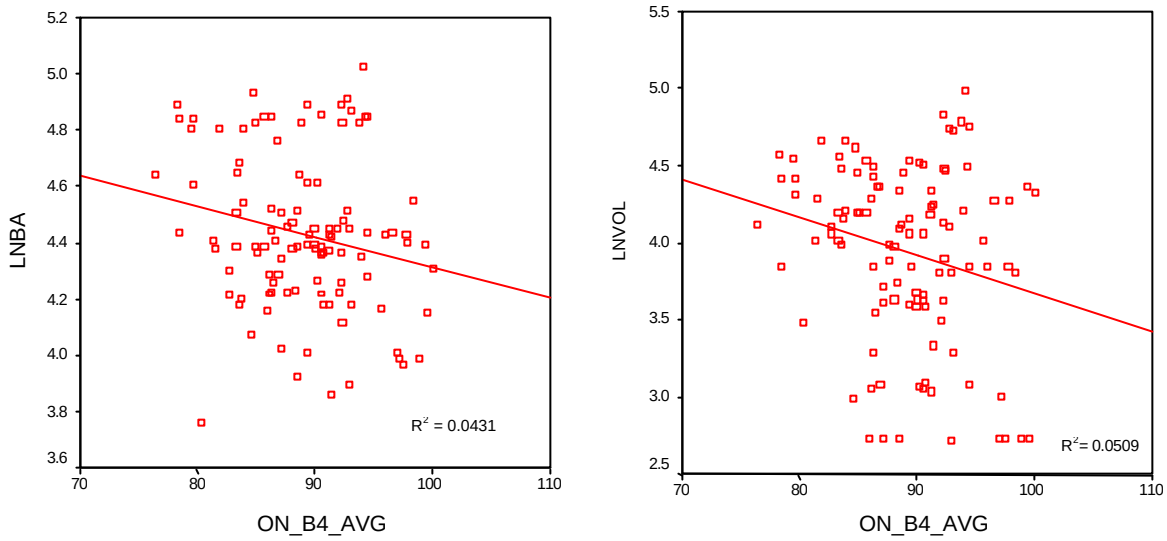


Figure 12.A. Scatter Plots for LTM-ON Band-4 with LN(BA) and LN(VOL)

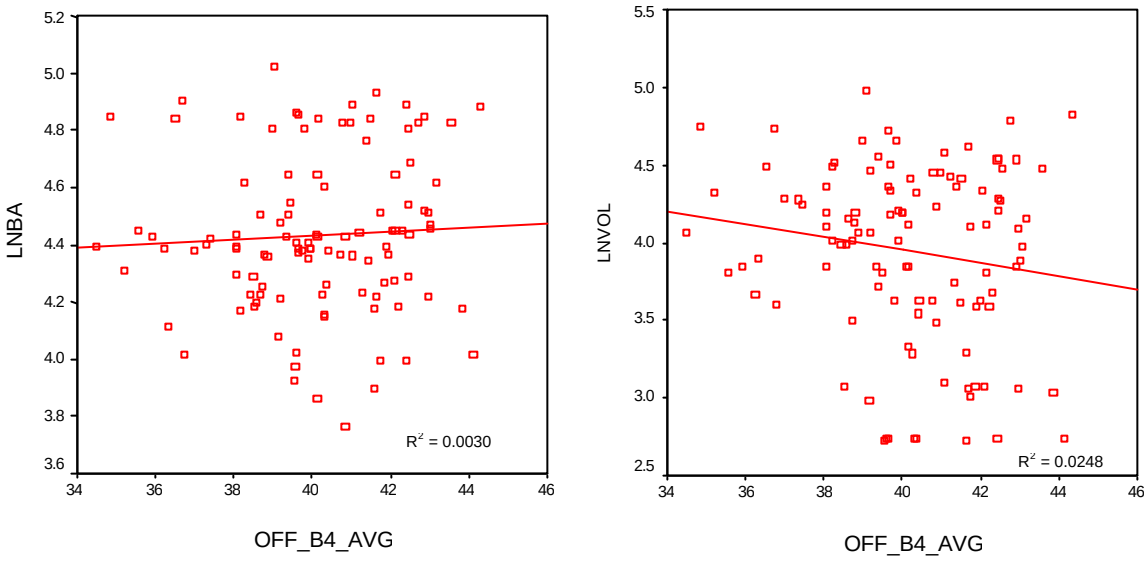


Figure 12.B. Scatter Plots for LTM-OFF Band-4 with LN(BA) and LN(VOL)

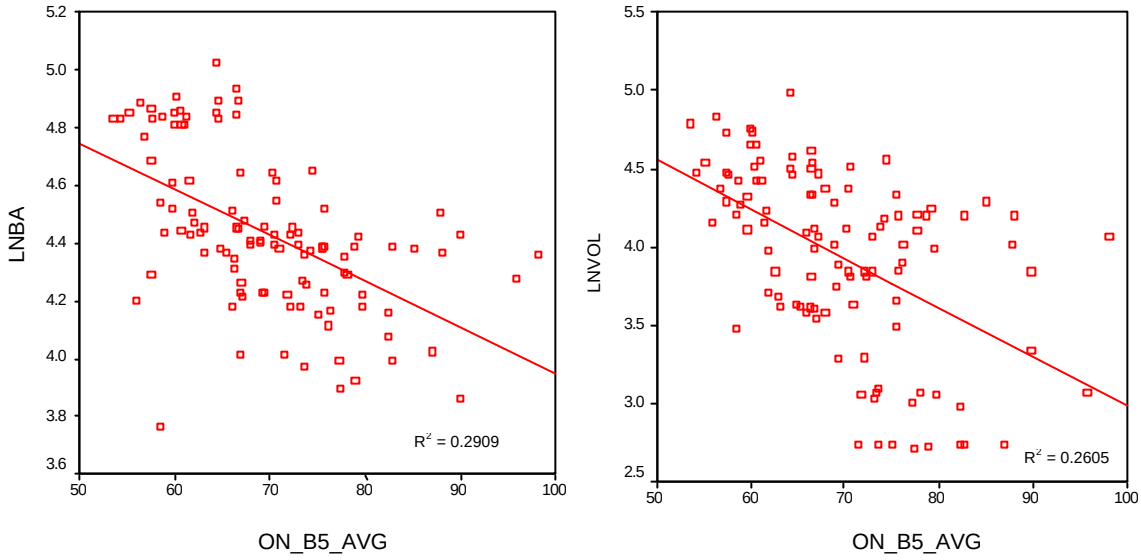


Figure 13.A. Scatter Plots for LTM-ON Band-5 with LN(BA) and LN(VOL)

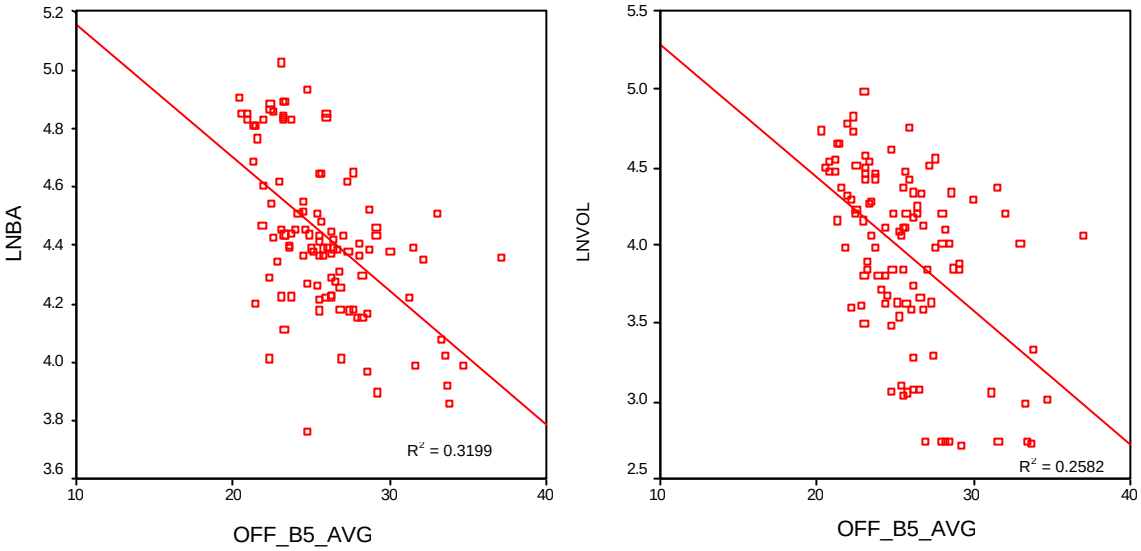


Figure 13.B. Scatter Plots for LTM-OFF Band-5 with LN(BA) and LN(VOL)

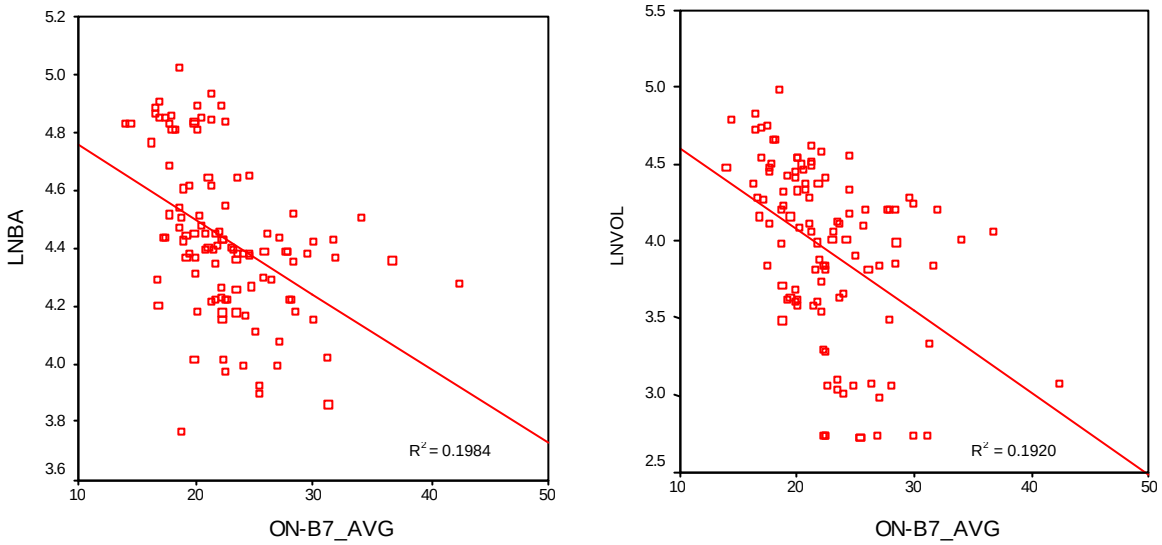


Figure 14.A. Scatter Plots for LTM-ON Band-7 with LN(BA) and LN(VOL)

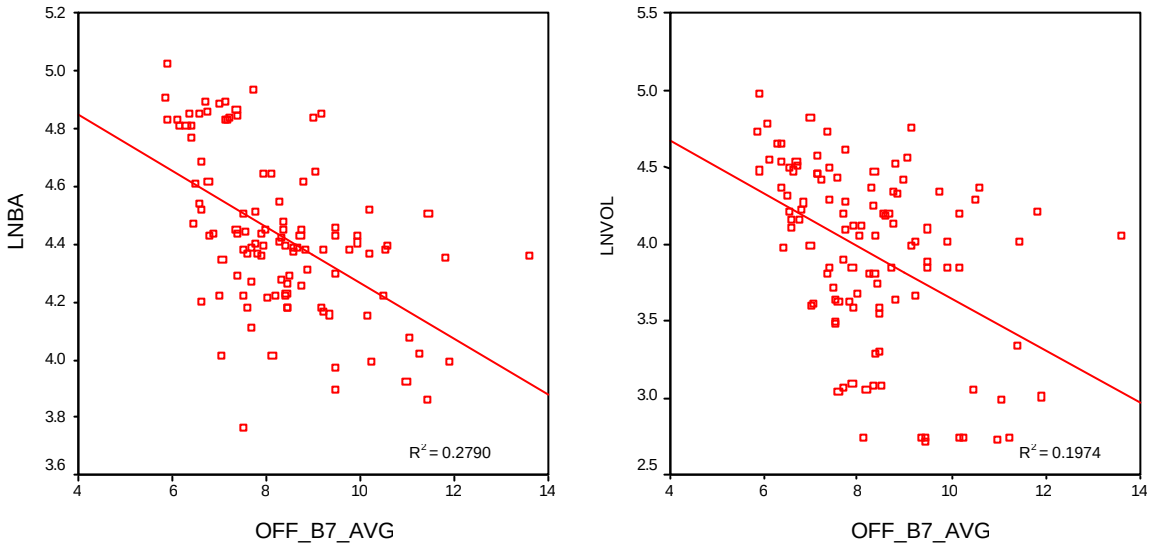


Figure 14.B. Scatter Plots for LTM-OFF Band-7 with LN(BA) and LN(VOL)

Examining SAR and LTM band interactions as possible explanatory variables required generating all possible ratio and product combinations from both SAR and LTM data independently of each other. No SAR and LTM cross-band ratios were examined. Available SAR data included the nine SAR mean band coefficients and their nine variances, seventy-two possible band ratio combinations as well as another seventy-two possible band product combinations. LTM data available were six leaf-on and six leaf-off mean band digital numbers and their according variances along with thirty combinations of band ratios. Band products were not generated for LTM data due to the specific sensitivity of each LTM sensor per specific ground feature.

STEPWISE, MULTIPLE-LINEAR REGRESSION MODEL DEVELOPMENT

Stepwise, multiple-linear regression was applied to derive possible predictive models from SAR and LTM data. Models were developed from selecting SAR and LTM bands based upon their direct statistical relationship with inventory data without regard to biological inference on the model. The stepwise regression selection method systematically reduces the dimension of candidate variables to generate a satisfactory model. With stepwise regression, the order of variable introduction is determined by a candidate variable's partial correlation coefficient with the dependent variable (Draper and Smith 1981). The process begins by selecting the independent variable that is most highly correlated with the model's dependent variable. Upon selection of the first introduced variable, a first-order linear regression is fitted and tested for significance. If the simple linear model is not significant then it is deemed $\hat{Y} = \overline{Y}$. If the model is indeed significant, the process continues and all possible independent variables not in the

equation are examined for their partial correlation coefficient. Examination of partial correlation coefficients from predictor variables not included in the model is equivalent to examining correlations between (1) residuals from the first linear model and (2) residuals from each of the single-termed regression models that are not actually performed. The predictor variable with the next highest partial correlation with the dependent variable is now selected and added to the model. Now, a two-term model is tested for significance, and partial F-values for each variable in the model are analyzed. Comparing the two variables, the lower F-value between the variables is tested for significance. According to the level of significance (?) predetermined by the modeler the corresponding variable is either retained or removed from the model. Testing the "least significant predictor" with the addition of a new variable is useful because a variable selected earlier as a best entry candidate may be insignificant at a later stage due to relationships between that variable and the newly introduced predictor variables. If the tested variable proves to be an insignificant contribution, it is removed from the model and regression model is fitted with only those variables remaining. The selection process then attempts to add an additional predictor variable depending upon its partial correlation with the dependent variable, given previously selected predictors are contained in the model. When there is no cause for present variables in the current model to be removed and the next best predictor variable fails to meet entry significance requirements, the stepwise selection process ceases.

For analyzing SAR data using stepwise, multiple-linear regression, a seven-layer modeling approach was applied, where within each layer a best *fit* model was developed from available predictor variables (Figure 16). Identical modeling designs were applied

for estimation of basal area and volume. The intent of this design was to determine which method of variable introduction for SAR data might be most beneficial for selecting optimal band(s) and/or secondary band interactions to more accurately estimate volume and basal area.

Table 4. Stepwise Regression Technique, SAR modeling design

MODEL	Independent Variables	Where:
1	A	A = BAND AVERAGES and VARIANCES B = BAND RATIOS C = BAND PRODUCTS
2	B	
3	C	
4	A - AND - B	
5	A - AND - C	
6	B - AND - C	
7	A, B, AND C	

A similar layered modeling approach was applied to analyze LTM data. Stepwise, multiple-linear models were developed from a three-layered approach with the band products eliminated. Band products were not suggested in the literature, and LTM's unique sensor characteristics, suggest band products would not be beneficial (Figure 17). Again, the intent of this layered approach strategy was to produce the best fitted models for estimating basal area and volume from various methods of introducing candidate variables.

