QUANTIFYING FOREST STANDS USING LANDSAT:

AN INITIAL EXPLORATION.

A Dissertation

by

VIRGIL STEVEN LEWIS

Submitted to the Graduate College of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1986

Major Subject: Forestry

QUANTIFYING FOREST STANDS USING LANDSAT:

AN INITIAL EXPLORATION.

A Dissertation

by

VIRGIL STEVEN LEWIS

Approved as to style and content by:

Robert D. Baker (Chairman of Committee)	Robert C. Maggio (Member)
J.J. O'Laughlin (Member)	William B. Smith (Member)
Kenneth L. White (Member)	J. Charles Lee (Head of Department)

May 1986

ABSTRACT

Quantifying Forest Stands using Landsat:

An Initial Exploration. (May 1986)

Virgil Steven Lewis, B.S., Colorado State University

Chairman of Advisory Committee: Dr. Robert D. Baker

Forest stand parameters were shown to be estimable, using data acquired by the Landsat satellite, at a statistically significant level of accuracy. When volume was regressed against spectral signature, the vegetation index which was found to be most significantly associated with volume was the Kauth and Thomas brightness measure. Average stand height was the forest stand parameter which had the highest level of statistical association with spectral signature. The greenness-brightness ratio was the vegetation index for which this association occurred. Average stand diameter, crown diameter, and per acre basal area were also shown to have statistically significant associations with spectral signature, as is represented on Landsat digital data. Canopy closure and stems per acre did not have significant associations with spectral signature.

A computerized classification analysis of spectral signature data for forest stands, based on mensurational characteristics obtained from a cluster analysis of conventional forest inventory data, separated the data into distinct classification categories, rather than grouping the data into a single category which would be expected in a uniform population.

An indirect approach to estimation of volume from spectral signature was done using spectral estimates of canopy closure, crown diameter, and average stand height. These estimates were input into an aerial photo volume table developed for the study area, and a volume estimate obtained. These indirect estimates, along with direct estimates, were not statistically different from the actual volume on a set of stands reserved from the initial model development.

Canopy closure was demonstrated not to be useful in the estimation of field based stand parameters, due to its confounded nature among stems per acre and tree diameter, which is similar to that of basal area.

In general, Landsat data can be used to enumerate forest stands over large areas with a sufficient degree of accuracy as to allow for its use in forest survey, as an example. The potential for substantial cost reductions in the application of Landsat to forest survey, the potential for the incorporation of remote sensing into the technology of geographic information systems as a component of forest survey and other applications are reviewed along with the potential applications of the results of this study with other remote sensing devices.

ACKNOWLEDGEMENTS

I have a huge number of people that I would like to recognize for their contribution to this degree. To begin, I would like to individually acknowledge my advisory committee members.

Without the guidance, suggestions, comments, and well timed criticisms of my chairman, Dr. Robert D. Baker, I could not have even approached the conclusion of this degree. His breadth of experience in the application potential of the results of my research are reflected throughout the text. At several times during my work, when I was at the point of exasperation and even quitting, his encouragement (and unique sense of humor) was a critical factor in my continuing on with work. Dr. Robert C. Maggio gave frequent suggestions on the use of remote sensing techniques, most of which were used in this study. His suggestions for the writing of this paper were extremely helpful in its completion. Dr. Kenneth L. White kept geographical term usage correct for the latter drafts of this paper. He took the time to review my performance in both the preliminary and final examinations for this degree, giving invaluable criticisms and recommendations that I will work to implement from this point of, and will not forget. His sense of humor was also a saving factor throughout this study, especially in concert with Dr's Baker and Maggio.

Dr. William B. Smith, the statistician on my committee, guided me through the intricacies of multivariate statistical techniques that I used in a large proportion of my study, and prevented me from using erroneous assumptions in designing and analyzing the major sections of this study. His comments and criticisms were adroit, well timed and invaluable. Dr. Jay O'Laughlin provided much needed criticisms on my writing style, along with suggestions most useful in their correction. The main point that he, along with Dr. White, emphasized forcefully, and necessarily, was that this study should be written with the non-specialist in mind. Not all readers in the audience will be versed in the jargon associated with remote sensing and biometrics. I would like to also acknowledge my Graduate Council Representative, Dr. Joe S. Ham of the Physics Department, for his suggested explanation of certain phenomena that I found during the course of this study.

I would like to thank Phil Higgins and Steve Grob of the St. Regis Paper Company for their allowing my use of inventory data from their lands in Polk County, Texas. Without this data, I could not have done this study. I would like to thank them for their patience during those long sessions at their copying machine.

I would like to thank the three professors at Colorado State who wrote reference letters for my admission to Texas A&M, Dr's Jay M. Hughes, Philip N. Omi, and Richard D Laven.

Being the last long term inmate at the Asylum, in the old Forestry building, to be let loose on "academic probation", I would be remiss in not recognizing my fellow inmates without whose humor suggestions, and camaraderie, I would not have finished this academic odyssey. Thank you Jerry Tuskan, Dave Cleaves, Dave Patterson, Jim Bailey, Danny Morrow, Tricia Craven, Art Wiselogel, and Tim Kusmertz. To all of the crowd at the new building, Dave Dignum, Doug Wunneburger, Mike Whatley, Clay Martin, Bill "Red" Dempsey, Clay Bassham, the secretaries, and the rest of the host, thank you.

I would like to thank Dr. Baker again, for his persistence in finding funds for my work. He kept this work going with his expert location of monies, since this research was unfunded from the beginning.

I would like to thank several friends in Baker, Oregon who gave encouragement for pursuing this degree. They are Tim Houlihan, Vicki Rockwell, Glen Hedgepeth, Phil Simonski, Katie Holleran, Katie Hogan, Al Colby, Bill Postlethwaite, and Rick Witzke.

Finally, I owe an incalculable debt to my parents, Samuel G. Lewis, who is not here to witness the conclusion of this work, and Marie S. Lewis, without whose love, support, and monetary help, I would have found it infinitely more difficult, if not impossible to finish this degree.

DEDICATION

To the memory of my father, Samuel G. Lewis. My only regret in undertaking this degree is that you were not here to witness its completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I.

INTRODUCTION

The principal thrust of computerized analysis in remote sensing(1) for the assessment of forest resources has occurred in two areas: first, in the development and refinement of computer software for mapping and classification of forest stands, and second, as an initial phase in the design and implementation of forest inventory sampling schemes (Aldrich 1979).

A number of techniques for enhancement of remote sensing data in forest classification and mapping have been developed, however, little work has been done beyond simple forest stand classification and mapping.

Statement of the problem

A need exists to evaluate the behavior of canopy closure and other remotely measurable forest stand parameters across their field measurable counterparts and those stand parameters which have not been studied for their potential as remotely measurable quantities. No such work has been done to analyze these behaviors in East Texas forest lands. Stand parameters estimated from remote sensing data need to be quantified in such a way as to account for combinations of canopy closure, crown diameter, height, and other forest stand parameters. To that end, the behavior of spectral signature, as various stand parameters fluctuate, needs to be examined, and quantified in a manner which is directly applicable to East Texas forests, and can be extended to other forest regions in the United States and beyond.