Table 5. Stepwise Regression Technique, LTM modeling design

MODEL	Independent Variables	Where:
1	A	A = BAND AVERAGES and VARIANCES
2	B	B = BAND RATIOS
3	A - AND - B	

A third strategy for developing fitted models using stepwise, multiple-linear regression was to determine of using both SAR and LTM data yielded better predictive equations than do either SAR or LTM data independently. The stepwise layered model approach was identical to that of the independent modeling with a three layer design (Figure 18). For this layered modeling design SAR band ratios and products were combined to reduce the dimension of the design.

Table 6. Stepwise Regression, Synergy of SAR and LTM modeling design

MODEL	Independent Variables	Where:
1	A	A = BAND AVERAGES and VARIANCES
2	B	B = BAND INTERACTIONS
3	A - AND - B	

In all instances of model development, band variances were of particular interest. It was hypothesized that variance terms may be beneficial predictor variables. This hypothesis is supported by the fact that there is a certain amount of unexplainable error incorporated with calculating a stand's mean band coefficient. Variance terms incorporate this variability. Introducing a measure of variability as an independent variable in the model may help *adjust* model fit.

MODEL SELECTION, STEPWISE MULTIPLE-LINEAR REGRESSION

Selection of the "best *regression equation*" is an intricate process. Several decision criteria have been suggested for model validation. Standard procedure uses (1) number of variables in the model, (2) Mean Squared Error (MSE), (3) R^2 and adjusted-R^2 terms, and (4) residual plot analysis. Each selection criterion has standing for model selection, and should briefly be discussed.

Degrees of freedom (d.f.) refers to the number of independent predictor variables associated with the regression equation, and denotes number of parameters selected as providing significant influence on predictability of the dependent variable. Assessment of Sum of Squares terms for a fitted regression model and its errors are partially determined by the *d.f.* term. Usefulness of degrees of freedom term as a selective tool is supported by a primary dilemma for linear regression. Fundamental theory behind regression analysis regarding predictor variables is that there are two conflicting selection criteria for independent variables. First, make predictive models as fully capable as possible. The number of predictor variables should be as great as possible where-in each is still significant and contributes to the predictive power of the model. In contrast to this, the number of variables should be as minimal as possible to avoid multicolinearity among the predictor variables (Neter *et al.* 1996).

A more valuable tool for determining predictive strength and stability of a model is to statistically and graphically examine the residuals. Residuals are defined as the difference between the observed (Y_i) and the predicted $(\hat{Y}_i)$ values:

$$\textbf{Residual} = (Y_i - \hat{Y}_i)$$

Residuals measure the difference between measured data and model prediction estimates. While the purpose of a regression model is to explain variability in the response variable, examining residuals provides information about the variability not explained by the regression model.

To examine residuals at a model level the summation of the squared residuals is calculated to produce residual mean square, or more commonly Mean Square Error (MSE):

$$\textbf{MSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i) \Big/ (n - p)$$

Where: n = number of sample units

p = number of parameters in the model

In evaluating a particular model, or several models as in this case, the lower the MSE, the better the model fits the data. MSE is influenced by the addition of predictor variables. The fact it is possible for MSE values to increase if improvement in residuals does not offset the additional loss of d.f.

Graphical representation of residuals is also a highly critical tool for determining the strength and form of a regression model. Graphical methods have an advantage over quantitative methods for model validation since they readily illustrate a broad range of complex aspects of the relationship between model and data. Numerical methods for model validation tend to be narrowly focused on a particular aspect of the relationship between model and data and often try to compress that information into a single descriptive number or test result (Neter *et al.* 1996).

There are certain fundamental linear regression assumptions that residuals must satisfy, and are they generally examined through graphical residual analysis. Linear regression analysis requires error terms (e_i) to comply with two assumptions: (1) errors (e_i) are assumed to have constant variance and (2) errors (e_i) are uncorrelated, independent, and identically distributed from a normal distribution. Thus any departures from the underlying assumptions of errors will be apparent in the residuals. Graphical analysis of residuals can be a powerful method to determine and/or identify any model deficiencies. Assuming the model fit from the data is correct, residuals approximate random errors that make the relationship between explanatory variables and the response variable a statistical relationship. Therefore, if residuals appear to behave randomly, it suggests that the model fit the data well. On the other hand, if non-random structure is evident in the residuals with respect to dependent and independent variables, it is a clear sign that there may be variability that is unexplained or could possibly be better explained in an alternative model form.

Coefficient of determination (R^2) examines the proportion of the total variation about $\bar{Y}$ explained by the predictor variables, and is another crucial model validation criterion (Draper and Smith 1981):

$$\mathbf{R^2} = 1 - \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \Big/ \sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2$$

Higher measures of R^2 imply that a regression model explains a higher percent of variability. The R^2 statistic should be approached with caution due to its underlying properties. Recall, the two fundamental principles of model development were to have a model with maximum number of independently contributing variables in the regression model and to have the least number of variables to avoid multicolinearity between

regressors. The R^2 statistic is susceptible to the effect of adding additional regressors and possible multicolinearity between explanatory variables. To account for inflation of the R^2 statistic due to addition of explanatory variables, a more suitable metric to determine what percent of variability the model explains may be an adjusted-R^2 statistic:

$$\mathbf{R_a^2} = 1 - (1 - R^2)\left(\frac{n-1}{n-p}\right)$$

This statistic has been *adjusted* for the model's increased amount of variability explained due to additional explanatory variables added to the regression model. Although there are discrepancies between use of either statistic, both R^2 and adjusted-R^2 will be compared for this analysis, given two diagnostic assumptions: 1) both terms present an overall value for the percent of variation accounted by the regression model, and 2) the difference between R^2 and adjusted-R^2 terms may help determine if the appropriate number of predictor variables have been selected. If the two statistics differ drastically, there is a probability that the regression model is overspecified and all included variables may not contribute meaningfully to the fitted model (Neter *et al.* 1996).

Tables 7-9 present model fit diagnostics using stepwise multiple-linear regression with SAR data. It is important to recognize that independent SAR band averages and variances produced better fitted models then band interaction terms. However, in all instances, combination of averages and variances with interaction terms increased the strength of the models' fit statistics. Very similar models for estimating volume were developed using SAR data among layers 4 – 7 of the design. Selection of Model 6 as the best fitted model is due to its relative high R^2= .07221 but higher adjusted-R^2=0.7114,

and more importantly, Model 6 is only a four parameter model compared to others which are six parameter models.

Table 7.A. Stepwise Regression, SAR Model Estimation of LN (*Volume*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	5	48.26	0.1033	0.3214	0.7008	0.6863
2	4	50.93	0.1156	0.3400	0.6621	0.6491
3	2	97.19	0.1184	0.3441	0.6471	0.6404
4	6	44.29	0.0967	0.3110	0.7226	0.7063
5	6	45.25	0.0952	0.3086	0.7269	0.7109
6	4	67.56	0.0950	0.3083	0.7221	0.7114
7	6	44.29	0.0967	0.3110	0.7226	0.7063

Nearly identical fit statistics are generated between the models, but due to the selection criteria, selection of the simpler model is warranted. Model 6 estimates of volume as:

$$\mathbf{LN(Vol) = b_0 + b_1(}C_{HH}/P_{HV}\mathbf{) + b_2(}P_{HV}/C_{HH}\mathbf{) + b_3(}P_{HH}\,C_{HH}\mathbf{) + b_4(}P_{HH}\,P_{HV}\mathbf{)} \qquad \mathbf{(4.1)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-91.6198	18.89889	<.0001
b_1	50.8993	12.19184	<.0001
b_2	39.0125	7.47290	<.0001
b_3	0.00113	0.00028	<.0001
b_4	-0.0007	0.00018	0.0005

SAR band selection for this model is consistent with past research that has found the product of C- and P-band, as well as ratio of P- and C-bands to be most highly correlated with stand volume (Dobson *et al.* 1995, Kasischke *et al.* 1995, Le Toan *et al.* 1992, Vona 2000). The model consisted of SAR band ratios and product interaction terms. It is interesting that combinations of ratios and products developed the most successful volume prediction model. Ratio and product terms are generally applied to capture

curvilinear relationships. Recall, scatter plots from individual SAR bands against volume (Fig 5-7) show no noticeable curvilinear relationship. However, no strong linear relationship were apparent either, leading to the assumption that there was likely some underlying curvilinear effect that this model best captured. The possibility of ratio or SAR band interaction capturing a curvilinear relationship can be addressed by SAR susceptible to a saturation point, where increased levels of volume are more difficult to differentiate due to SAR's inability to proportionately capture these increased levels of volume (Dobson *et al.* 1995, Vona 2000).

Table 7.B. Stepwise Regression, SAR Model Estimation of LN (*Basal Area*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	6	16.77	0.0402	0.2005	0.4965	0.4669
2	2	33.82	0.0469	0.2165	0.3896	0.3780
3	2	31.94	0.0479	0.2189	0.3760	0.3643
4	6	19.50	0.0372	0.1928	0.5342	0.5068
5	5	19.40	0.0407	0.2018	0.4851	0.4601
6	2	33.82	0.0469	0.2165	0.3896	0.3780
7	5	20.60	0.0395	0.1988	0.4999	0.4757

For estimating basal area, Model 4 that contains the combination of SAR band averages and variances in addition to band ratios yielded the best fitted model. Model 4 estimated basal area as the following:

$$LN(BA) = b_0 + b_1(\mathbf{C_{HH}}) + b_2(\mathbf{C_{HH}\text{-}var}) + b_3(\mathbf{L_{VV}\text{-}var}) + b_4(\mathbf{P_{HV}}) + b_5(\mathbf{C_{VV}/L_{VV}}) + b_6(\mathbf{P_{HV}/C_{HH}}) \quad \textbf{(4.2)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-20.64550	7.71721	0.0087
b_1	0.20480	0.25282	0.0105

b_2	0.00374	0.35077	0.0027
b_3	-0.00138	0.72539	<.0001
b_4	-0.14804	0.31023	0.0047
b_5	8.20616	1.20608	<.0001
b_6	15.22704	0.33430	0.0034

Again, C- and P-bands heavily influence model estimation, supporting the premise that these specific frequencies interact and are sensitive to attributes that help explain structural relationships. Selection of the variance terms as significant predictor variables supports the objective to determine if variances from aggregated pixels' band coefficients may lend significant explanatory contribution to the regression model. Selection of the variances from the C_{HH} and L_{VV} bands suggests that the shorter and mid-wavelengths may be useful in detecting variability in crown and understory layers that would help better predict basal area.

Unlike SAR data, use of LTM data provided considerable differences in ability to estimate volume among model development design layers. Examining Tables 8 A&B, shows these ranges of predictability are quite apparent.

Table 8.A. Stepwise Regression, LTM Model Estimation of LN (*Volume*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	5	22.61	0.1646	0.4057	0.5233	0.5002
2	4	27.17	0.1672	0.4089	0.5110	0.4922
3	6	28.71	0.1297	0.3601	0.6281	0.6062

Model 3 produced the best fitted model for estimating volume using LTM data, and consisted of combining band averages and variances with band ratios. Combination of both LTM bands and their ratio interactions considerably increased the fit of the model

for estimating volume. Comparing fit statistics, Model 3 produced highest R^2=0.6281, adjusted-R^2=0.6062, and lowest MSE=0.1297 despite having the largest number of parameters model has a relatively small difference between the R^2 and adj-R^2 leading to the conclusion R^2 is not overly inflated due to additional variables. Model 3 estimates volume as the following:

$$\mathbf{LN(VOL) = b_0 + b_1(on\text{-}B4\text{-}var) + b_2(on\text{-}B5\text{-}var) + b_3(off\text{-}B5\text{-}var) + b_4(on\text{-}B2/on\text{-}B1) + b_5(on\text{-}B3/on\text{-}B7) + b_6(off\text{-}B3/off\text{-}B4)} \quad \textbf{(4.3)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	6.71224	1.47880	0.001
b_1	-0.00468	1.75058	0.0004
b_2	0.00151	1.79276	0.0003
b_3	-0.01371	2.70907	<.0001
b_4	-19.71396	3.86877	<.0001
b_5	2.27694	4.85013	<.0001
b_6	7.89817	3.85030	<.0001

Again it is apparent that variances from the aggregated band coefficients have significant predictive influence on the model. The major advantage of LTM ratios is they effectively compensate for scene brightness variation caused by sources of forest condition variability. Since LTM data is based upon absorption and reflectance of radiation, a ratio value may best define how similar conditions with a managed forest affect the spectral responses from the LTM sensors. Defining these relationships in terms of reflected radiation ratios may facilitate the interpretation of subtle reflectance differences.

Table 8.B. Stepwise Regression, LTM Model Estimation of LN (*Basal Area*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	4	17.83	0.0464	0.2155	0.4068	0.3840
2	3	25.58	0.0448	0.2117	0.4222	0.4057
3	3	25.58	0.0448	0.2117	0.4222	0.4057

Model 2 and Model 3 for estimating basal area selected identical predictor variables. This is a significant finding that suggests LTM band ratio terms are more successful for predicting basal area than are band averages and variances. The fitted regression equation for estimating basal area from Model 2 and Model 3.

$$\mathbf{LN(BA) = b_0 + b_1(on\text{-}B3/on\text{-}B5) + b_2(off\text{-}B1/off\text{-}B5) + b_3(off\text{-}B5/off\text{-}B4)} \quad \textbf{(4.4)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-0.04435	1.03192	0.9658
b_1	4.05054	1.13010	0.0005
b_2	1.00539	0.24490	<.0001
b_3	1.52790	0.61748	0.0149

Model 2 and Model 3 produce a slightly higher R^2=0.4222, adjusted-R^2=0.4057 with a slightly lower MSE=0.2117 and less parameters than Model 1. Residual plots exhibit no violations of regression assumptions.

Modeling both SAR and LTM data independently allows comparison between the two types of remotely sensed data for attribute estimation. It is evident that for both volume and basal area, SAR data explains more variability in stand volume and basal area. This ability is consistent with previous findings which have indicated SAR may be more successful than LTM data due to its ability to penetrate the crown layer and interact directly with the stand's stems and understory layers. Also, in both instances variances from aggregated band coefficients played an integral role in fitting estimation models.

This is a new finding for model development of biophysical attribute estimation. Although variance may be foreseen as a step-away from the original band's spectral response and as a result fails to accurately represent spectral response. However, it can be shown that calculating the variance term does mathematically account for the original pixels' spatial resolution.

A third approach using stepwise multiple-linear regression incorporated combining both SAR and LTM data as available predictor variables. Earlier findings from developing predictive models applying SAR and LTM data independently have found that both have potential for accurate attribute estimations, and suggest using both SAR and LTM data may increase overall estimation fit. This approach is also supported by literature which have found that the each of the remotely sensed data provides unique spectral interpretations and hence modeling capability (Kuplich *et al.* 2000, Leckie 1990, Pohl and Van Genderen 1998). Combining these two data types is also a new approach for biophysical attribute estimation. Previous attempts to use both SAR and LTM data have been for classification techniques, and have required resampling image pixels to a uniform spatial resolution. The need to obtain consistent spatial resolution prior to classification is required since classification algorithms are based upon interpreting spectral response pattern recognition (Franklin 2001). However, resampling prior to SAR and LTM data calibration is not warranted for this application because the differing spatial resolution from SAR and LTM data are as useful for modeling purposes as are their unique, spectral interpretation. Differing pixel spatial resolution was accounted for by internally resampling underlying stand pixels to a 0.5 meter resolution when aggregating pixels. This approach allowed original spatial resolution from SAR and LTM to be maintained, while not violating spatial resolution assumptions.

Results from combining SAR and LTM data are presented in Tables 9 A&B. In examining Table 9.A, it is immediately evident that the combination of SAR and LTM data explains more variability in volume.

Table 9.A. Stepwise Regression, Combined SAR and LTM Model Estimation of LN (*Volume*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	11	32.96	0.0774	0.2782	0.7889	0.7650
2	4	67.56	0.0950	0.3083	0.7221	0.7114
3	8	48.70	0.0726	0.2695	0.7957	0.7794

Model 1, consisting of SAR and LTM band averages and their corresponding variances, provides a better estimation of stand volume than does any independent SAR or LTM model. The fitted model consists of a full range of variables including SAR band averages and variances with LTM band averages and variances. However, this model is not the best model based on the combined data. Model 3, which consists of all possible SAR and LTM band averages and variances and their band interaction terms, provides the greatest fitted model with an R^2=0.7957 and an adjusted-R^2=0.7794 with an MSE=0.0726. Residual plots show no violation of regression assumptions and support selection of this model as the best fitted model for estimating of volume. Model 3 fits the estimation of volume:

$$\mathbf{LN(Vol) = b_0 + b_1(C_{HV}\text{-var}) + b_2(L_{VV}\text{-var}) + b_3(L_{VH}\text{-var}) + b_4(P_{HV}) + b_5(P_{HH}/C_{HH}) + b_6(\text{on-B3/on-B7}) + b_7(\text{on-B7/on-B2}) + b_8(\text{off-B1/off-B4})} \quad \textbf{(4.5)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-11.37544	1.61704	<.0001
b_1	0.00199	0.00098	0.0457
b_2	-0.00312	0.00078	0.0002
b_3	0.00968	0.00423	0.0243
b_4	0.02279	0.00324	<.0001
b_5	1.79389	0.33461	<.0001
b_6	2.76590	0.54644	<.0001
b_7	2.24641	0.66426	0.001
b_8	2.40768	0.51066	<.0001

This fitted model differs from either model developed from SAR and LTM independently. Estimating volume using SAR data, it became apparent the P-band weighed heavily on the model with interaction terms with the C-band. Such interactions are selected here again, but different, SAR band variances are selected as significant predictors. The C- and L-bands might be measuring variability within the stand that was not previously accounted for by the SAR model itself. Selection of the LTM ratios is consistent with the findings of modeling LTM data separately.

Table 9.B. Stepwise Regression, Combined SAR and LTM Model Estimation of LN(*Basal Area*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	4	21.98	0.0424	0.2060	0.4581	0.4373
2	4	20.85	0.0434	0.2084	0.4450	0.4237
3	4	26.84	0.0385	0.1963	0.5079	0.4890

Combining SAR and LTM data for estimating basal area was not as successful as it was for predicting stand volume. Model 3 was the best model from this approach, and included all possible SAR and LTM band mean and interaction variables. Basal area was estimated as the following:

$$\mathbf{LN(BA) = b_0 + b_1(C_{VV}\text{-var}) + b_2(L_{VV}\text{-var}) + b_3(L_{HH}/C_{VV}) + b_4(\text{on-B1/on-B5})} \quad \textbf{(4.6)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	6.33662	0.76378	<.0001
b_1	0.00629	0.00130	<.0001
b_2	-0.00150	0.00034	<.0001
b_3	-1.34182	0.28794	<.0001
b_4	0.86595	0.22740	0.0002

Although Model 3 provided the most suitable fitted equation for basal area for combining SAR and LTM data, it was not the overall best fitted model for estimating basal area. The model that best fit the data for estimating basal area in terms of R^2, adjusted-R^2, and MSE was generated from (Equation 4.2) with use of SAR data independently. However, these developed models are very comparable with their prediction and the resulting fit statistics. The model from combining SAR and LTM data could be justified as a better fitted model since it has fewer predictor variables than does the model using SAR independently, despite having lower fit statistics. Regardless which model is selected as the better model (Equation 4.2) or (Equation 4.6), it is imperative to note that applying SAR data independently does as good, if not better, for predicting basal area as does it in conjunction with LTM data. This finding suggests that although for estimating volume combining SAR and LTM data is a beneficial approach with the stepwise regression method, for estimating stand basal area it may or not be influential in improving model predictive capability. Likely, the unimproved model fitness is due to both SAR and LTM's limited ability to capture basal area structural components independently. Combining the two data sources may introduce more *"noise"* than useable information.

Use of stepwise, multiple-linear regression has several drawbacks and concerns that need to be addressed since they may be salient for this research. Greatest of these concerns is multicolinearity between predictor variables selected by stepwise selection method. Model development could possibly produce a fitted model that uses all three polarizations of a specific SAR band, however, it is likely that these bands would contain high levels of multicolinearity that might not be addressed in the stepwise regression method. Secondly, the statistical process and individual estimate hypothesis testing is

burdensome, and possibly damages the end-model. Finally, and most troublesome, models developed from stepwise regression do not represent any inherent biological meaning to explain the relationship between remotely sensed data and stand parameters. As a result, model explanation becomes a reactionary process rather than a model building approach. Extrapolation of this model to an independent dataset cannot be warranted since statistical relationships were developed strictly for the present data. Limitations of stepwise multiple linear regression make use of this selection process a questionable technique for model selection, and should be recognized by the modeler.

BIOLOGICALLY MOTIVATED MODEL DEVELOPMENT

Stepwise multiple-linear regression was used to evaluate selection of SAR and LTM candidate variables that statistically had the greatest predictive potential. Fitted estimation models were developed from statistical relationships exhibited between SAR and LTM data with forest characteristics, but developed models failed to attach any biological significance of SAR and LTM bands entered into these models. Stepwise regression examines the correlation between candidate variables with respect to the response variable to determine selection and order of introduced independent variables. A more appropriate model fitting technique to overcome this limitation of stepwise regression is to develop prediction models from SAR and LTM channels that had highest correlations with stand level characteristics. Previously, developed whole-stand models found basal area and volume some linear function of stand characteristics such as age, height, and stand density. Selecting SAR and LTM bands that exhibit the highest correlation with those measured stand characteristics attempts to bring biological

significance to the model development process, rather than developed solely from statistical procedure. The biologically motivated model determines selection and order of entrance for candidate variables with respect to their correlation with stand level characteristics.

Stand growth and yield models have been extensively researched. Predictor variables in these models have precise inference for modeling a stand's biological development. Selecting SAR and LTM bands that have highest correlation and explain the greatest variability between growth and yield model variables may better provide accurate estimates and allow developed models to be more appropriately extrapolated outside present data compared to models developed from the stepwise procedure. In addition to models becoming more biologically significant, examining correlations between individual SAR and LTM bands and stand level characteristics may help further determine how different frequencies and sensors interact with and are sensitive to specific forest attributes. Defining these relationships may bring better understanding for how to apply these relationships to predict forest attributes. While it is strongly suggested that empirical models extracted from such an image calibration will not remain constant as sites and conditions change, it is hypothesized that the biophysical relationships defined should rarely differ from locale (Franklin 2001). Relationships defined to estimate forest attributes by use of statistical selection process will likely deviate more drastically from image to image.

An inherent assumption for developing biologically motivated predictive models is that significant correlation exists between stand–level characteristics and the measured response variables (basal area and volume). To explain the dependence of spectral

response with basal area and volume also requires an understanding how these dependent variables are functions of their structural components. To examine biological relationships between stand characteristics, Pearson correlation coefficients were determined for stand characteristics with basal area and volume (Tables 10.A & B).

Correlation coefficients explain both strength and direction of linear relationships between variables. Relationships are presented between values of –1 and +1, and interpreted as the closer coefficient is to | 1 | the stronger the linear relationship. A negative correlation coefficient implies variables have an inverse relationship, suggesting that a high value of one variable is associated with a low value of the other. While a positive correlation indicates variables have either a direct positive or negative relationship, where a high value of one variable is associated with a high value of the other, or vice-versa with low values.

Table 10.A. Correlation Coefficients for Ln(*Basal Area*) and Stand Characteristics

	Age^{-1}	Ln(HD)	Ln(TPA)
Ln(BA)	-0.4083	0.4097	0.1640

Table 10.B. Correlation Coefficients for Ln(*Volume*) and Stand Characteristics

	Age^{-1}	Ln(HD)	Ln(TPA)	Ln(BA) Age^{-1}	Ln(HD) Age^{-1}	Ln(TPA) Age^{-1}
Ln(Vol)	-0.8399	0.84413	0.4307	-0.7371	-0.8147	-0.8022

As expected, consistently high positive relationships exist between the response variables and their components. A negative correlation with Age^{-1} by definition means that as Age^{-1} decreases, the relationship decreases. Yet, since Age^{-1} is an inverse, the

relationship is inverted. A negative correlation still maintains an absolute positive relationship, where as Age^{-1} decreases similarly basal area increases. In examining correlation coefficients for volume, it is clear that age, height, and basal area have the highest correlations – which should be expected. An important finding is that allowing height and stand density to be transformed with a function of age significantly increased their corresponding correlation.

Simple linear relationships between stand characteristics of age, height, basal area, and stand density from the developed whole-stand models and SAR and LTM bands (including interaction terms) were generated to determine the strength of these relationships (Tables 11 A&B). Correlation coefficients for SAR and LTM band interaction terms were also calculated but are not presented here due to sere size of the correlation matrix. Examining Table 11.A it is evident that the C- and P-band had the highest correlation coefficients with stand variables. This finding is consistent with examination for linear relationships between SAR frequencies and response variables (Figures 5-7), which identified C- and P-bands having the strongest linear relationships. Specifically, P_{HV} and P_{HH} had the overall highest correlation coefficients with the given stand characteristics and are likely to be the most promising predictor variables among SAR bands based upon their high correlations. However, it is important to mention that overall, SAR band interaction terms consistently had higher correlation coefficients, and as a result, had stronger linear relationship with stand characteristics than the simple linear relationship for SAR bands individually. A second point to mention is that SAR band variance terms had considerably low correlation coefficients. But this should be expected since the variance term addresses the deviation of aggregated SAR pixels being

calibrated to a response variable. If high variances existed, then either SAR data did not adequately capture measured ground data, or ground data is highly variable. Small variances suggest that SAR data both captured on-ground measured data accurately and little variation exists within the stand.

Table 11.A. Correlation Coefficients Between SAR Bands and Stand Variables

SAR BAND	Age^{-1}	LN(ba)	LN(tpa)	$LN(tpa)Age^{-1}$	$LN(ba)Age^{-1}$	LN(hd)	$LN(hd)Age^{-1}$
C_{HH}	-0.4573	0.2149	-0.3582	-0.4617	-0.4260	0.3435	-0.4557
C_{VV}	-0.4419	0.5060	-0.1830	-0.4180	-0.3584	0.3893	-0.4264
C_{HV}	-0.4742	0.4220	-0.2406	-0.4536	-0.4114	0.3780	-0.4724
L_{HH}	-0.2970	-0.1491	-0.3642	-0.3127	-0.3372	-0.2422	-0.3097
L_{VV}	-0.2933	-0.0315	-0.2592	-0.2875	-0.3234	0.2734	-0.3076
L_{HV}	-0.3753	-0.03881	-0.4157	-0.3919	-0.4050	0.3794	-0.3744
P_{HH}	-0.6866	0.4401	-0.4320	-0.6733	-0.6275	0.5943	-0.6727
P_{VV}	-0.4437	0.3990	-0.2925	-0.4351	-0.3895	0.4262	-0.4332
P_{HV}	-0.7785	0.5103	-0.5531	-0.7679	-0.7237	0.7485	-76290

LTM band correlations were considerably lower than those from SAR bands, particularly with stand density and height. Temporal variation with the LTM bands does capture seasonal change in foliage conditions within a stand as seen by higher correlation coefficients for LTM-OFF data. LTM band ratios similarly had higher correlations than individual LTM bands. Again, increased correlations for ratios can be attributed to a ratio helping define a spectral relationship between band and forest characteristics.

Table 11.B. Correlation Coefficients Between LTM Bands and Stand Variables

LTM BAND	Age^{-1}	LN(ba)	LN(tpa)	$LN(tpa)Age^{-1}$	$LN(ba)Age^{-1}$	LN(hd)	$LN(hd)Age^{-1}$
ON-BAND 1	0.1933	-0.3277	0.0228	0.1827	0.1421	-0.1924	0.1722
ON-BAND 2	0.3717	-0.4632	0.1461	0.3612	0.3004	-0.3363	0.3486
ON-BAND 3	0.2850	-03946	0.0708	0.2686	0.2296	-0.2880	0.2645
ON-BAND 4	0.1390	-0.2076	0.05625	0.1354	0.0965	-0.1162	0.1270
ON-BAND 5	0.3049	-0.5394	0.0192	0.2738	0.2224	-0.3277	0.2811
ON-BAND 6	0.2748	-0.4454	0.0416	0.2509	0.2117	-0.3118	0.2521
OFF-BAND 1	0.1024	-0.3503	-0.0425	0.0828	0.0386	-0.1267	0.0869
OFF-BAND 2	0.1312	-0.3888	-0.00245	0.1168	0.0623	-0.2105	0.1040
OFF-BAND 3	0.1987	-0.4805	-0.0012	0.1754	0.1196	-0.2722	0.1725
OFF-BAND 4	0.2693	0.0552	0.3332	0.2873	0.2915	-0.2791	0.2641
OFF-BAND 5	0.2673	-0.5656	-0.0035	0.2339	0.1708	-0.2917	0.2447
OFF-BAND 6	0.2842	-0.3448	0.2070	0.2743	0.2275	-0.3759	0.2638

SAR and LTM bands and interaction terms that had the highest correlation coefficients with stand level characteristics were selected as candidate independent variables for the biologically motivated models. Although a high correlation does not guarantee that a specific band will be significant for the predictive model, it does assist in data reduction by selecting only those bands that have a certain known level of correlation. An important side note is that a low correlation coefficient does not automatically mean that there is no correlation. An array of possible relationships could result in a low correlation. However, analysis of all possible transformations falls outside the scope of this project and graphical interpretation of correlations between SAR and LTM bands with independent stand variables did not warrant further attention. Hence, criteria for selecting only those SAR and LTM bands being those having highest correlations was a valid selection method.

An identical layered model development design was applied for analyzing SAR and LTM data with the biologically motivated model technique.

MODEL SELECTION, BIOLOGICALLY MOTIVATED MODEL

Model selection from the biologically motivated models was identical to that previously performed for stepwise regression models using: (1) number of predictive variables, (2) Mean Squared Error (MSE), (3) R^2 and adjusted-R^2, and (4) residual plot analysis as selection criteria. Tables 12 A&B presents model fit statistics from models developed from the biologically motivated technique using SAR data.

Table 12.A. SAR Biologically Motivated Models for Estimating LN(*Volume*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	4	51.28	0.1151	0.3392	0.6635	0.6506
2	3	57.55	0.1281	0.3579	0.6218	0.6110
3	6	40.93	0.1023	0.3199	0.7066	0.6893
4	8	37.07	0.0897	0.2995	0.7478	0.7276
5	5	53.36	0.0962	0.3101	0.7215	0.7079
6	8	37.66	0.0886	0.2977	0.7508	0.7309
7	10	32.48	0.0841	0.2901	0.7682	0.7445

In examining Table 9.A it is clear that biologically motivated models generate a wider range of predictive model capability for estimating volume then did stepwise regression. This can be directly attributed to candidate variable selection criterion being the correlation with the independent stand level variables, rather than partial significance with the response variable as with the stepwise technique. For estimating volume, Model 6 and Model 7 are the most successful. Both models produce good fit statistics, yet

again, selection is based upon number of parameters within the model. Model 7 contains ten independent variables, while Model 6 contains only eight. Differences in R^2 and adjusted-R^2 support this decision. Also, Model 6 parameter estimates are all significant (p=.0001) and appears to be more parsimonious than Model 7. Model 6 consists of SAR band ratios and product interaction terms, and present best diagnostic statistics with an R^2 = 0.7508, adjusted-R^2=0.7309 and the lowest MSE=0.0886. Model 6 estimates volume as:

$$\mathbf{LN(Vol) = b_0 + b_1(C_{HH}\,P_{HH}) + b_2(C_{HH}\,P_{HV}) + b_3(L_{HV}P_{HV}) + b_4(P_{HH}P_{HV}) + b_5(P_{HV}/C_{VV}) + b_6(C_{VV}/\,L_{HH}) + b_7(L_{HH}/P_{HV}) + b_8(P_{HV}/\,L_{HV})} \qquad \mathbf{(4.7)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-206.975	37.09313	<.0001
b_1	0.00252	0.00044	<.0001
b_2	-0.00375	0.00060	<.0001
b_3	0.00233	0.00037	<.0001
b_4	-0.00154	0.00028	<.0001
b_5	31.46985	6.34672	<.0001
b_6	106.3931	21.07392	<.0001
b_7	39.78638	7.49638	<.0001
b_8	56.64722	10.08861	<.0001

Residual analysis illustrates that the assumption of constant variance has not been violated, that model form is appropriate, and supports the selection of Model 6 due to tighter fit and range of the residuals. Close inspection of the model shows that the P-band (specifically P_{HV}) has the most influence. But it is also interesting to notice that there is a full array of band interaction terms within the model. Clearly, wavelength properties do allow for capturing different aspects within a stand's composition.