---

(1) - Appendix D contains definitions of terms, used in this study, which could cause the reader confusion.

Citations, graphics, and text follows the style of: Photogrammetric Engineering and Remote Sensing.

Stand parameter estimation using digital remote sensing data has received scant attention since the launching of Landsat-1 in 1972. Research, prior to and since its launch, investigating utilization of spectral data for inferences concerning forest stand parameters has been restricted to applications related to forest calssification and mapping

Canopy closure, crown diameter, and height are the stand parameters most directly visible from a remote sensing device. They are also surrogates for the individual tree volume equation of diameter squared, multiplied by height, and by stocking for a volume estimate on an areal basis (Husch et al. 1972, Dilworth 1977). Since they are also the common inputs to an aerial photo volume table (APVT) equation, using an APVT approach with digital remote sensing data lends itself to investigation (Aldrich 1979, Avery 1978).

Traditionally used remote sensing data, such as color infrared aerial photography at a scale of 1:15,840, provides a level of information in excess of what an operational forest manager needs. The digital nature of satellite imagery offers potential flexibility in natural resource data management heretofore unrealized. For example, the midcycle update of regional forest surveys can be expidited by incorporating multispectral imagery into the survey process.

Upon combining multispectral methods for estimation of stand parameters with the emerging technology of geographic information systems (GIS), as an example, a substantial proportion of summary statistics that are reported in a forest survey can be obtained using a combined data management system, Separating a major proportion of summary statistic compilation from field data collection should result in significant cost savings, and minimal, if any, degradation in accuracy, bias, and precision in summary statistic reporting.

Finally, the deficiency of previous work in utilizing spectral signature to draw inferences concerning forest stand parameters suggested the potential for work in this area.

Objectives

The overall objective of this study was to answer the question: Is there a quantifiable relationship between spectral signature, as is represented by Landsat digital count data, and timber volume or those stand parameters that can be used to estimate volume through an aerial photo volume table (APVT) approach? Spectral signature can be defined as the observed solar radiation reflected by a forest stand. The three inputs to an APVT are canopy closure, crown diameter, and average stand height. Within the overall objective, four subobjectives were considered:

1. Using statistical methods, quantify the relationship of stand volume versus canopy closure, crown diameter, and average stand height.
2. Using statistical methods, quantify the behavior of canopy closure as basal area, stems per acre, average tree diameter, average height, and crown diameter are varied.
3. Determine if a quantative relationship exists between spectral signature and timber volume.
4. Using statistical methods, quantify the behavior of spectral signature as basal area, stems per acre, average stand height, average tree diameter, canopy closure, and crown diameter are varied.

Chapter II.

LITERATURE REVIEW

Limited research has been done to partition digital data in order to obtain information concerning forest stand parameters. Work which has been done was in conjunction with other objectives, and provided limited results pertinent to this study. The following areas will be considered as a framework for developing and supporting the premise of this study: examples of compendia on remote sensing methodologies, general application of Landsat data to forest research, results of a study investigating the application of Landsat data for estimation of rangeland brush canopy densities, the uses and development of aerial photo volume tables, and the uses of vegetation indices for information extraction from Landsat data.

Syntheses of remote sensing methods

Aldrich (1979) reviewed the applications of remote sensing for wildland resources. He recommended useful journals for particular application areas, and appropriate remote sensing platforms. He pointed out the limitations associated with specific sensors and future potential research areas.

The prominent synthesis of remote sensing techniques, as applied to a wide array of disciplines, is presented in the Manual of Remote Sensing (American Society of Photogrammetry 1983). The two volumes assemble the literature on remote sensing, presenting it in a summarized, concise form. These sources served as guides to those areas of research most closely related to this study, due to the lack of previous work directly applicable to it.

Landsat imagery research applied to forestry

The Landsat satellite, launched in July of 1972, carried two imaging systems of which the Multispectral Scanner (MSS) is the one pertinent to this study. The device is an optical-mechanical scanner, working through the use of rotating mirrors that record the reflected solar radiation from an area of interest. The analog signal is digitized into numerical form, and then transmitted to the ground. Four channels were selected for use in the MSS, corresponding approximately to the wavelengths for which color infrared aerial photography film was designed. Each band is sensitive to the approximate wavelength for which a corresponding film layer is sensitive in color infrared film. Channel 4, designated in this study as MSS4 or B4, corresponds to the green region of visible light, ranging from 0.5 to 0.6 micrometers in wavelength (a micrometer is 1.0×10^{-6} meters). MSS5, or B5, corresponds to red light, ranging from 0.6 to 0.7 micrometers in wavelength. MSS6, or B6, corresponding to near infrared wavelengths, ranges from 0.7 to 0.8 micrometers. MSS7, or B7, also in the near infrared region, ranges from 0.8 to 1.1 micrometers.

The principal area of forestry research involving the use of Landsat imagery concerns the use of classification and mapping of forest lands as either a preliminary stage in forest survey, or as a stand alone product.

Early examples of image classification work will be discussed to illustrate to primary emphasis of Landsat research applied to forestry. Landgrebe et al. (1972) clustered Landsat digital data into 11 descriptive land use/land cover categories and used training sets from each land use/ land cover category to implement a maximum likelihood classification routine. An approximate classification accuracy of 75-80 percent was achieved. Heath (1974) reported results from a clustering routine on the Sam Houston National Forest. An overall classification and mapping accuracy of 74 percent was achieved. Two classification algorithms were compared in a U.S.D.A., Forest Service study in Georgia, Colorado, and South Dakota (Hoffer and LARS Staff 1973, Heller 1975). In a pointwise comparison of forestland classification, the nearest neighbor method achieved an 80 percent accuracy, and the maximum likelihood a 90 percent accuracy rate.

The utility of Landsat as a sole source for land cover classification and mapping has been well documented. Additional useful examples include: Texas Natural Resources Information System (1979), Finley (1979), and Mazade (1981).

Ancillary data, such as aerial photography, multitemporal data, digital terrain maps, ownership and political boundaries have been shown, when combined with Landsat data, to improve classification accuracies. Illustrative examples include: Strahler et al. (1978), Fleming and Hoffer (1979), Hoffer et al. (1979), and Nelson and Hoffer (1979).

A number of results in the area of "pure" classification have been negative. An example of this is found in Colwell and Titus (1976). While attempting to incorporate Landsat data into a forest inventory plan for the Sam Houston National Forest, they found the forestland too homogeneous to allow for stratification into what they termed "surveyable" units. Helms and Shain (1981) pointed out a series of problems associated with the use of remote sensing data as applied to forest inventory. Ground verification and remotely sensed data must be matched temporally to avoid classification confusion resulting from land use/land cover change. Precise ground plot location and registration of the ground plot location on a map using Landsat data is difficult due to its digital nature.

Positive results for the use of Landsat data as an element of forest surveys have been obtained in the Pacific Northwest and California. Harding and Scott (1978) used Landsat land use/land cover classification results as a means to stratify a forested study area as a precursor to optimal plot size determination, photo plot location, and finally field verification. Ownership and political boundaries were digitized and merged with Landsat multispectral data into a combined data file. The output products consisted of color coded maps and tabular statistics arrayed by ownership category.

In Douglas County, Oregon, a "semi-supervised" classification was used to establish nine broad vegetative categories, which, when combined with photo interpretation, ownership, and location description, allowed subdivision by species and density classes. Ground plots were then visited as the final stage in the overall sampling scheme. The output products from the inventory consisted of small scale vegetative maps, intermediate scale "vegetative treatment opportunity" maps, and tabular descriptive statistics (Aggers and Kelley 1976, Oregon State Dept. of Forestry 1978). On the Klamath National Forest, in northern California, Defense Mapping Agency Digital Terrain Map data were merged with Landsat multispectral imagery into a composite file, which was then classified and used as a preliminary stage in a forest inventory operation (Strahler et al. 1978, 1981).

Research directed toward using Landsat multispectral data to draw inferences concerning forest stand parameters, for example cubic foot volumes, has not been attempted beyond classification of canopy closures into descriptive categories. A number of works have shown that open versus closed canopies can be distinguished in ponderosa pine (Pinus ponderosa) due to browning of understory vegetation in autumn (Hoffer and LARS staff 1973, 1975a, 1975b, Strahler et al. 1978, Mead et al. 1979a). Open versus closed canopies were also shown to be visible in Douglas-fir (Pseudotsuga menziesii), white fir (Abies concolor), and red fir (Abies magnifica) forests (Strahler et al. 1978). Williams (1976) showed that three broad categories of canopy closure can be distinguished in the pine-hardwood forests of eastern North Carolina. It was suggested that these differences in canopy closure are a result of an increase in observed infrared radiance from the understory vegetation visible through a forest canopy.

A major point must be remembered when considering previous work using Landsat data to draw inferences concerning forest stand parameters. The results obtained have been primarily qualitative in nature, adjunct to mapping and classification of forests, or as an initial stage in forest survey. To the best knowledge of the author, no previous studies have attempted to directly quantify forest stand parameters based on satellite data.

Rangeland brush canopy results

Boyd (1984) demonstrated a statistically significant relationship between vegetation indices derived from Landsat data and rangeland brush canopy percentages. An example of a vegetation index is the simple ratio of Landsat band 7 divided by band 5. The vegetation index found to be the least sensitive to soil background was the greenness index (Kauth and Thomas 1976).

Since percent canopy cover is a quantity that can be estimated in forests, the implication follows that Landsat can provide data from which this quantity can be estimated. When height and crown diameter are accounted for, it follows that an APVT-like approach can be used with Landsat data. A cautionary result concerning the practical equivalence of certain groups of vegetation indices will be examined in the section of vegetation indices and their use.

Aerial photo volume tables

The majority of previous work on APVT's has concentrated on their development and use in forest inventory. Avery (1958) demonstrated the feasibility of constructing composite aerial volume tables in mixed pine-hardwood standsc compiled from individual tree measurements. A composite aerial volume table can be entered to estimate gross volume for a stand without accounting for the individual species within that stand. The tradeoff here is a larger standard error of the estimate, depending on the accuracy, precision, and bias of the input measurements. Composite tables can provide volume estimates within ten percent of the actual field volume, provided the inputs: canopy closure, crown diameter, and height are measured reliably.

Avery and Myhre (1959) developed a composite table for Southern Arkansas pine-hardwood forests using multiple regression techniques, excluding crown diameter as an independent variable. The local nature of aerial volume tables was re-emphasized as well as the points brought out in Avery (1958).

Pope (1962) presented a systematic technique for use in constructing aerial volume tables using regression methods. Avery (1978) presented a section containing suggestions for deciding whether or not a composite table should be used as opposed to an individual species table. Mead and Smith (1979b) developed a table for use in pine forests infested with southern pine beetle, using multiple regression techniques with height and canopy closure as the independent variables.

The most common form of the aerial volume equations developed in the literature, applied to southern forests, involve height, crown diameter, canopy closure, their squares, and some function of the interaction of the previous terms. The form of the model selected in this study used Avery (1958), Avery and Myhre (1959), and Mead and Smith (1979) as guides for variable selection. A detailed discussion concerning the actual development of the aerial volume table used in this study is presented in the section on APVT model development and analysis.

Vegetation indices

Numerous works, primarily in crop and rangeland analysis, have shown that Landsat data relate closely with vegetation density indicators such as biomass, leaf area index, and percent canopy coverage (Richardson and Wiegand 1977).

The premise for developing a vegetation index is to suppress insofar as possible the soil background effect by employing a transformation to the digital data. Numerous vegetation indices have been developed for seasonal vegetation condition analysis using vegetation amounts and conditions, plant canopy models for yield estimation, or classification techniques for crop and soil condition surveys. A comparative treatise on this subject was done by Lautenschlager and Perry (1981).

Kauth and Thomas (1976), using a sequential orthogonalization of the Gram-Schmidt matrix process (Rao 1973), derived linear combinations reducing the mathematical dimensions of Landsat gray shade data. Using matrix notation and the descriptive terms coined by them, they are as follows:

let SR = (B4, B5, B6, B7) be the vector of digital counts for a Landsat picture element at a given location, then,

"Greenness" (G):

$G = SR'(-.283, -.66, .577, .388)$,

"Brightness" (B):

$B = SR'(.332, .603, .675, .262)$,

"Yellowness" (Y): (1)

$Y = SR'(.723, -.597, .206, -.278)$

"Nonesuch" (N):

$N = SR'(.404, -.039, -.505, .762)$.

Wheeler et al. (1976) and Misra et al. (1977b) used principal component analysis to independently derive linear combinations which reduce the dimensionality of a data set, while preserving as much of the original sample variation as possible. They achieved the following results, using the same matrix notation:

$$\begin{aligned} G_p &= SR'(-.386,\ -.536,\ .535,\ .532), \\ B_p &= SR'(.406,\ .600,\ .645,\ .243), \\ Y_p &= SR'(.723,\ -.597,\ .206,\ -.278) \\ N_p &= SR'(.404,\ -.039,\ -.505,\ .762). \end{aligned} \tag{2}$$

Misra et al. (1977a) also rescaled the principal component analysis into spectral density space, followed by transformation back into digital count space, achieving similar results to that of Kauth and Thomas (1976).

The remarkable similarity between the three works is likely due to the "greenness" component accounting for over 80 percent of the total sample variation, in all cases. Principal component analysis assumes no prior ordering or interpretation of input variables, whereas the Gram-Schmidt orthogonalization indirectly allows some freedom of choice in variable ordering. Another example of the numerous vegetation indices that have been developed appears in Hay et al. (1979).

Aaronson and Davis (1979) showed a significant correlation among the numerous vegetation indices developed up to that time. Correlation coefficients ranged from .81 up to 1.00. Lautenschlager and Perry (1981) performed a cluster analysis on the vegetation indices available and showed that two large clusters existed, having intracluster correlations of greater than .90, and several smaller groupings. The two large clusters were based on: (1), bands 7 and 5, and (2), bands 6 and 5. The second group also includes linear combinations of all four Landsat bands.