Table 12.B. SAR Biologically Motivated Models for Estimating LN(*Basal Area*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	6	13.24	0.0449	0.2118	0.4379	0.4048
2	2	33.82	0.0469	0.2165	0.3896	0.3780
3	7	11.24	0.0453	0.2129	0.4378	0.3988
4	8	14.30	0.0380	0.1949	0.5336	0.4963
5	7	13.21	0.0421	0.2051	0.4780	0.4418
6	7	13.24	0.0420	0.2050	0.4785	0.4424
7	8	10.84	0.0436	0.2088	0.4643	0.4215

For estimating basal area, Model 4 is clearly the most successful producing the strongest fit statistics, and uses SAR band averages and variances with band ratio terms. Model 4 defines the prediction of basal area as:

$$\mathbf{LN(BA) = b_0 + b_1(C_{HH}) + b_2(P_{HV}) + b_3(C_{HH}\text{-var}) + b_4(L_{HV}\text{-var}) + b_5(P_{HV}/C_{VV}) + b_6(C_{VV}/L_{HH}) + b_7(P_{HV}/C_{HH}) + b_8(L_{HH}/P_{HV})} \quad \textbf{(4.8)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-70.59930	21.62000	0.0015
b_1	0.38082	0.10685	0.0006
b_2	-0.26020	0.06879	0.0003
b_3	0.00331	0.00134	0.0152
b_4	-0.00415	0.00126	0.0013
b_5	6.85188	2.72361	0.0135
b_6	30.41594	9.06391	0.0011
b_7	26.28725	6.89036	0.0002
b_8	8.36444	3.41165	0.01600

It is interesting to note that the two models for predicting volume and basal area consist of nearly identical predictor variables, leading to the finding that certain bands are consistently better predictive variables for estimating physical attributes. This finding is unique to this technique, where stepwise had little consistency with band selection, biologically motivated models for SAR do. SAR bands C_{HH} and P_{HV} are consistently selected as predictive variables either as mean band coefficient or as interaction terms.

Selection of these bands is not limited to the models discussed as best models, but also are selected the in additional models generated with SAR data.

Models developed using LTM data with the biologically motivated technique to estimate volume and basal area are presented in Tables 10 A&B. Like the stepwise regression technique, LTM data does not do as well for predicting either basal area or volume as does SAR data.

Table 13.A. LTM Biologically Motivated Models for Estimating LN(*Volume*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	4	30.00	0.1588	0.3985	0.5357	0.5179
2	4	25.51	0.1726	0.4155	0.4952	0.4758
3	9	18.34	0.1347	0.3670	0.6251	0.5910

For estimating volume, Model 3 produced the best fitted model. The combination of LTM band mean and variance terms with band ratios had more success at developing an estimation model that might have future potential. Model 3 produces a considerably better fit statistics of R^2=0.6251, adjusted-R^2=0.5910 with an MSE=0.1347. Residual analysis supports selection of this model as no assumptions of errors are violated. Model 3 predicts volume as:

$$\mathbf{LN(VOL) = b_0 + b_1(on\text{-}B4\text{-}var) + b_2(off\text{-}B3) + b_3(off\text{-}B5) + b_4(off\text{-}B7) + b_5(on\text{-}B1/on\text{-}B2) + b_6(on\text{-}B3/on\text{-}B5) + b_7(off\text{-}B1/off\text{-}B4) + b_8(off\text{-}B1/off\text{-}B7)\ \ b_9(off\text{-}B3/off\text{-}B5)} \qquad \textbf{(4.9)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-6.26091	3.71360	0.0965
b_1	-0.00466	0.00135	0.0008
b_2	0.84650	0.30188	0.0061
b_3	-0.57265	0.17862	0.0018
b_4	0.78896	0.22966	0.0009

b_5	2.71817	0.82618	0.0014
b_6	7.66831	2.14671	0.0005
b_7	2.88306	0.77383	0.0003
b_8	1.09104	0.29295	0.0003
b_9	-22.1044	7.42897	0.0037

LTM band selection from the biologically motivated model resembles those selected by the stepwise procedure. Both methods consistently select Band-3, -4, -5, and -7 for both LTM-ON and LTM-OFF. These bands' sensitivities to chlorophyll, vegetation vigor and growth cause them to be repeatedly selected. Band-1 and –2 are heavily influenced by moisture and atmospheric haze and may likely be unsuitable individually, but a ratio term of Band-1 and Band-2 is likely to be more beneficial.

Table 13.B. LTM Biologically Motivated Models for Estimating LN(*Basal Area*)

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	3	21.59	0.0479	0.2190	0.3815	0.3638
2	6	16.24	0.0408	0.2020	0.4886	0.4585
3	5	18.58	0.0416	0.2039	0.4742	0.4487

Both Model 2 and Model 3 produce nearly identical fit statistics except for number of predictor variables. Model 2 was chosen as the best fitted model for estimating basal area. Visual analysis of residuals from both Model 2 and Model 3 illustrate a random and independent distribution. Neither model violates the assumption of constant variance nor displays a pattern of heteroscedasticity but Model 2's range of residuals is tighter around zero and may suggest a better model form. Model 2 for estimating basal area using LTM data developed the following model:

$$\mathbf{LN(BA) = b_0 + b_1(on\text{-}B1/on\text{-}B2) + b_2(on\text{-}B3/on\text{-}B5) + b_3(off\text{-}B2/off\text{-}B5) +} \qquad \mathbf{(4.10)}$$

b_4(off-B2/off-B6) + b_5(off-B3/off-B5) + b_6(off-B3/off-B6)

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-0.65507	0.89056	0.4637
b_1	1.18935	0.40303	0.0039
b_2	3.30930	1.02999	0.0018
b_3	-32.31390	8.51172	0.0002
b_4	10.15415	2.65651	0.0002
b_5	40.50002	10.39954	0.0002
b_6	-12.24753	3.24028	0.0003

Specific SAR bands were discussed as consistently being contributing variables. In terms of LTM data, there are similar relationships that are consistently identified as significant regressors. Ratio terms from Band-3 and –5 is an example of such and their significance will be addressed later.

The combination of both SAR and LTM data were also examined for estimating basal area and volume using the biologically motivated technique. Fit statistics from models developed from this approach are presented in Tables 14 A&B.

Table 14.A. Combination of SAR and LTM, Biologically Motivated Model for Estimating LN(*Volume*).

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	8	35.71	0.0922	0.3037	0.7407	0.7199
2	8	37.66	0.0886	0.2977	0.7508	0.7309
3	9	36.46	0.0833	0.2889	0.7682	0.7472

Model 3 for predicting volume presents the overall highest R^2=0.7682 and adjusted-R^2=0.7472 with an MSE=0.0833 and is selected as the best model for estimating volume using both SAR and LTM data. Model 2 is very similar in fit statistics and has one less predictor variable. Selection of Model 3 is supported by two considerations. First, the

adjusted-R^2 is higher for Model 3, and shows that the R^2 is not inflated die to additional predictor variable. And secondly, Model 2 failed to capture significance influence from LTM bands. Since a principle purpose was to determine the effect of combing both SAR and LTM data for estimating volume, it was deemed appropriate to select Model 3 which incorporates both sources of data. Residual comparison between the two models also supported selection of Model 3. Residuals from Model 2 indicate that the model tends to have more variability at the mid- and high- levels of volume, whereas Model 3 has less variability and tighter distribution of residuals. Model 2 fit the data for estimating volume as:

$$\mathbf{LN(Vol) = b_0 + b_1(C_{VV}) + b_2(C_{HV}) + b_3(P_{HH}) + b_4(C_{HV}/P_{HH}) + b_5(C_{VV}/L_{HH}) + b_6(L_{HH}/P_{HV}) + b_7(C_{HV}P_{HH}) + b_8(\text{on-B3/on-B5}) + b_9(\text{off-B1/off-B4})} \quad \mathbf{(4.11)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-116.51139	25.27435	<.0001
b_1	-0.43761	0.11814	0.0003
b_2	1.42307	0.30803	<.0001
b_3	0.34717	0.07090	<.0001
b_4	-91.67849	20.97441	<.0001
b_5	84.27483	22.43455	0.0003
b_6	35.62128	8.93598	<.0001
b_7	-0.00406	0.00085	<.0001
b_8	6.72527	1.70632	0.0002
b_9	1.87993	0.57835	0.0016

The estimation model for volume from combining both SAR and LTM is a promising finding for combining imaging and optical imaging systems. SAR frequencies selected are consistent with earlier models that used both the C- and P-band individually, as well as, band interactions. Repeated selection of specific SAR bands may possibly be due to a unique band-intensity relationship being defined that is beneficial for modeling volume. The model also selects two functioning LTM bands that have likely been previously been

beneficial for estimating volume. The combination of the two sources, presumably combines SAR's ability to penetrate and LTM's ability to differentiate and separate foliage layers that may be disrupting SAR backscatter.

Table 14.B. Combination of SAR and LTM, Biologically Motivated Model for Estimating LN(*Basal Area*).

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
1	5	16.79	0.0435	0.2087	0.4491	0.4224
2	7	13.24	0.0420	0.2050	0.4785	0.4424
3	7	15.27	0.0392	0.1979	0.5142	0.4806

Examination of combining SAR and LTM data for estimating basal area, combination of mean band coefficients and variances with band interaction terms produces the strongest predictive model. Model 3 incorporates use of both SAR and LTM data as predictive variables, and estimates basal area as the following:

$$\mathbf{LN(BA) = b_0 + b_1(C_{HH}) + b_2(P_{HV}) + b_3(C_{VV}\text{-var}) + b_4(L_{HH}\text{-var}) + b_5(C_{VV}/L_{HH}) + b_6(P_{HV}/C_{HH}) + b_7(on\text{-}B1/on\text{-}B2)} \quad \textbf{(4.12)}$$

Parameter	Estimate	Std. Error	Pr > |t|
b_0	-23.1779	8.0785	0.005
b_1	0.21205	0.08254	0.0107
b_2	-0.15141	0.05342	0.0055
b_3	0.00374	0.00157	0.0186
b_4	-0.00452	0.00132	0.0009
b_5	6.82295	1.53974	<.0001
b_6	15.43381	5.29599	0.0044
b_7	1.1594	0.48427	0.0185

Comparing fit statistics from models to estimate basal area developed from combining SAR and LTM data, it is apparent that models developed through the layered design progressively increase in ability to account for variation in basal area. Selection of Model 3 as the most appropriate model is supported by it highest R^2, adjusted- R^2 and

lowest MSE. Residual analysis implies that model is appropriate and no violations of errors are apparent.

Important for model selection where correlation analyses were used to examine relationships between SAR and LTM spectral responses and forest stand characteristics, are difficulties that may arise due to possible multicolinearity between various stand level attributes. This interdependence, partially unique to forestry purposes, is desired and expected with modeling forest growth and yield. Recognition of this condition with model forest attributes may require this to be addressed according to modeling objectives.

CHAPTER V.

RESULTS AND DISCUSSION

BEST MODEL PRESENTATION

The success of models for estimating basal and volume have been addressed using both stepwise multiple-linear regression and a more biologically motivated multiple-linear model approach.

Comparing SAR and LTM data independently, it has been shown that SAR data explained more variation in both basal area and volume. Comparison between stepwise regression and biologically motivated models indicate that the biologically motivated models produce slightly better model fit statistics than did stepwise regression. This is an important finding because the development of prediction models from a biologically motivated model form provides more confidence in the model's extrapolation behavior than do those developed based solely upon statistical correlation (stepwise procedures). Tables 15 – 17 present fit statistics from the best models from the each modeling approach.

STEPWISE REGRESSION VS. BIOLOGICALLY MOTIVATED MODEL FOR ESTIMATING VOLUME USING SAR DATA

The best fitted model to estimate volume using stepwise multiple-linear regression included four parameters consisting of SAR band ratios and product interaction terms and produced an R^2=0.7221 and adjusted- R^2=0.7114 with an

MSE=0.095 (Equation 5.1). The biologically motivated technique improved model fit to R^2=0.7508 and adjusted-R^2=0.7309 with an MSE=0.2977 (Equation 5.2).

$$\mathbf{LN(Vol) = b_0 + b_1(C_{HH}/P_{HV}) + b_2(P_{HV}/C_{HH}) + b_3(P_{HH}C_{HH}) + b_4(P_{HH}P_{HV})} \quad \mathbf{(5.1)}$$

$$\mathbf{LN(Vol) = b_0 + b_1(C_{HH}P_{HH}) + b_2(C_{HH}P_{HV}) + b_3(L_{HV}P_{HV}) + b_4(P_{HH}P_{HV}) + b_5(P_{HV}/C_{VV}) + b_6(C_{VV}/ L_{HH}) + b_7(L_{HH}/P_{HV}) + b_8(P_{HV}/ L_{HV})} \quad \mathbf{(5.2)}$$

Table 15.A. Stepwise Regression vs. Biologically Motivated Model to Estimate Volume Using SAR Data

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	4	67.56	0.0950	0.3083	0.7221	0.7114
BIO	8	37.66	0.0886	0.2997	0.7508	0.7309

Estimates from these models are similar with findings presented from published literature that estimated volume as a function of SAR data. Vona (2000) is the most valid comparison. Using the identical SAR scene Vona (2000) developed a system of simultaneous equations to predict volume from managed slash pine plantations. Strong linear relationships were found and modeled with the system of equations, and concluded with a R^2=84% for predicting volume. Vona's (2000) research was designed for a plot level analysis rather than a stand level approach as used in this study so results are not directly comparable. However, the two studies provide corroborating evidence for SAR's ability to estimate volume.

The biologically motivated models account for more variation and produce a higher R^2 and adjusted-R^2 with a lower MSE for estimating volume than do models derived using the stepwise regression technique. One might attribute the increase in R^2 to be inflated due to additional parameters, but interpretation of the adjusted-R^2 accounts for the additional regressors within the model and supports this finding. There may, however, be a simple explanation to why biologically motivated models contain more contributing predictor variables than do stepwise regression models. Biologically motivated models were developed from correlation analysis with unique stand level characteristics. In contrast, stepwise regression models selected candidate variables due to their partial correlation with the response variable. The whole-stand model to estimate volume consisted of seven parameters, using correlation analyses with these seven stand characteristics introduced a larger number of candidate variables into the predictive model, and these were reduced according to their significance. Stepwise regression models developed through correlation with the response variable, had greater restriction for entrance of variables due to a predetermined level of significance (a=.15). Loosening this entrance requirement may likely have increased the number of significant variables introduced into the model.

Examining the predictive influence of SAR bands, it is clear that overall the P-band most heavily influences a model's ability to estimate volume and accounts for the majority of the variability. This finding is consistent with past literature and supports the premise that P-band's ability to interact directly with the stem and more structural stand components is most effective for estimating volume (Dobson *et al.* 1995, Kasischke *et al.* 1995, Le Toan *et al.* 1992, and Vona 2000). Specifically, P_{HV} has the greatest sensitivity

to volume components, which can be supported by having the highest simple linear correlation (R^2=62%) with volume. This finding indicates that not only is the P-band most significant, but also that the cross-polarized –HV orientation is overall most effective for returning stronger intensity of backscatter measurement. Consistently, the best SAR band predictor variables were band ratio and product terms. SAR band ratios and product terms are used to capture information provided not only by the bands themselves but also to define a backscatter intensity relationship between two bands. The most commonly used interaction terms consisted of both the C- and P-band. Interaction between a short wavelength (C-band) and a long wavelength (P-band) may provide information for defining a backscatter intensity relationship that is not obvious when comparing the bands independently, and this intensity relationship better models the response variable.

STEPWISE REGRESSION VS. BIOLOGICALLY MOTIVATED MODEL FOR ESTIMATING BASAL AREA USING SAR DATA

SAR data were not as useful for estimating basal area as they are for estimating volume. Stepwise multiple-linear regression developed a six parameter model with an R^2=0.5342 and adjusted- R^2=0.5068 with an MSE=0.0372 (Equation 5.3.). A very similar model was developed using the biologically motivated technique which produced an R^2=0.5336 and adjusted-R^2=0.4963 with an MSE=0.0379 (Equation 5.4.).

$$\mathbf{LN(BA) = b_0 + b_1(C_{HH}) + b_2(C_{HH}\text{-var}) + b_3(L_{VV}\text{-var}) + b_4(P_{HV}) + b_5(C_{VV}/L_{VV}) + b_6(P_{HV}/C_{HH})} \quad \mathbf{(5.3)}$$

$$LN(BA) = b_0 + b_1(C_{HH}) + b_2(P_{HV}) + b_3(C_{HH}\text{-var}) + b_4(L_{HV}\text{-var}) + b_5(P_{HV}/C_{VV}) + b_6(C_{VV}/L_{HH}) + b_7(P_{HV}/C_{HH}) + b_8(L_{HH}/P_{HV}) \quad (5.4)$$

Table 15.B. Stepwise Regression vs. Biologically Motivated Model to Estimate Basal Area Using SAR Data

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	6	19.50	0.0372	0.1928	0.5342	0.5068
BIO	8	14.30	0.0379	0.1949	0.5336	0.4963

However, inability to account for more variation may be equally attributed to the measured inventory data as it is with SAR's ability to accurately capture variation within basal area components. Generally, basal area is more difficult to predict than volume using remotely sensed data. Forest structure heavily influences stand basal area, and it is quite possible, for equal basal area (ft^2) per acre to have considerably different stand structure. Recall the previously fitted whole-stand model to predicting basal area, measured inventory data only accounted for 85% of the variation in basal area. Although a 'proportional' comparison is not legitimate, it is strongly hypothesized that while possible perhaps, it is highly unlikely that SAR data would estimate basal area better than field sampled measurements.