Tucker (1979) showed that for monitoring vegetation, combinations of red and infrared wavelengths are superior to green-infrared combinations. The properties of red-infrared combinations are similar for biomass, leaf area index, chlorophyll content, and leaf water content estimation. He showed that the asymptotic properties of the red-infrared combinations are similar in prediction of high amounts of green biomass, and that standing dead vegetation has a linearizing effect on biomass estimation, combined with a corresponding decrease in the coefficient of determination.

In summary, numerous vegetation indices used in monitoring crop and rangeland vegetation condition and production have been developed. Each index displays differing degrees of sensitivity to soil background effects and vegetation condition parameters such as leaf area index. Perry and Lautenschlager (1984) showed, however, that for decision making in the context of classification, vegetation indices within two large categories can be shown to be functionally equivalent. The two groupings consist of those indices based on ratios, and those based on differences, to exploit distinctions between vegetation and soil background effects. A number of vegetation indices, in the works previously described, will be used to statistically relate timber volume to spectral signature. These indices will be formally defined in the methods chapter.

A note on the analogy between volume and biomass

No attempt has apparently been made to use vegetation indices to estimate tree volume, and other stand parameters, from satellite imagery. However, due to the relationship between volume and biomass, and the relationship between certain stand parameters, such as individual tree diameter and height versus volume, the techniques for estimating biomass should be applicable to volume and other stand parameter estimation. Consider the following empirical relationship:

$$\text{Biomass} = \text{volume} \times \text{density per unit volume}. \qquad (3)$$

If an analyst has the appropriate volume to biomass conversions, translation from volume to biomass should be possible.

Chapter III.

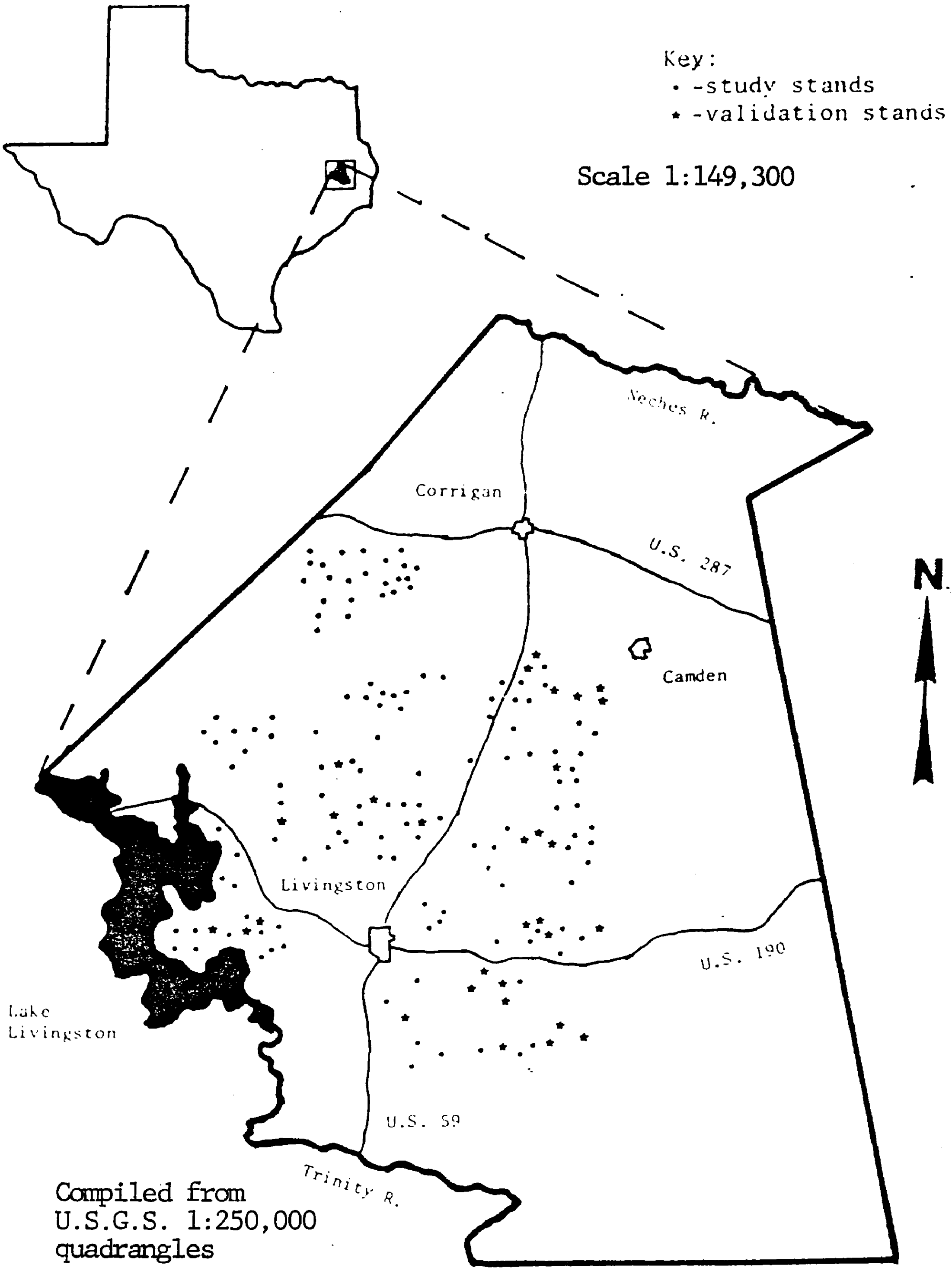
THE STUDY AREA

The study area for this research is located in Polk County, Texas. The area is adjacent to Lake Livingston. Figure 1 indicates the location of those stands selected for analysis and verification.

Land cover in the area is a mosaic of forest, urban, and agricultural areas. The topography of the area is typical of the upper Gulf Coastal plain, consisting of gently rolling plains, interrupted by low lying cuestas and flood plains. Soils within the area are also typical of the East Texas forest region, consisting of mostly sandy soils on uplands, and clays to silty clays in bottomland areas.

Forest types in the area are a mixture of pine, hardwood, and mixed pine-hardwood stands. The prediminant mixture in the area is one of oak and pine, with the predominant pine being loblolly pine (*Pinus taeda*). Slash pine (*Pinus elliotii*), shortleaf pine (*Pinus echinata*), and longleaf pine (*Pinus palustris*) occur sporadically in the area as well. Upland hardwoods are primarily a mixture of oak and hickory (*Quercus spp.-Carya spp.*), with sweetgum (*Liquidambar styraciflua*), hackberry (*Celtis occidentalis*), and hawthorne (*Craetagus spp.*) as examples of sporadically occurring components. Bottomland forests are typified by elm (*Ulmus*), tupelo (*Nyssa spp.*), and ash (*Fraxinus spp.*).

Forest management in the area covers the entire spectrum of potential management scenarios, from intensive pine plantation management conversion practiced by industrial forest managers to custodial forms of management practiced by non-industrial private timber owners such as farm woodlots.



POLK COUNTY, TEXAS

Figure 1. The study area

Chapter IV.

MATERIAL USED AND DATA OBTAINED

The premise for this study, to examine to potential relationships between spectral signature versus timber volume and other forest stand parameters, necessitated the acquisition of data on three levels before proceeding with the quantitative analysis. The three levels were the ground data, aerial photography data, and satellite data.

The ground data, acquired from and with the permission of a cooperating landowner, consisted of stand maps by compartment and stand and stock tables on an individual stand basis (Figure A1, Appendix A, contains an example of both stand maps and it's associated data). The stands averaged between 50 and 110 feet in estimated height, 0 to 300 stems per acre, and 7 to 16 inches in average tree diameter (Tables 1 and 2 present the stand distribution by height, stems per acre, and average tree diameter class).

The aerial photography data were extracted from color infrared aerial reprints, at a scale of 1:20,000, obtained on loan from the Texas Forest Service (TFS). TFS obtained these reprints by permission of Kirby Forest Industries who was the original contractor with Park Aerial Surveys of Lexington, Kentucky.

Satellite data were obtained from two Landsat MSS images, partitioned into subscenes that registered to the quadrangles covering the study area (Table C22, Appendix C). They were obtained from the image library of the Texas Natural Resources Information System at Austin, Texas.

Each of the three levels of data contributed a portion of a single observation used in the quantitative analyses. In addition to the four subobjectives of this study, a separate, detailed narrative on data extraction preceeds the analytical description and follows.

Data extraction

Field data

A critical point must be emphasized at this juncture. Because ground data were obtained from conventional forest inventory techniques, forest

Table 1. Number of Study Stands
by Height Class and Stems per Acre.

Height Class (feet)	Stems per Acre					
	0-60	60-120	120-180	180-240	240-300	more than 300
50-60	1					
60-70	2	4	8	1		1
70-80	4	23	22	2	1	1
80-90	5	20	25	4		
90-100	1	10	8	2		

Table 2. Number of Study Stands by Height Class and Dbh Class.

Height Class (feet)	Dbh Class (inches)					
	7-8	8-9	9-10	10-11	11-12	12-16
50-60		1	1			
60-70	5	9	1	1		
70-80	7	23	16	5	2	2
80-90	1	14	20	10	5	4
90-100			5	8	5	4
100-110			1		1	

management activities are continually occurring on the area. Only one set of aerial photographs at a sufficiently large scale was available, and only two Landsat scenes of adequate quality were available near the time when the field inventory was taken. As a consequence of this, the following criteria were used in study stand selection:

1. Study stands must be reliably located on the aerial photography.
2. Field data must indicate no silvicultural activities since the time of aerial photography.
3. No stands with an extremely elongated shape will be examined. Elongated stands, when examined on digital imagery, present the problem of excessive number of mixed pixels, being both in and out of the stand simultaneously.
4. Field and satellite data will be matched as closely in time as is possible.

Field data, including total basal area, volume, number of stems were recorded on a per acre basis for each individual stand. Average diameter at breast height (denoted in this study as Dbh) for an individual stand was estimated in the following manner:

1. Average basal area per stem was estimated from the quotient of basal area per acre over stems per acre.
2. Average Dbh was then estimated by solving:

$$BA = .005454 \text{ X } (Dbh^2) \tag{4}$$

for Dbh and computing the Dbh estimate.

To determine average height for a stand, a circuitous method was necessary. Individual plot tally records were not available from the cooperating landowner due to the sheer volume of data requested and the fact that a substantial portion of the records had been destroyed after a short archiving period within the company. As a consequence, an estimate was made using site index, recorded for a minimum of three trees per stand. This estimate, less desirable in terms of statistical reliability, was the only one available.

Schumacher and Coile (1960) presented an equation for predicting height based on age and site index for southern pines. A requisite for recording of site trees within a stand during the field inventory was to record site index based upon loblolly pine wherever possible. The presumption was made that successful location of loblolly site trees was achieved, facilitating the use of the Schumacher and Coile equation as shown below:

$$\ln(\text{height}) = \ln(\text{site index}) - 6.258 \times (1/\text{age} - 1/50), \qquad (5)$$

where ln denotes the natural logarithm.

The estimates of average stand diameter and height were then appended to their respective observation containing the rest of the field data.

Aerial photography data

Accompanying stand and stock table data were stand maps for the study area, on a compartmental basis. Study stands were transferred by a Kail reflecting projector to the 7½ minute U.S. Geological Survey quadrangles covering the study area. Photo centers from the coverage on the area were also located on the quadrangles. Acetate overlays were prepared for each photo in the study area containing study stands, according to the criteria discussed earlier for study stand selection.

Field inventory plots were distributed in a systematic manner on a 161 by 322 meter (8 by 16 chain) grid. A dot grid corresponding to the same grid size was constructed for use with aerial photography. The first inventory plot was located in the stand 80 meters south and 80 meters east (4 chains) of the northwest corner of the stand. Circular .4046 hectare (one acre) photo plots were located on the photography in such a manner as to closely approximate the original field inventory as possible. Canopy closure and crown diameter were the two measurements desired at each photo plot. Canopy closure was measured three times, twice by the "ocular" method discussed in Avery (1978), and a third time using a crown closure comparator (Aldrich 1967, Aldrich and Norick 1969). Crown diameter was measured using a Michigan Photo Interpreters Aid. Measurements were taken for the smallest visible, largest visible, and approximate median visible crown for each photo plot. Averages were taken for both canopy closure and crown diameter per stand and then appended to the per stand field data.

A critical point concerning study stand selection is explained below. The selection criteria for study stands combined with the need for data which is temporally matched, resulted in a distinctly nonrandom distribution of study stands within the study area. Summaries of stand volume and stocking, at a coarser level of resolution, were available from the cooperating landowner as an ancillary data source. Each study stand was segregated from the remaining stands and treated as a separate classification category. A comparison between covariance matrices for basal area was performed using the version of Box's M test implemented on the DISCRIM procedure of the Statistical Analysis System (SAS) at the Data Processing Center, Texas A&M University (Mardia et al. 1979, SAS 1982). Appendix B contains details of this test.

The key point to remember concerning this test of sampling adequacy, or "representativeness", is that the study stands appear to be adequately representative of the natural stands in the area. At the .10 level, no significant difference in covariance matrices, based on pine and hardwood basal area, existed.

The significant difference between mean basal area vectors (Table B3, Appendix B) can be attributed to the ongoing conversion of the area from natural stands to a pine plantation form of forest management. This is reflected in the significant difference between the two pine basal area values and the lack of significance in hardwood basal areas. Since the selection of study stands emphasized natural stands, those stands which were pine plantations were pooled into the ramaining category, forcing the significant difference.

Satellite data

As was previously discussed, quadrangle sized subscenes were extracted from the two Landsat scenes available for the study area. Study stands were located on the quadrangles. An example of a subscene, the Blanchard quadrangle, is given in Figure 2. The following procedure was used for extraction of spectral data from each quadrangle.

1. The individual quadrangle was digitized, capturing individual stand boundaries, using the video camera digitizer located at the Remote Sensing Center (RSC) at Texas A&M University.