Comparing the results from the two developed models with past research, the findings are slightly lower than published literature that have estimated basal area as a function of SAR data. Vona (2000) reported validated model prediction estimates of R^2=80% for slash pine plantation basal area. Harrell *et al.* (1997), in a comparative study applying four attribute estimation techniques compiled from previous research efforts (

by Dobson *et al.* 1995, Kasischke *et al.* 1995, and Ranson and Sun 1994), reported validated model estimates from the various techniques ranging from R^2=63% to 74%.

Both model development techniques selected near identical predictor variables. This finding has an important implication. Selection of near identical predictor variables ensures that statistical inference is not being lost using correlation analyses with stand characteristics for the biologically motivated technique rather than the correlation with response variable as with stepwise regression.

Selection of SAR bands as predictor variables to estimate basal area differed from those selected to estimate volume. The most obvious difference is in the change of SAR bands that most heavily influenced the fitted model. To estimate basal area, short and mid-length C- and L-bands were more predictive. Whereas, to estimate volume the longer P-band was most influential for model development. Examining correlation coefficients for C- and P-bands show that they are more highly correlated with the measure of stand density which basal area is more dependent upon than with other stand characteristics.

Interestingly, SAR band variance terms were significant contributing predictor variables to estimate basal area. Identical SAR bands, C_{HH}-variance and L_{HH}-variance, were selected for inclusion with both model development techniques. Addition of variance terms from aggregating pixels is a new approach for forest attribute estimation, and subsequently these findings indicate this approach deserves further exploration.

STEPWISE REGRESSION VS. BIOLOGICALLY MOTIVATED MODEL FOR ESTIMATING VOLUME USING LTM DATA

Fitted models to estimate volume using LTM data were not as successful with either development technique as were those using SAR data. Employing the stepwise regression technique, a six parameter model was developed with an R^2=0.6281 and adjusted-R^2=0.6062 with an MSE=0.1297 (Equation 5.5). In comparison of the biologically motivated model producing an R^2=0.6251 and adjusted-R^2=0.5910 with an MSE=0.1347 (Equation 5.6). Comparing fit statistics suggests that the stepwise regression model better fit the data to estimate volume than the biologically motivated model.

$$\mathbf{LN(VOL) = b_0 + b_1(on\text{-}B4\text{-}var) + b_2(on\text{-}B5\text{-}var) + b_3(off\text{-}B5\text{-}var) + b_4(on\text{-}B2/on\text{-}B1) + b_5(on\text{-}B3/on\text{-}B7) + b_6(off\text{-}B3/off\text{-}B4)} \quad \textbf{(5.5)}$$

$$\mathbf{LN(VOL) = b_0 + b_1(on\text{-}B4\text{-}var) + b_2(off\text{-}B3) + b_3(off\text{-}B5) + b_4(off\text{-}B7) + b_5(on\text{-}B1/on\text{-}B2) + b_6(on\text{-}B3/on\text{-}B5) + b_7(off\text{-}B1/off\text{-}B4) + b_8(off\text{-}B1/off\text{-}B7)\ \ b_9(off\text{-}B3/off\text{-}B5)} \quad \textbf{(5.6)}$$

Table 16.A. Stepwise Regression vs. Biologically Motivated Model to Estimate Volume Using LTM Data

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	6	28.71	0.1297	0.3601	0.6281	0.6062
BIO	9	18.34	0.1347	0.3670	0.6251	0.5910

However, comparing residuals from the two models, residuals from the stepwise regression technique have more troublesome distribution than do residuals from the biologically motivated model, which present a tighter distribution of errors. Further

examination suggests a single observation is negatively influencing the regression line for the biologically motivated model, and removal of this likely outlier would significantly improve fit, and consequently further reduce residual distribution.

Comparing predictor variables from the two models developed to estimate volume, it is evident there is consistency with certain LTM bands selected particularly, LTM Band –3 and –5. LTM Band – 3 is highly sensitive to chlorophyll absorption, and as a result, forest conditions with healthy vegetation and foliage layers reflect smaller amounts of radiation than do those areas that are less healthy or have overall lower levels of vegetation. LTM Band – 5 is similarly heavily absorbed by increased levels of healthy vegetation (Lillesand and Kieffer 1994). Sensitivity to vegetation density and health, leaf moisture, and canopy shading provokes reflectance and absorption variability for forest conditions. Generally, increasing vegetation results in increased absorption and subsequently less reflectance among these bands. Applying this rationale, stands with lower levels of density will have lower ratio relationships with Band – 3 and –5 , while increasingly dense stands might have higher ratio relationships. Another common LTM band ratio was exhibited between LTM Band –1 and Band – 2. Both bands have a negative relationship with volume (Figures 9 and 10). Yet, Band – 1 and - 2 have a positive linear correlation with each other. The range for spectral reflectance digital numbers for conifers for these bands makes the ratio driven by LTM Band – 1 and regulated by Band – 2. For both bands, as vegetation increases, absorption increases and reflectance decreases. So the ratio is generally highly correlated with growth.

Comparing fitted models from the two development techniques suggests that they may not be fully capable of estimating volume with LTM data. Ripple *et al.* (1991) using

LTM-5 recorded results of R^2=70% for predicting volume at the plot level. Lowe (2002) applied a multiple-linear technique to an clustered plot level analysis, and derived a validated R^2=76% for predicting volume (ft^3) for pine plantation, in southeast Georgia. However, it is worth mentioning that the objective of this research was to compare SAR and LTM data from two model development techniques, not to specifically optimize prediction estimates using LTM data. As a result, a full exploration of data transformations and specific algorithms for LTM processes were not investigated.

STEPWISE REGRESSION VS. BIOLOGICALLY MOTIVATED MODEL FOR ESTIMATING BASAL AREA USING LTM DATA

Similar to the results to estimate volume, using LTM data to fit models to estimate basal area was not as successful as from using SAR data. Stepwise regression model produced an R^2=0.4222 and adjusted-R^2=0.4057 with an MSE=0.0448 (Equation 5.7). The biologically motivated model technique accounted for more variation, and improved fit statistics to an R^2=0.4886 and adjusted-R^2=0.4585 with an MSE=0.0408 (Equation 5.8). Comparative residual analysis from the two models show that residuals from the stepwise regression model exhibit more pattern structure than do those from the biologically motivated model, suggesting the biologically motivated model may be a better model form.

$$\mathbf{LN(BA) = b_0 + b_1(on\text{-}B3/on\text{-}B5) + b_2(off\text{-}B1/off\text{-}B5) + b_3(off\text{-}B5/off\text{-}B4)} \quad \mathbf{(5.7)}$$

$$\mathbf{LN(BA) = b_0 + b_1(on\text{-}B1/on\text{-}B2) + b_2(on\text{-}B3/on\text{-}B5) + b_3(off\text{-}B2/off\text{-}B5) + b_4(off\text{-}B2/off\text{-}B6) + b_5(off\text{-}B3/off\text{-}B5) + b_6(off\text{-}B3/off\text{-}B6)} \quad \mathbf{(5.8)}$$

Table 16.B. Stepwise Regression vs. Biologically Motivated Model to Estimate Basal Area Using LTM Data

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	3	25.58	0.0448	0.2117	0.4222	0.4057
BIO	6	16.24	0.0408	0.2020	0.4886	0.4585

Fitted models to estimate basal area using LTM data were overall the least successful. The issue of measured inventory data may once more be pertinent for model performance. Again, selection of LTM band ratios between Band – 3 and – 5 heavily influencing the fitted model. Secondary band ratios exhibited are between Band – 4 and – 5 and also between Band – 3 and – 6. Both ratios are commonly cited for coniferous classification and attribute estimation retrieval, and they generally perform very similarly. Band – 5 is sensitive to leaf-moisture in the crown and exhibits an inverse relationship. Conversely, Band– 4 is unaffected by leaf-moisture. Higher values in Band –5 indicates drier foliage layers, less canopy, and lower moisture content leading to lower ratios values.

Comparing the results from the fitted models supports the premise that if LTM data is fully optimized, it does have better potential for forest attribute estimation than found in this analysis. Lowe (2002) and Phelps (2000) reported R^2=78% and 72% respectively for predicting basal area for pine plantations applying a clustered, plot level analysis.

STEPWISE REGRESSION VS. BIOLOGICALLY MOTIVATED MODEL FOR ESTIMATING VOLUME AND BASAL AREA WITH COMBINED SAR AND LTM DATA

Developing models to estimate basal area and volume from combining SAR and LTM data were also explored. "*Best*" model fit statistics from the two model development technique for estimating basal area and volume are presented in Tables 17 A&B.

In comparing Tables 17.A&B, the stepwise regression technique accounts for more variability to estimate volume than does the biologically motivated technique. The stepwise regression model produced an R^2=0.7957 and adjusted-R^2=0.7794 with an MSE=0.0726 (Equation 5.9), while the biologically motivated model produced an R^2=0.7682 and adjusted-R^2=0.7472 with an MSE=0.0833 (Equation 5.10).

$$\mathbf{LN(Vol) = b_0 + b_1(C_{HV}\text{-var}) + b_2(L_{VV}\text{-var}) + b_3(L_{VH}\text{-var}) + b_4(P_{HV}) + b_5(P_{HH}/C_{HH}) + b_6(\text{on-B3/on-B7}) + b_7(\text{on-B7/on-B2}) + b_8(\text{off-B1/off-B4})} \quad \mathbf{(5.9)}$$

$$\mathbf{LN(Vol) = b_0 + b_1(C_{VV}) + b_2(C_{HV}) + b_3(P_{HH}) + b_4(C_{HV}/P_{HH}) + b_5(C_{VV}/L_{HH}) + b_6(L_{HH}/P_{HV}) +} \quad \mathbf{(5.10)}$$
$$b_7(C_{HV}P_{HH}) + b_8(\text{on-B3/on-B5}) + b_9(\text{off-B1/off-B4})$$

Table 17.A. Best Stepwise Regression Model from Combined SAR and LTM data to Estimate Volume

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	8	48.70	0.0726	0.2695	0.7957	0.7794
BIO	9	36.46	0.0833	0.2886	0.7682	0.7472

Further examination of ANOVA's from the two regression models showed the stepwise regression procedure, which requires predictor variables only to be significant at a=.05, has allowed two questionable variables considerably near a=.05 to remain in the model. Use of a predetermined significance level that regressors must satisfy for entrance and removal differs from the more iterative, biologically motivated technique which applied a more stringent approach to ensure the model was as parsimonious as possible. It is hypothesized that removal of these variables in question would likely reduce the overall fit of the stepwise regression model and make the two models more similar in proportion of variation explained. However, the increase attributed to combining SAR and LTM can be better exhibited when comparing the individual models from the layered model design presented in Tables 4, 9 and 11. In examining these tables, the addition of LTM to the SAR model increased the R^2 by nearly 8% for estimating volume both model development techniques – a clear indicator that the two sensor due lend complementary information that when combined yield a greater success for accurate volume estimates.

Combining SAR and LTM data to estimate basal area was also successful for improving model fitness from the stepwise regression models. Applying stepwise regression to estimate basal area produced a model having an R^2=0.5079 and adjusted-R^2=0.4890 with an MSE=0.0385 (Equation 5.11). Similarly, the biologically motivated technique developed a fitted model with an R^2=0.5142 and adjusted-R^2=0.4806 with an MSE=0.0392 (Equation 5.12). Examination of the fit statistics from the two models, stepwise regression model has a slightly higher adjusted-R^2. With such minimal

differences in fit statistics, the biologically motivated model would be preferred for extrapolation purposes due to its more forest biological inference for model development.

$$\mathbf{LN(BA) = b_0 + b_1(C_{VV}\text{-}var) + b_2(L_{VV}\text{-}var) + b_3(L_{HH}/C_{VV}) + b_4(on\text{-}B1/on\text{-}B5)} \quad \mathbf{(5.11)}$$

$$\mathbf{LN(BA) = b_0 + b_1(C_{HH}) + b_2(P_{HV}) + b_3(C_{VV}\text{-}var) + b_4(L_{HH}\text{-}var) + b_5(C_{VV}/L_{HH}) + b_6(P_{HV}/C_{HH}) + b_7(on\text{-}B1/on\text{-}B2)} \quad \mathbf{(5.12)}$$

Table 17.B. Best Stepwise Regression Model from Combined SAR and LTM data to Estimate Basal Area

MODEL	D.F.	F-VALUE	MSE	RMSE	R^2	ADJ-R^2
STEP	4	26.84	0.0385	0.1963	0.5079	0.4890
BIO	7	15.27	0.0392	0.1979	0.5142	0.4806

Closer examination of the two models identifies that vastly different predictor variables were selected. The stepwise regression model consists of a combination of SAR and LTM bands, whereas the biologically motivated model is comprised only SAR bands. The biologically motivated technique fails to capture possible LTM inference for modeling basal area. This finding indicates that independently, SAR data is better suited for estimating basal area, but also, combining SAR and LTM data may be unproductive for increasing the ability to explain more variation applying these conditions.

CHAPTER VI.

CONCLUSIONS

Forest inventory systems are depend upon accurate and efficient forest estimates, primarily to predict and project forest growth. Remotely sensed data retrieved over various forest types provides unique spectral responses that are determined by the forest's dynamics and characteristics. To investigate the potential of integrating new remote sensing techniques with forest inventory systems, relationships between SAR and LTM data with slash pine attributes were examined to assess potential capability to estimate volume and basal area. Models were developed using multiple-linear regression. In a new approach for forest attribute estimation, SAR and LTM data were also combined to determine if spectral properties of SAR and LTM are complementary with each other to produce better prediction models. Models were developed using two variable selection techniques to determine whether pure statistical relationships or a more biologically motivated model produce more appropriate predictive models.

Models derived using SAR and LTM data to predict stand basal area and volume was found to explain more variation in both volume and basal area. Models fit to SAR data consistently produced higher R^2 and adjusted- R^2 with lower MSEs. This finding is important for possible integration into forest inventory systems and management implications because of the biological development experienced within coniferous forests. A coniferous canopy layer develops within the formative years, and generates a complex crown layer that generally reaches full-closure by approximately age eight.

Conversely, basal area and volume slowly develop through formative years, but continue to develop well after crown closure. Optical imaging sensors like LTM may not well suited to capture this continuous growth since they are more sensitive to crown composition and generally reach a saturation level similar in time to complete crown closure. SAR's unique ability to penetrate the crown and interact with the woody and stem systems allows for more complete information on an attribute's structural composition to be extracted from the stands. A second explanation of SAR's ability to more accurately estimate volume and basal area is its higher ground resolution. This finer resolution reduces introduced heterogeneity and allows for more 'pure' signatures to be extracted that are more representative of the forest conditions.

In an attempt to capture the unique spectral and spatial sensitivities from both SAR and LTM data the two data sources were combined to estimate volume and basal area. Combining radar and optical imaging data is a new approach for forest attribute estimation. In nearly all instances, combining SAR and LTM data produced sizable improvements in explaining variation of both volume and basal area. The overall best model for estimating volume using the synthesized data produced significantly higher R^2 and adjusted-R^2 with lower MSE than from either SAR or LTM data individually. These findings suggest the two sources do lend unique information, which when combined may allow for more accurate and precise attribute estimates. This is a promising finding for forest attribute estimation and may warrant further attention.

Models were also developed using two model development techniques – stepwise multiple-linear regression and a biologically motivated multiple-linear regression technique. Developing predictive models from correlation analysis of SAR and LTM

bands with basal area and volume components allowed for biological inference to be associated with the fitted model, and consequently better predicted stand attributes. Models from the biologically motivated model approach consistently explained more variation in basal area and volume. Extrapolation of models developed from known biological relationships of stand development promotes confidence in model behavior relative to those developed through stepwise regression.