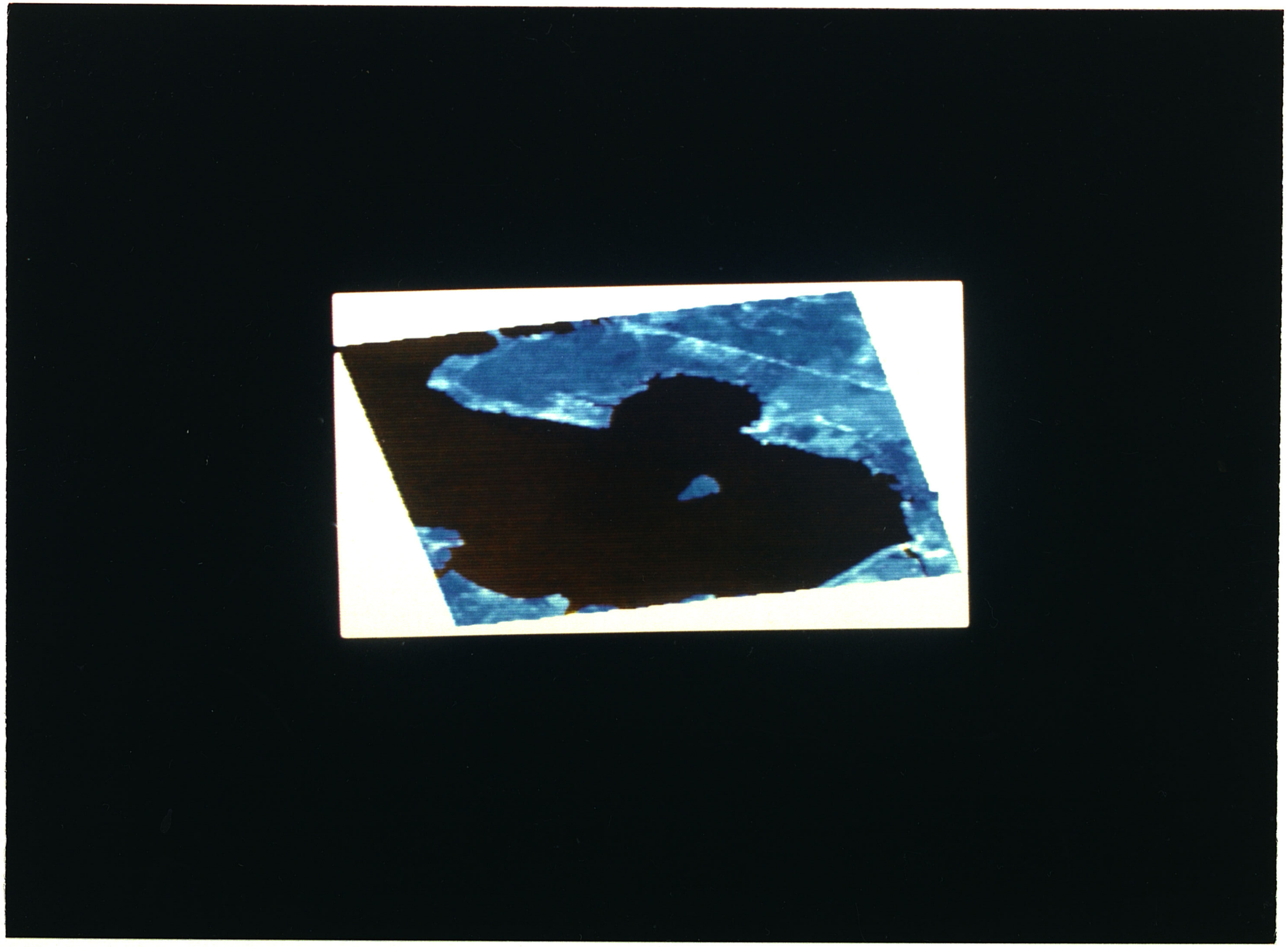


Figure 2. The Blanchard Quadrangle

2. The Landsat subscene was entered into the image processor at the RSC.
3. The quadrangle was geometrically transformed to align it with the subscene, followed by actual registration.
4. Means and standard deviations were recorded, by Landsat band, for each study stand within the quadrangle.
5. These values were appended to the overall data base containing ground and aerial photography data.

An ideal image should possess the property of consistent registering with neighboring images, irrespective of the time of imaging. Due to the orbital drift inherent in the satellite, this ideal is almost never achieved.

A considerable portion of the spectral data was lost due to a 10-15 kilometer gap between the two scenes acquired. No other scenes from the time desired met the restrictions placed on scene quality which were minimum cloud cover and maximum contrast. Seventy nine of the original 146 study stands were able to have spectral data extracted from them.

An ideal image processor should allow extraction of an unlimited number of training sets for classification. This ideal is almost never achieved due to the exorbitant amount of central processing unit (CPU) time involved.

Full covariance matrices for each stand were unobtainable because the routine used in their extraction is designed for training set extraction prior to classification. A maximum of five training sets were premitted by this routine, therefore individual bands had to be extracted using an algorithm which obtained statistics by individual band only.

Chapter V.

METHODS USED

In developing the premise for this study, four analyses, which corresponded to the subobjectives of this study, were developed, followed by a sysnthesis to answer the overall objective question. Initially, an aerial photo volume table (APVT) equation was developed, followed by an examination of the relationships between canopy closure and the stand parameters of interest in this research: stems per acre, average stand Dbh, basal area, and crown diameter.

Spectral analysis consisted of two sections: first, an analysis of the relationships between spectral signature versus volume and those stand parameters needed for input into the APVT equation, and second, an examination of the relationship between spectral signature and those stand parameters not examined with respect to estimating volume.

The development of the APVT equation served two purposes: first, to provide the second step in the indirect volume estimation from spectral signature, and second, to present a volume table for use in East Texas forests. The analysis of canopy closure relationships was for the purpose of determining if ground stand parameters can be estimated using it, and as an indication of its relative contribution to both spectral signature and the development of the aerial photo volume table.

Spectral analysis of volume relationships, central to the overall purposes of this study, was partitioned into three subapproached. First, cluster analysis was done using the inherent properties of stand level data to determine natural "clumpings" within the data, followed by a classification analysis using the spectral signature associated with each stand. Second, volume was regressed against spectral signature, and last, the inputs to the derived APVT equation were regressed against spectral signature, estimated from imagery, and then the estimates of canopy closure, crown diameter, and average stand height were input into the APVT equation to obtain an estimate of volume.

Stand parameters, not associated with volume estimation methods, were regressed against spectral signature to see if complete forest enumeration using remote sensing devices is possible.

This section will be discussed in two parts. First, the sequence in which analyses were done will be outlined, and second, the methods used within was separate analysis.

Analytical sequence

This section will outline the order in which analyses were done. The test for sampling adequacy, or "representativeness", was first (Appendix B). Since no additional samples were taken, the aerial photo volume table was developed for Polk County, Texas. The final equation was held in reserve for use in the indirect portion of the spectral analysis.

Canopy closure relationships with ground level stand parameters was done next. Spectral extimation methods were performed simultaneously, after the analysis of canopy closure relationships. The three methods used were the clustering/classification approach, the direct regression approach, and the indirect regression approach. The indirect regression approach used the aerial photo volume table equation, reserved from the first analysis.

Verification analysis, comparing estimated volumes versus actual field volumes for a set of timber stands reserved from initial analysis, was done following spectral analysis. Finally, spectral estimation of those stand parameters, basal area, average tree diameter, and stems per acre, which were not used to draw inferences concerning timber volume was done using regression methods.

The next section will outline the specific methods used for each of the separate analyses. A brief outline of the statistical methods not commonly used in forestry will be included at the appropriate point in the discussion.

Individual analysis methods

APVT analysis methods

Standard multiple regression techniques were used in formulating a model describing stand volume as a function of height, canopy closure, and crown diameter. Initial model formulation incorporated canopy closure, crown diameter, height, their squares, and all interactions of the above six variables. The use of a quadratic plus interaction model follows similar lines of reasoning as set out by Avery (1958), Avery and Myhre (1959), and Mead and Smith (1979b). In brief, their models considered the strong association between crown diameter and Dbh, using this association in an analogy to the relationship between tree diameter squared multiplied by height (Dbh^2 X HT). The resulting formulation described individual tree volume by terms similar to crown diameter squared times height (CD^2 X HT). On an areal basis, stand canopy closure can then be considered as analogous to stand basal area, expanding the individual tree volume to a per acre basis.

The initial search among the variables just discussed was done using stepwise regression using the MAXR and MINR variable selection criteria (SAS 1982, Neter et al. 1983). Suitable candidates were then screened using both Mallows C_p statistic and R^2 (Hocking 1976, Neter et al. 1983), and those variable combinations making practical sense in terms of the input variables.

Those models which optimized the aforementioned criteria were then run on the REG procedure on the Statistical Analysis System (SAS 1982), estimating regression coefficients and extracting the appropriate diagnostic statistics.

A second iteration employed the use of a logarithmic transformation to volume, and regressing against the independent variables from the selected candidate models. Two reasons effected this decision: first, the tendency of the transformations to mitigate heteroscedasticity in residual behavior (Neter et al. 1983), and second, to evalute its predicitive potential concomitant with the reasoning set out by Mead and Smith (1979b). The logarithmic aerial photo volume table equation was used to develop the table for use in Polk County, Texas.

The third iteration involved two presumptions: first, that canopy closure is the more reliable of the two photo measurements, and second, repeating the same notion with crown diameter. Each model in this case incorporated as independent variables expected crown diameter given canopy closure, or expected canopy closure given crown diameter. The expected variables were estimated as follows:

$$ECD = (554.623 \times (CC/STPA))^{\frac{1}{2}}, \text{ and} \quad (6)$$

$$ECC = .0018 \times CD^2 \times STPA, \text{ where:} \quad (7)$$

ECC = expected canopy closure in percent,

ECD = expected crown diameter in feet,

CC = canopy closure in percent,

CD = crown diameter in feet, and

STPA = stems per acre.

Each of the above formulation alternatives was estimated for each of the candidate models. Final model selection was based on the tradeoff between R^2, multicollinearity effects, "good" diagnostic statistics (specifically the Prediction REsidual Sum of Squares, or PRESS), and its "explainability" in terms of the analogy to conventional volume equations.

Canopy closure analysis

This analysis attempted to examine potential relationships between canopy closure and certain ground stand parameters. The premise for this analysis can be established with an abbreviated and generalized review of stand development over time. As a stand develops, one should expect that the number of stems per acre decreases and that height, Dbh, and basal area increase, implying a relatively constant canopy closure in stands that are unmanaged. Most industrial forestlands in East Texas have had some form of forest management activities performed on them, and as a consequence cannot be strictly considered as unmanaged. The following hypothesis was posed: in managed stands, differential stocking should be detectable through changes in canopy closure.

To that end, two approaches were considered: multiple regression and response surface fitting (Neter et al. 1983, Myers 1976), with canopy closure as the independent variable. The use of response surface analysis facilitates examination of the effect of independent variables on the dependent variable in an optimizing approach with a quadratic and interaction regression.

Spectral data analysis: volume estimation

Three approaches were used for analyzing relationships between timber volume and spectral signature: A) clustering/classification, B) an indirect regression approach, and C) a direct regression approach. Each approach will now be briefly outlined.

A) clustering/classification approach

Two steps were used in this approach: (1), field data clustering to enumerate qualitative categories based on volume, stocking, and height, and (2), following clustering by classification using stand concomitant spectral data to attempt separation of the categories derived from clustering (Mardia et al. 1979). Field data, compiled into total volume, total stocking, and average size summaries were clustered using the FASTCLUS procedure in SAS (SAS 1982). This routine produced cluster summary statistics, and a cluster membership listing by individual stand. Cluster membership listings were not directly available on the other clustering routines available. Clustering was done on both principal components of the raw data and the raw data itself. Cluster membership was then identified for each stand, and an artificial variable identifying cluster membership appended to the overall data set.

The classification step involved utilizing the spectral signature mean vectors as the variable with which the discriminant analysis was done, and the qualitative cluster membership codes as the variable into which the classification was done. Linear and quadratic classification functions were used depending upon the homogeneity of covariance matrices, or variances with single indices. Linear classification functions, used when variances are similar, use a pooled variance or covariance matrix. A quadratic classification function, used when variances or covariance matrices are heterogeneous, employs the within category covariance matrices. The decision to use a linear or quadratic classification function was based on a test for covariance matrix homogeneity.

B) indirect regression approach

The aerial photo volume table equation is brought into play in this approach. The following models were posed for this approach:

CC, CD, or HT = f(SPS), where: (8)

CC = canopy closure in percent,

CD = crown diameter in feet,

HT = average stand height in feet, and

SPS denotes spectral signature.

The premise here is that if one could estimate the three inputs to the aerial photo volume table equation, all that would be necessary is entering the crown diameter, canopy closure, and height estimates into the APVT equation in order to obtain a volume estimate. Univariate regression and canonical correlation (Johnson and Wichern 1982, Mardia et al. 1979) were the two techniques employed here. In addition to using unaltered spectral signature, the following transformations were made: greenness, brightness, the greenness-brightness ratio (Kauth and Thomas 1976), the simple ratios of Landsat band 7 to band 5 and band 6 to band 5, the normalized differences, differences, and difference-difference (Richardson and Wiegand 1977). Formal definitions of these transformations, called vegetation indices, will be presented shortly. In each case, the dependent variables, crown diameter, canopy closure, and average stand height were maintained in an untransformed state due to the time limitations on this study. Recommendations for further analysis are discussed in the summary discussion chapter.

C) direct regression approach

In this case, volume was directly related to spectral signature in a similar fashion as with canopy closure, crown diameter, and average stand height in the indirect approach. The same formulation was used here as with the indirect approach, substituting volume on the left hand side.