Prediction models developed at a stand level rather than a plot level produced very comparable results in ability to estimate basal area and volume. For integration into a forest inventory system, stand level models are preferred, and will greatly reduce calibration time and expense associated with developing prediction models.

CHAPTER VII.

REFERENCES CITED

Ahern, F.J., D.G. Leckie, D. Werle: Applications of RADARSAT SAR Data in Forested Environments (volume 19, no 4, novembre-décembre 1993) p. 330

Baker, R.D. and D.V. Smith. 1970. Aerial Photography Use In Industrial Forest Management In the South, 1970 and 1974. Texas Forestry Paper, School of Forestry, Stephen F. Austin State University.

Beaudoin, A., T. LeToan, S. Goze, E. Nezry, A. Lopes, E. Mougin, C. C. Hsu, H. C. Han, J. A. Kong, and R. T. Shin, 1994, Retrieval of forest biomass from SAR data, Int. J. Rem. Sens., vol. 15, no. 14, pp. 2777-2796.

Carter, D.R., L.G. Arvantes, D. Brackett, V.Boychera, and S.Sager. 1997. A decision support system for timber harvest scheduling. J. For., 97:12-18.

Chipman, J.W., T.M. Lillesand, J.D. Gage, and S. Radcliffe. 2000/ Spaceborne Imaging Radar in Support of Forest Resource Management. *Photogrammetric Engineering and Remote Sensing.* 19(11): 1357-1366.

Clutter, J.L., J.C. Fortson, L.V. Pienaar, G.H. Brister, and R.L. Bailey. 1983. *Timber Management: A QuantitativeApproach.* Krieger Publishing Company. 333pp.

DeMers M.N, 1997, Fundamentals of Geographic Information Systems: John Wiley and Sons Inc.

Dobson, M.C., F. T. Ulaby, T. Le Toan, E.S. Kasischke, and N. Christensen (1991), Dependence of Radar Backscatter on Coniferous Forest Biomass, *IEEE Trans. Geosci. Remote Sens*. 30:412-415.

Dobson, M.C., F.T.Ulaby, T. Le Toan, A. Beaudoin, E.S. Kasischke, and N. Christensen, 1992, Dependence of radar backscatter on coniferous forest biomass, *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 412-415.

Dobson, M.C., F.T.Ulaby, and L. E. Pierce, (1995), Landcover classification and estimation of terrain attribute using synthetic aperture radar. *Remote Sens. Environ*. 51:199-214.

Dobson, M.C., L.E. Piercce and F.T. Ulaby, 1996. Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites, *IEEE Trans. Geosci. Remote Sens.*, 34(1): 83-99.

Draper, N. R. and Smith, H. (1981). Applied Regression Analysis. Wiley, New York.

ERDAS IMAGINE, 1997. "ERDAS Field Guide and Tour Guides, version 8.3". Fourth edition. ERDAS, Inc. Atlanta, Georgia, USA.

EOS SAR Facility Instrument Team, (1991), The Earth observing system synthetic aperture radar (EOS SAR). Report to the Space Science and Applications Advisory Committee for the Office of Space Science and Applications 1992 Strategic Plan. Pasadena, California: Jet Propulsion Laboratory.

Franklin, Steven E., Remote Sensing for Sustainable Forest Management, CRC Press LLC. Florida, 2000.

Fransson, J. E. S., and H. Israelsson, (1999), Estimation of stem volume in boreal forest Using ERS-1 C- and JERS-1 L-band SAR data, *Int. J. Remote Sens*. 20:123-137.

Harrell, P., E. Kasischke, L. Borgeau-Chavez, E. Haney, and N. Christensen, 1997, Evaluation of approaches to estimate aboveground biomass in southern pine forests using SIR-C data, *Remote Sens. Environ*., vol. 59. no. 2, pp. 223-233.

Hussin, Y.A., Reich, R. M., Hoffer, R. M. (1989). Relationships between multipolarized radar backscatter and slash pine stand parameters. *Proc. 1989 ASPRS-ACSM Fall Convention, Clevland, Ohio pp.427-431.*

Hussin, Y.A., Reich, R. M., Hoffer, R. M. (1991). Estimating Slash Pine Biomass Using radar Backscatter, *IEEE Trans. Geosci. Remote Sens*. 29:427-431.

Iverson, L.R., E.A. Cook, and R.L. Graham. 1990. Estimating forest cover over southeastern United States using TM-calibrated AVHRR data. Pages 1252-1262 in Volume 3 *Proc. Int. Conf. Workshop Global Natural Resour. Monit. Assess.*, ASPSRS, Bethesda, MD.

Izzawati, P. Lewis and J. McMorrow (1998) 3D model simulation of polarimetric radar backscatter and texture of oil-palm plantations. *Proc. IEEE Trans. Geosci & Rem. Sens.*, (IGARSS '98), Jul.6th-10th, Seattle, USA

Kasischke, E. S., L. L Bourgeau-Chavez, Jr., and Haney, E. (1994), Observations on the sensitivity of ERS-1 SAR image intensity to changes in aboveground biomass in young loblolly pine forests. *Int. J. Remote Sens*. 15:3-16

Kasischke, E. S., N. L. Christensen, Jr., and L. L. Bourgeau-Chavez, (1995), Correlating radar Backscatter with components of biomass in loblolly pine forests. *IEEE Trans. Geosci. Remote. Sens*. 33:643-659.

Kasischke, E. S., J. M. Melack, and M. C. Dobson, (1997), The Use of Imaging Radar for ecological Applications – A Review, *Remote Sens. Environ*. 59:141-156.

Kupplich, T.M., C.C. Freitas, and J.V. Soares. (2000). The study of ERS-1 SAR and Landsat TM synergism for land use classification *Remote Sens. Environ., 21(10): 2101-2111.*

Lang, R. H., N. S. Chauhan, K. J. Ranson, and O. Kilic, (1993), Modeling P-Band SAR returns from a Red Pine Stand, *Remote Sens. Environ*. 47:132-141.

Leckie, D.G. (1990). Synergy of Synthetic Aperture Radar and Visible/Infrared Data for Forest Type Discrimination. *Photomgrammetric Engineering and Remote Sensing*. 56: 1237–1246.

Le Toan, T., A. Beaudoin, J. Riom, and D. Guyon, (1992). Relating Forest Biomass to SAR Data. *IEEE Trans. Geosci. Remote. Sens*. 30:403-411.

Lillesand, Thomas M., and Kiefer, Ralph W., Remote Sensing and Image Interpretation, 3rd Ed. Wiley and Sons, Inc. New York, 1994.

Lowe, Roger C. (2002). Masters Thesis, "Predicting Pine Basal Area Using Landsat Thematic Mapper in the Georgia Piedmont". Daniel B. Warnell School of Forest Resources, The University of Georgia.

Luckman, A., J. Baker, T.M. Kuppich, C.C.F Yanasse, and A.C. Fery, 1997. A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments. *Remote Sens. Environ*. 60:1-13.

Neter, J., M.H. Kutner, C.J. Nachtsheim, W. Wasserman. 1996. *Applied Linear Statistical Models*, WCB McGraw Hill. 1408pp.

Phelps, Brian L. (2001). Masters Thesis, "Predicting Pine Basal Area Using LTM data in the Coastal Plain of Georgia". Daniel B. Warnell School of Forest Resources, The University of Georgia.

Pierce, L.E., K.M. Bergen, M.C. Dobson and F.T. Ulaby, 1998. Multispectral land-cover classification using SIR-C/X-SAR imagery, *Remote Sens. Environ*. 64(1):20-33.

Pohl, C., and J.L. Van Genderen, 1998. Multisensor image fusion in remote sensing: concepts, methods, and applications. *Int. J. Remote Sens., 19: 823-854.*

Poso, S., T.Häme, and R. Paananen. 1984. A method of estimating the stand characteristics of a forest compartment using satellite imagery. *Silva Fenn*., 18:261-292.

Ranson, K. and G. Sun, (1993). Northern Forest Classification Using Temporal Multifrequency and Multipolarimetric SAR Images. *Remote Sens. Environ*. 47:142-153.

Ranson, K. J., S. Saatchi, and G. Sun, 1995, Boreal forest ecosystem characterization with SIR-C/X-SAR, *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 867-876.

E. Rignot and R. Chellappa. Segmentation of polarimetric synthetic aperture radar data. *IEEE-IP*, 1(3):281-300, July 1992.

Rignot, E., J. B. Way, C. Williams, L. Viereck, J. Yarie, and K. McDonald, 1994, Radar mapping of aboveground biomass in boreal forests of interior Alaska, *IEEE Trans. Geosci. Remote Sens.*, vol. 32, pp. 1117- 1124.

Sader, S.A. 1987. Forest biomass, canopy structure, and species composition relationship with multipolarization L-band synthetic aperture radar data. *Photogram. Engin. Remote Sens*, 53:193-202.

Shiver, B.D. and B.E. Borders. 1996. *Sampling Techniques for Forest Resource Inventory*. John Wiley & Sons, Inc. 356 p.

Stellingwerf. D. A. and Y. A. Hussin. 1997. *Measurements and Estimation of Forest Stands Parameters Using Remote Sensing*. Ridderprint bv, Ridderkerk. 272p.

Van Zyl, J.J. and C. F. Burnette, 1992, Baysian classification of polarimetric SAR images using adaptive a priori probabilities, *Int. J. Rem. Sens*., vol. 13, pp. 835-840.

Vona, John D. (2000). Masters Thesis, "Modeling Slash Pine using Synthetic Aperture Radar in Georgia's Coastal Plain". Daniel B. Warnell School of Forest Resources, The University of Georgia.

Wang, Y., E. S. Kasischke, J. M. Melack, F. W. Davis, and N. L. Christensen, Jr., 1994, The Effects of changes in Loblolly pine biomass and soil moisture on ERS-1 SAR backscatter, *Reotem. Sens. of Environ*, vol. 49, no. 1, pp. 25-31.

Wang, Y., F. W. Davis, J. M. Melack, E. S. Kasischke, and N. L. Christensen, Jr., 1995, The effects of changes in forest biomass on radar backscatter from tree canopies, *Int. J. Remote Sens..*, vol. 16, no. 3, pp. 503-513.

Wang, Y., J.L. Day, F.W. Davis (1998), Sensitivity of Modeled C- and L-Band Radar Backscatter to Ground Surface Parameters in Loblolly Pine Forest. *Remote Sens. Environ*. 66:331-342.

Whiffen, H., C. Cieszewski, B. Borders, R. Lowe, B. Izlar, M. Zupko, and W. Cooke, *Using 30-Meter Satellite Data to Estimate Basal Areas of Intensively Managed Loblolly Pine stands*. PMRC Technical Report 2000-6. Daniel B. Warnell School of Forest Resources.

Wong, Y. F. and E. C. Posner, 1993. A new clustering algorithm applicable to multispectral and polarimetric SAR images. *IEEE Trans. Geosci. and Remote Sens.*, 31(3), pp. 634-644.

Forest Facts: Economic Impact on Georgia. Georgia Forestry Association.

Accessed May 25, 2001. <http://gfagrow.org./facts.htm.

(www.auslig.gov.au/acres/prod_ser/landdata.htm,

edcwww.crusgs.gov/glis/hyper/guide/landsat_tm). Accessed January 20, 2002

APPENDIX

MODEL ANOVA AND RESIDUAL PLOTS

WHOLE-STAND MODEL FOR PREDICTING LN(VOLUME)

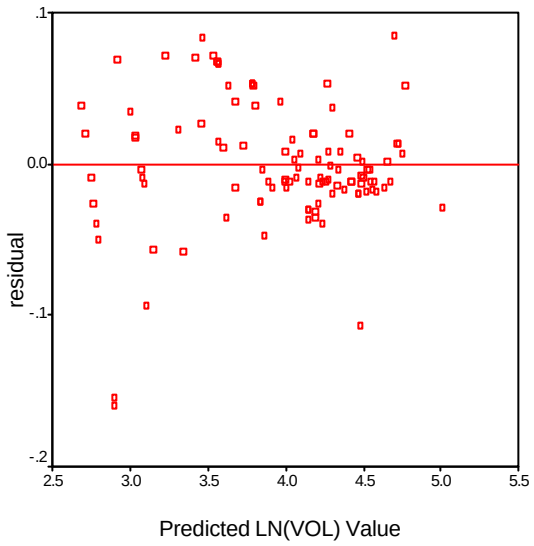
Dependent Variable: LNVOL
Analysis of Variance

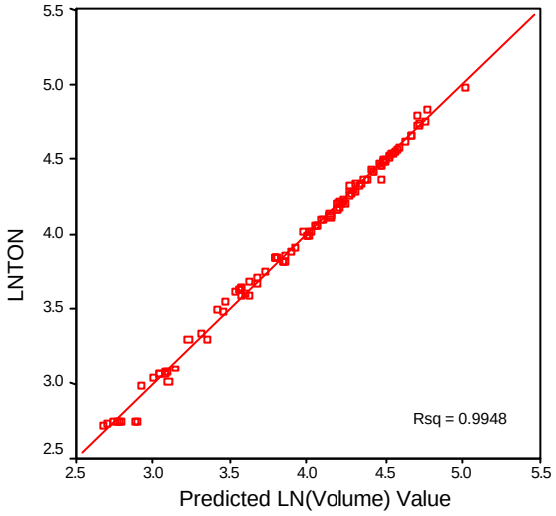
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	35.37986	5.89664	3225.16	<.0001
Error	102	0.18649	0.00183		
Corrected Total	108	35.56635			

Root MSE	0.04276	R-Square	0.9948
Dependent Mean	3.93957	Adj R-Sq	0.9944
Coeff Var	1.08537		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-8.28224	0.67938	-12.19	<.0001
INVAGE	1	77.66729	13.30513	5.84	<.0001
LNH	1	2.01454	0.09720	20.73	<.0001
LNTPA	1	0.75622	0.05690	13.29	<.0001
LNBA_AGE	1	20.80919	0.55822	37.28	<.0001
LNH_AGE	1	-19.62295	1.98162	-9.90	<.0001
LNTPA_AGE	1	-16.05024	1.20143	-13.36	<.0001





WHOLE-STAND MODEL FOR PREDICTING LN(BASAL AREA)

Dependent Variable: LNBA

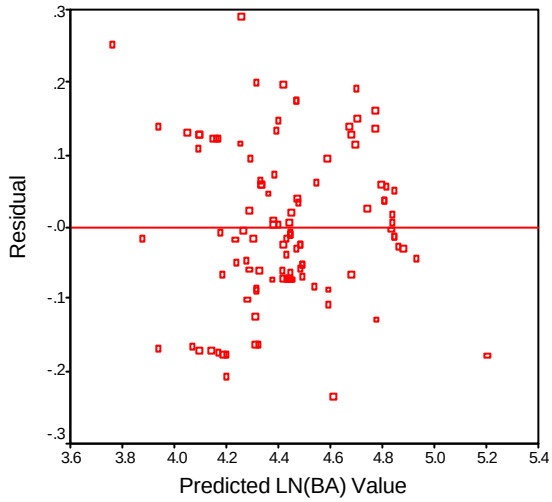
Analysis of Variance

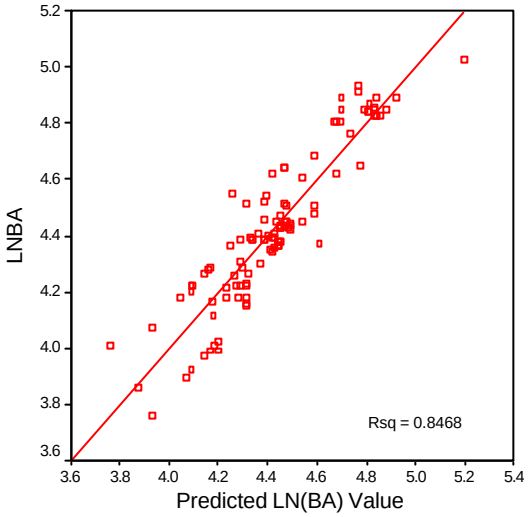
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6.89427	2.29809	193.40	<.0001
Error	105	1.24764	0.01188		
Corrected Total	108	8.14192			

Root MSE	0.10901	R-Square	0.8468
Dependent Mean	4.43406	Adj R-Sq	0.8424
Coeff Var	2.45838		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.00088	0.60461	-4.96	<.0001
INVAGE	1	-8.74191	1.44888	-6.03	<.0001
LNH	1	0.96199	0.11631	8.27	<.0001
LNTPA	1	0.71691	0.03338	21.48	<.0001





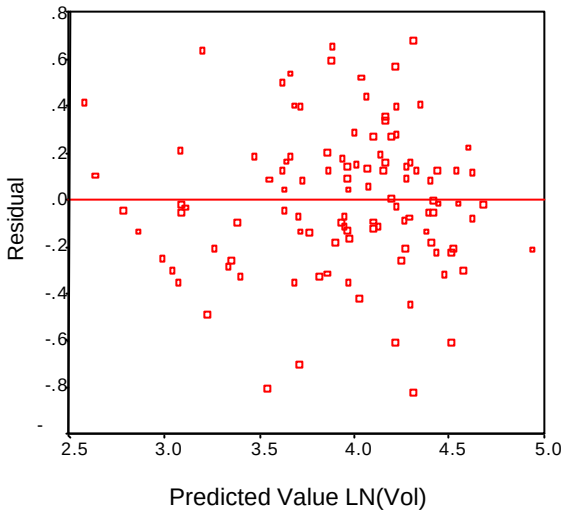
Equation 4.1. Stepwise Regression For Estimating LN(Volume)

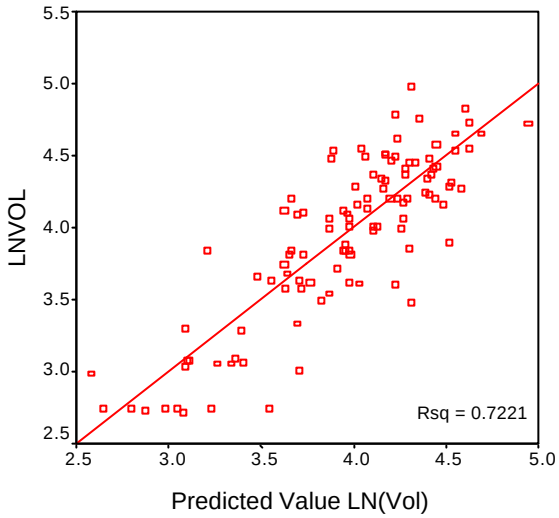
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	25.68229	6.42057	67.56	<.0001
Error	104	9.88405	0.09504		
Corrected Total	108	35.56635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-91.61977	18.89889	2.23361	23.50	<.0001
sarb1b9	50.89930	12.19184	1.65648	17.43	<.0001
sarb9b1	39.01250	7.47290	2.59019	27.25	<.0001
psarb7b1	0.00113	0.00027781	1.56139	16.43	<.0001
psarb7b9	-0.00066050	0.00018236	1.24672	13.12	0.0005

All variables left in the model are significant at the 0.0500 level.





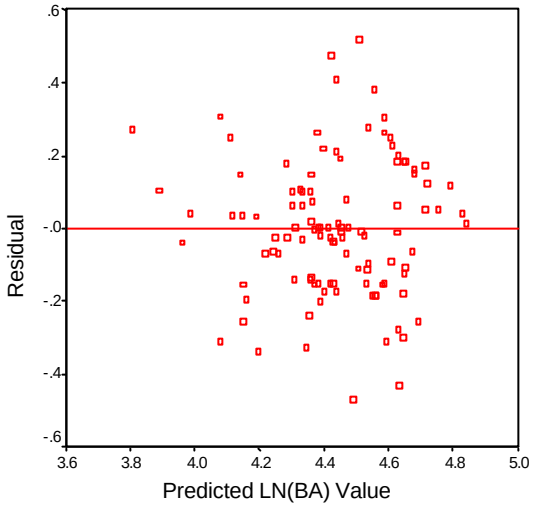
Equation 4.2. Stepwise Regression For Estimating LN(Basal Area)

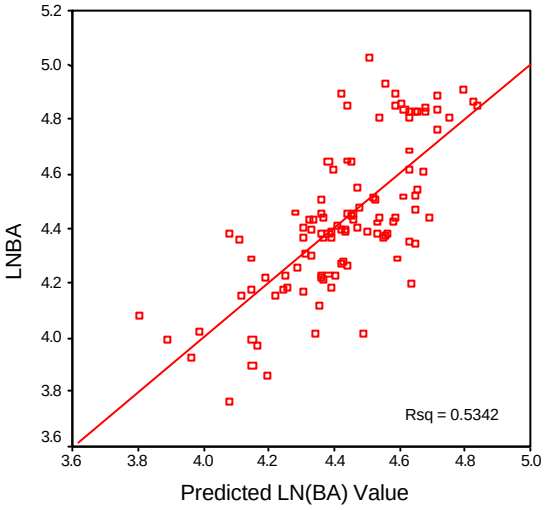
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	4.34930	0.72488	19.50	<.0001
Error	102	3.79262	0.03718		
Corrected Total	108	8.14192			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-20.64548	7.71721	0.26611	7.16	0.0087
SAR_B1_AVG	0.20480	0.07854	0.25282	6.80	0.0105
SAR_B1_VAR	0.00374	0.00122	0.35077	9.43	0.0027
SAR_B5_VAR	-0.00138	0.00031199	0.72539	19.51	<.0001
SAR_B9_AVG	-0.14804	0.05125	0.31023	8.34	0.0047
sarb2b4	8.20616	1.44086	1.20608	32.44	<.0001
sarb9b1	15.22704	5.07826	0.33430	8.99	0.0034

All variables left in the model are significant at the 0.0500 level.





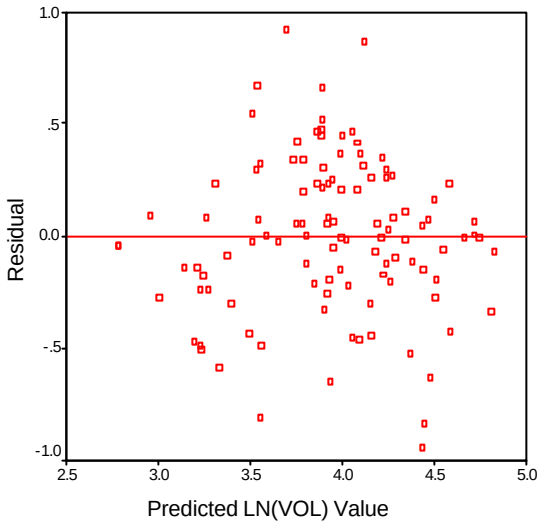
Equation 4.3. LTM, Stepwise Regression For Estimating LN(Volume)

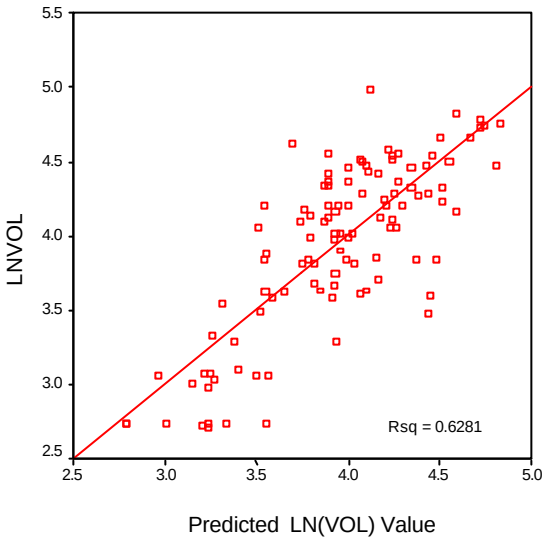
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	22.33978	3.72330	28.71	<.0001
Error	102	13.22657	0.12967		
Corrected Total	108	35.56635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	6.71224	1.98763	1.47880	11.40	0.0010
ON_B4_VAR	-0.00468	0.00127	1.75058	13.50	0.0004
ON_B5_VAR	0.00151	0.00040613	1.79276	13.83	0.0003
OFF_B5_VAR	-0.01371	0.00300	2.70907	20.89	<.0001
onb2b1	-19.71396	3.60920	3.86877	29.84	<.0001
onb3b6	2.27694	0.37230	4.85013	37.40	<.0001
offb3b4	7.89817	1.61927	3.08503	23.79	<.0001

All variables left in the model are significant at the 0.0500 level.





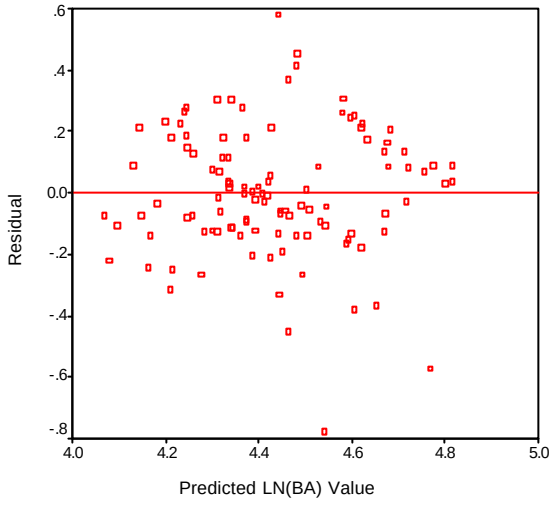
Equation 4.4. LTM, Stepwise Regression For Estimating LN(Basal Area)

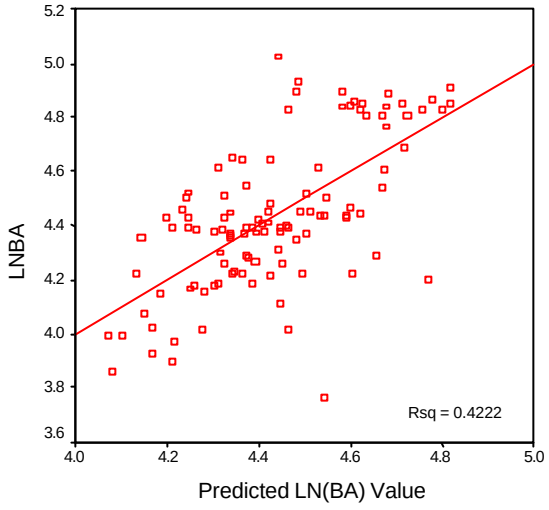
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.43757	1.14586	25.58	<.0001
Error	105	4.70435	0.04480		
Corrected Total	108	8.14192			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-0.04435	1.03192	0.00008276	0.00	0.9658
onb3b5	4.05054	1.13010	0.57557	12.85	0.0005
offb1b5	1.00539	0.24490	0.75509	16.85	<.0001
offb5b4	1.52790	0.61748	0.27432	6.12	0.0149

All variables left in the model are significant at the 0.0500 level.





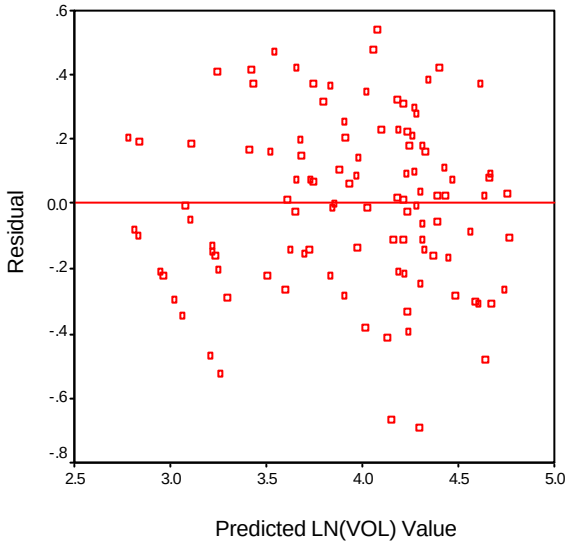
Equation 4.5. SAR and LTM Synergy, Stepwise Regression For Estimating LN(Volume)

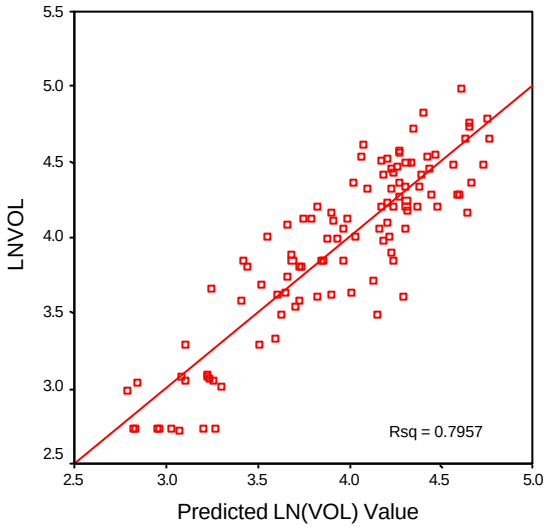
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	28.30146	3.53768	48.70	<.0001
Error	100	7.26489	0.07265		
Corrected Total	108	35.56635			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-11.37544	1.61704	3.59519	49.49	<.0001
SAR_B3_VAR	0.00199	0.00098464	0.29730	4.09	0.0457
SAR_B5_VAR	-0.00312	0.00079194	1.12625	15.50	0.0002
SAR_B6_VAR	0.00968	0.00423	0.38016	5.23	0.0243
SAR_B9_AVG	0.02279	0.00324	3.59530	49.49	<.0001
sarb7b1	1.79389	0.33461	2.08805	28.74	<.0001
onb3b6	2.76590	0.54644	1.86133	25.62	<.0001
onb6b2	2.24641	0.66426	0.83086	11.44	0.0010
offb1b4	2.40768	0.51066	1.61499	22.23	<.0001

All variables left in the model are significant at the 0.0500 level.





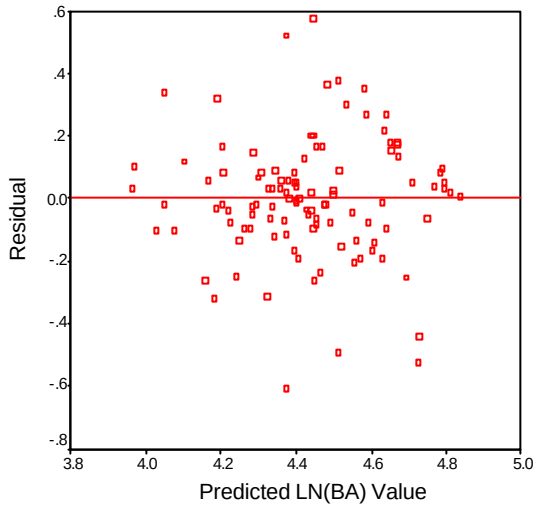
Equation 4.6. SAR and LTM Combined, Stepwise Regression For Estimating LN(Volume)

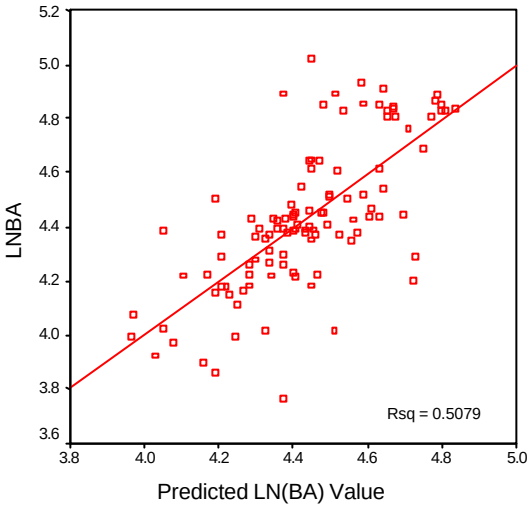
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	4.13532	1.03383	26.84	<.0001
Error	104	4.00660	0.03852		
Corrected Total	108	8.14192			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	6.33662	0.76378	2.65168	68.83	<.0001
SAR_B2_VAR	0.00629	0.00130	0.90434	23.47	<.0001
SAR_B5_VAR	-0.00150	0.00033720	0.75937	19.71	<.0001
sarb4b2	-1.34182	0.28794	0.83664	21.72	<.0001
onb1b5	0.86595	0.22740	0.55867	14.50	0.0002

All variables left in the model are significant at the 0.0500 level.





Equation 4.7. SAR Biologically Motivated Model For Estimating LN(Volume)

Dependent Variable: LNTON

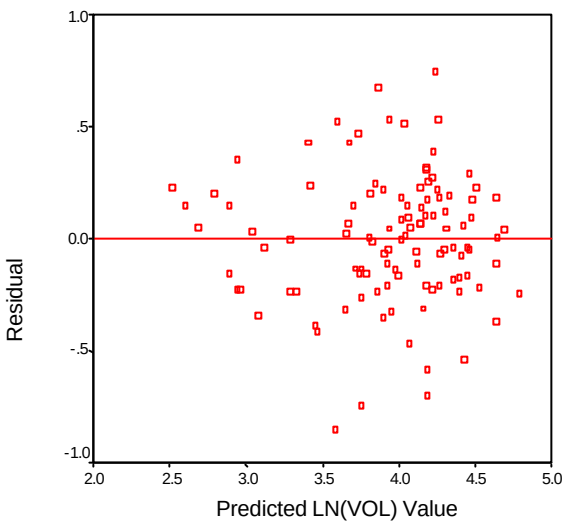
Analysis of Variance

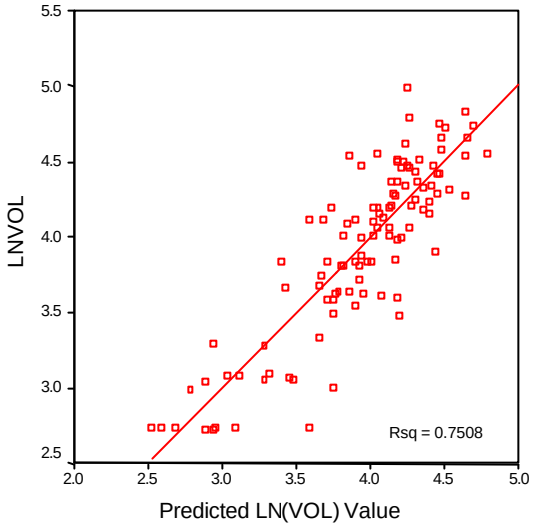
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	26.70312	3.33789	37.66	<.0001
Error	100	8.86323	0.08863		
Corrected Total	108	35.56635			

Root MSE	0.29771	R-Square	0.7508
Dependent Mean	3.93957	Adj R-Sq	0.7309
Coeff Var	7.55697		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	TypeII SS
Intercept	1	-206.97466	37.09313	-5.58	<.0001	1691.70089	2.75956
psarb1b7	1	0.00252	0.00043711	5.77	<.0001	14.46138	2.95221
psarb1b9	1	-0.00375	0.00060445	-6.21	<.0001	4.41903	3.41364
psarb6b9	1	0.00233	0.00037268	6.26	<.0001	3.31084	3.47256
psarb7b9	1	-0.00154	0.00028383	-5.41	<.0001	1.29068	2.59882
sarb9b2	1	31.46985	6.34672	4.96	<.0001	0.03364	2.17913
sarb2b4	1	106.39308	21.07392	5.05	<.0001	0.19706	2.25906
sarb4b9	1	39.78638	7.49638	5.31	<.0001	0.19609	2.49665
sarb9b6	1	56.64722	10.08861	5.61	<.0001	2.79439	2.79439





Equation 4.8. SAR Biologically Motivated Model For Estimating LN(Basal Area)

Dependent Variable: LNBA

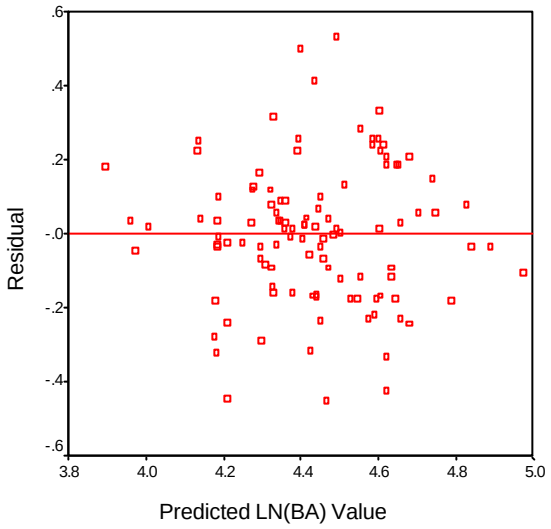
Analysis of Variance

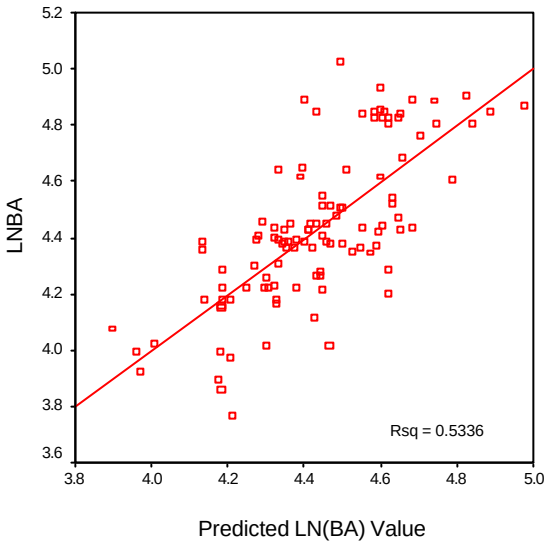
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	4.34427	0.54303	14.30	<.0001
Error	100	3.79764	0.03798		
Corrected Total	108	8.14192			

Root MSE	0.19488	R-Square	0.5336
Dependent Mean	4.43406	Adj R-Sq	0.4963
Coeff Var	4.39496		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	-70.59929	21.62000	-3.27	0.0015	2143.04154	0.40495
SAR_B1_AVG	1	0.38082	0.10685	3.56	0.0006	0.37612	0.48240
SAR_B9_AVG	1	-0.26020	0.06879	-3.78	0.0003	1.86636	0.54342
SAR_B1_VAR	1	0.00331	0.00134	2.47	0.0152	0.06586	0.23179
SAR_B4_VAR	1	-0.00415	0.00126	-3.30	0.0013	0.23845	0.41329
sarb9b2	1	6.85188	2.72361	2.52	0.0135	0.67527	0.24035
sarb2b4	1	30.41594	9.06391	3.36	0.0011	0.56890	0.42765
sarb9b1	1	26.28725	6.89036	3.82	0.0002	0.32503	0.55274
sarb4b9	1	8.36444	3.41165	2.45	0.0160	0.22827	0.22827





Equation 4.9. LTM Biologically Motivated Model For Estimating LN(Volume)

The REG Procedure
Model: MODEL1
Dependent Variable: LNTON

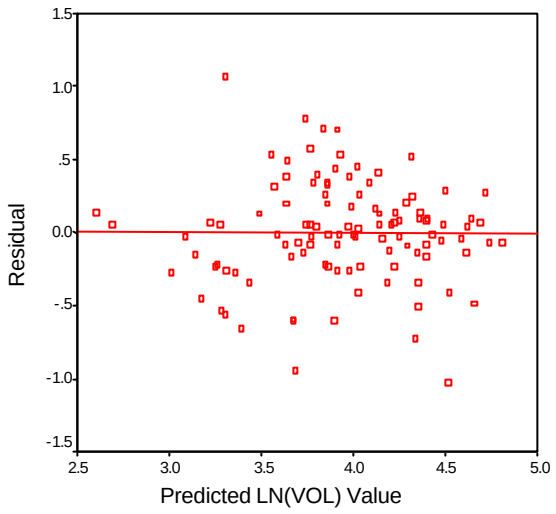
Analysis of Variance

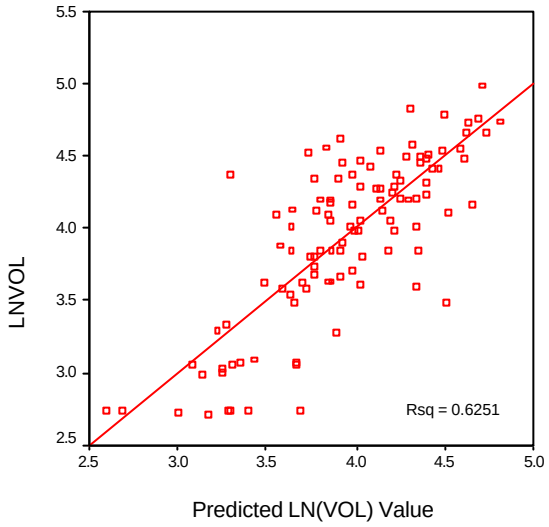
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	22.23274	2.47030	18.34	<.0001
Error	99	13.33360	0.13468		
Corrected Total	108	35.56635			

Root MSE	0.36699	R-Square	0.6251
Dependent Mean	3.93957	Adj R-Sq	0.5910
Coeff Var	9.31553		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	-6.26091	3.73136	-1.68	0.0965	1691.70089	0.37919
ON_B4_VAR	1	-0.00466	0.00135	-3.46	0.0008	5.73031	1.60898
OFF_B3_AVG	1	0.84650	0.30188	2.80	0.0061	6.57493	1.05902
OFF_B5_AVG	1	-0.57265	0.17862	-3.21	0.0018	1.71143	1.38437
OFF_B6_AVG	1	0.78896	0.22966	3.44	0.0009	1.14362	1.58953
onb1b2	1	2.71817	0.82618	3.29	0.0014	2.10237	1.45785
onb3b5	1	7.66831	2.14671	3.57	0.0005	0.72393	1.71856
offb1b4	1	2.88306	0.77383	3.73	0.0003	2.37419	1.86949
offb1b6	1	1.09104	0.29295	3.72	0.0003	0.67957	1.86808
offb3b5	1	-22.10440	7.42897	-2.98	0.0037	1.19237	1.19237





Equation 4.10. LTM Biologically Motivated Model For Estimating LN(Basal Area)

The REG Procedure

Model: MODEL1

Dependent Variable: LNBA

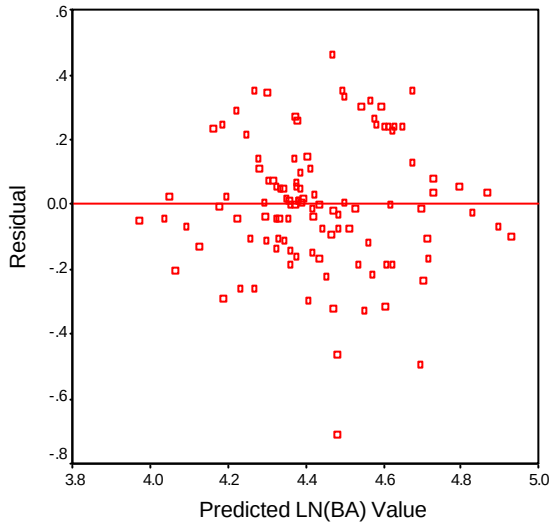
Analysis of Variance

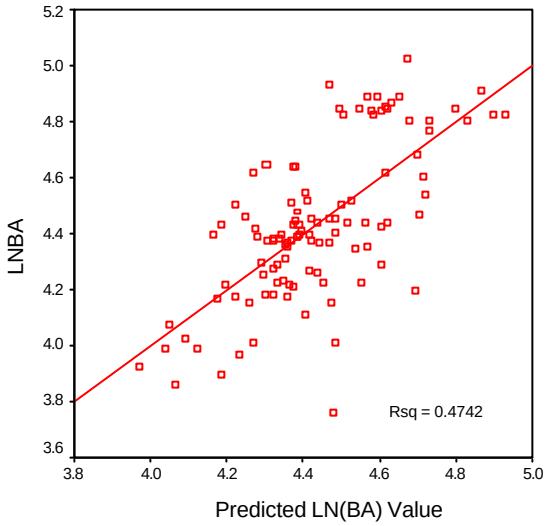
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3.86127	0.77225	18.58	<.0001
Error	103	4.28065	0.04156		
Corrected Total	108	8.14192			

Root MSE	0.20386	R-Square	0.4742
Dependent Mean	4.43406	Adj R-Sq	0.4487
Coeff Var	4.59763		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	4.87362	1.30497	3.73	0.0003	2143.04154	0.57966
OFF_B5_AVG	1	-0.26549	0.06289	-4.22	<.0001	2.60462	0.74057
OFF_B6_AVG	1	0.67567	0.15814	4.27	<.0001	0.02551	0.75872
onb3b5	1	2.13332	0.99982	2.13	0.0352	0.49518	0.18921
offb3b5	1	-8.34879	2.46361	-3.39	0.0010	0.02046	0.47728
offb3b6	**1**	**2.60643**	**0.62817**	**4.15**	**<.0001**	**0.71551**	**0.71551**





Equation 4.11. SAR and LTM Synergy, Biologically Motivated Model

For Estimating LN(Volume)

Dependent Variable: LNTON

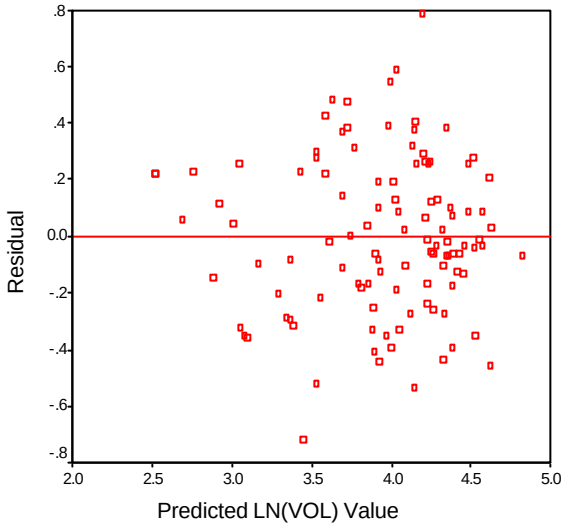
Analysis of Variance

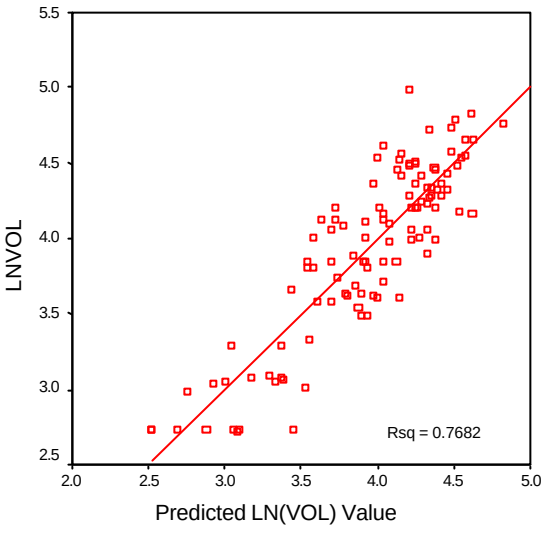
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	27.32282	3.03587	36.46	<.0001
Error	99	8.24352	0.08327		
Corrected Total	108	35.56635			

Root MSE	0.28856	R-Square	0.7682
Dependent Mean	3.93957	Adj R-Sq	0.7472
Coeff Var	7.32471		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	-116.51139	25.27435	-4.61	<.0001	1691.70089	1.76952
SAR_B2_AVG	1	-0.43761	0.11814	-3.70	0.0003	11.76841	1.14256
SAR_B3_AVG	1	1.42307	0.30803	4.62	<.0001	0.07673	1.77730
SAR_B7_AVG	1	0.34717	0.07090	4.90	<.0001	6.69287	1.99663
sarb3b9	1	-91.67849	20.97441	-4.37	<.0001	5.02559	1.59086
sarb2b4	1	84.27483	22.43455	3.76	0.0003	0.03895	1.17500
sarb4b9	1	35.62128	8.93598	3.99	0.0001	0.00044196	1.32316
psarb3b7	1	-0.00406	0.00085547	-4.74	<.0001	2.30144	1.87342
onb3b5	1	6.72527	1.70632	3.94	0.0002	0.53859	1.29354
offb1b4	1	1.87993	0.57835	3.25	0.0016	0.87978	0.87978





Equation 4.12. SAR and LTM Combined, Biologically Motivated Model

For Estimating LN(Basal Area)

Dependent Variable: LNBA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	4.18688	0.59813	15.27	<.0001
Error	101	3.95504	0.03916		
Corrected Total	108	8.14192			

Root MSE	0.19789	R-Square	0.5142
Dependent Mean	4.43406	Adj R-Sq	0.4806
Coeff Var	4.46285		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	1	-23.17789	8.07850	-2.87	0.0050	2143.04154	0.32234
SAR_B1_AVG	1	0.21205	0.08154	2.60	0.0107	0.37612	0.26480
SAR_B9_AVG	1	-0.15141	0.05342	-2.83	0.0055	1.86636	0.31457
SAR_B2_VAR	1	0.00374	0.00157	2.39	0.0186	0.12977	0.22395
SAR_B4_VAR	1	-0.00452	0.00132	-3.42	0.0009	0.16247	0.45848
sarb2b4	1	6.82295	1.53974	4.43	<.0001	1.12517	0.76891
sarb9b1	1	15.43381	5.29599	2.91	0.0044	0.30253	0.33257
onb1b2	1	1.15944	0.48427	2.39	0.0185	0.22446	0.22446