Verification analysis

Twenty two stands within the study area were selected and reserved from analysis to test the results obtained in the first spectral data analysis section. Each estimation technique calculated a volume estimate for each stand. The volumes actually occuring in the reserved stands acted as a control against which the estimated volumes could be compared. Each estimation technique was considered as an individual treatment, and the estimated volumes as an observation for that technique. An analysis of variance was performed comparing volumes obtained from estimation using spectral data and volumes actually occuring in the reserved stands.

Spectral data analysis: other stand parameters

Stand parameters not used in volume estimation were regressed against spectral signature data, with spectral data as the independent variable. The models hypothesized for this section are as follows:

Stems per acre, basal area, or Dbh = f(SPS), (9)

where: SPS denotes spectral signature.

In this instance, simple correlation, partial correlation, canonical correlation, and univariate regression were the techniques used to evaluate this set of relationships. In addition to the raw spectral data, the transformations discussed in the previous section and to be defined shortly were also employed. Independent variables were kept in an untransformed state due to the time limitations on this study. As a consequence, recommendations for further analysis are presented in the summary discussion chapter.

A note on the use of vegetation index acronyms

The large number of vegetation indices in the spectral analysis portions of this study could potentially produce confusion for the reader. To that end, throughout the remainder of this study, the following acronyms will refer to their respective vegetation indices, defined below.

R7: The ratio of Landsat B7 divided by B5.

R6: Landsat B6 / B5.

ND7: (B7 - B5) / (B7 + B5).

ND6: (B6 - B5) / (B6 + B5).

D7: B7 - B5.

D6: B6 - B5.

DD: 2 X (B7 + B6) - (B5 + B4).

G: (-.283 X B4) - (.66 X B5) + (.577 X B6) + (.388 X B7). (10)

B: (.332 X B4) + (.603 X B5) + (.675 X B6) + (.262 X B7).

GB: G / B.

PRIN(I): The I'th principal component from the raw data used for the experiment. I = 1, 2, 3, 4.

B4, B5, B6, B7: Individual Landsat MSS bands or channels.

Chapter VI.

RESULTS AND DISCUSSION

The three inputs to the development of the aerial photo volume table equation were canopy closure, crown diameter, and average stand height (CC, CD, HT respectively). Each of these variables are correlated with one another at a high level of significance (Table 3). As a consequence of this phenomenon, a natural property of any regression model fitted to them will be one of some degree of multicollinearity (Neter et al. 1983). Multicollinearity is an unavoidable consequence of using numerous combinations of few original variables. Photographically measured mensurational variables, being few in number available, will nearly always have some degree of multicollinearity associated with them in regression analyses.

Canopy closure, crown diameter, height, their squares, and all possible combinations of these six were used in a stepwise regression to explore for likely candidates for further analysis. A total of twenty three variable combinations were explored (Table 4). The MAXR and MINR options of the STEPWISE procedure on SAS (SAS 1982) were used as the exploration techniques. The background for both exploration criteria, and others, are discussed in the survey article on variable selection in regression analysis by Hocking (1976). Higher power terms and cross-products were not considered because the typical pattern involved with them usually presents bizarre behavior in prediction beyond the range of the data (Pope 1962, Neter et al. 1983).

The nine "best" candidate models, from the available combinations of variables were examined (Table 5). The "best" model choice, when considering an number of potential candidates, involves a considerable degree of subjectivity on the part of an analyst. Two subjective criteria beyond mathematical considerations were used in final model selection.

First, height, crown diameter, and canopy closure were to be included in the final model as separate variables. The second criteria was to include the AB2C term in the final model (see Table 4). The reasons for these two criteria are: (a) to mitigate the multicollinearity resulting

Table 3. Partial correlations among the aerial photo volume table input variables, with the significance probability of the correlation coefficient in the row immediately below the coefficient itself.

		APVT Input Variables		
		CC	CD	HT
	CC	1.0000	.4649	.2459
Prob(r) =		0.0000	.0001	.0027
	CD		1.0000	.2700
Prob(r) =			0.0000	.0009
	HT			1.0000
Prob(r) =				0.0000

Note: CC denotes canopy closure in percent,
CD denotes crown diameter in feet, and
HT denotes average stand height in feet.

Table 4. Definitions of the variables used in building the aerial photo volume table equation with stepwise regression.

HT = height in feet	CC = canopy closure (pct.)
CD = crown diameter in feet	CDCC = CD X CC
HTCC = HT X CC	HTCD = HT X CD
H2 = HT^2	CC2 = CC^2
CD2 = CD^2	H2CC = H2 X CC
HCC2 = HT X CC2	H2CD = H2 X CD
HCD2 = HT X CD2	CD2CC = CD2 X CC
CDCC2 = CD X CC2	ABC = CC X CD X HT
ABC2 = CC X CD X H2	AB2C = CC X CD2 X HT
A2B2C = CC2 X CD2 X HT	A2BC2 = CC2 X CD X H2
AB2C2 = CC X CD2 X H2	A2B2C2 = CC2 X CD2 X H2
A2BC = CC2 X CD X HT	

Table 5. Stepwise aerial photo volume table regression variable selection: results for the first nine model candidates, giving the variables included first, then highlights of diagnostic statistics.

Number of Variables in the Model	"best" Variable Combination
1	H2CC
2	HTCC, H2CC
3	CD2, CD2CC, A2B2C
4	CC, CD, CDCC, AB2C
5	CC, CD, CDCC, HCC2, AB2C
6	HCC2, H2CD, CD2CC, ABC, A2BC2, A2B2C2
7	HT, HCC2, H2CD, CD2CC, AB2C, A2BC2, A2B2C2
8	HT, CD, HCC2, H2CD, CD2CC AB2C, A2BC2, A2B2C2
9	HT, CC, CC2, HTCD, CDCC, H2CD, CD2CC, A2BC2, A2B2C2

Table 5. (cont.) APVT variable selection from stepwise regression: diagnostic highlights.

Number of Variables	R^2	C_p	F, d.f. error
1	.4171	52.48	103.77, 145
2	.4692	37.03	63.64, 144
3	.5008	28.40	47.83, 143
4	.5079	28.03	36.64, 142
5	.5098	29.41	29.33, 141
6	.5276	25.41	26.07, 140
7	.5395	23.41	23.27, 139
8	.5418	24.68	20.40, 138
9	.5517	23.34	18.74, 137

Note: refer to Table 4. for definitions of the variables used in this table.

from the naturally high degree of correlation occuring in crossproducts of the kind used here, and (b) to maintain consistency with the analogy between tree diameter squared multiplied by height and its conocmitant term using crown diameter.

Several patterns of behavior appeared in the stepwise analysis. When the number of variables in the model is increased, the C_p statistic will approach the number of variables in the model, indicating an unbiased model estimate (Hocking 1976, Neter et al. 1983). This particular statistic stabilized in the vicinity of 20 when the number of variables equaled 5, and remained there as additional variables were incorporated into the model. This statistic does not give an indication of multicollinearity. When the number of variables equaled six and beyond, the variance inflation factor (VIF) increased at an almost exponential rate. A definition of VIF can be found in Neter et al. (1983). At 20 variables, the smallest approximately "unbiased" model, the average VIF was 100,000 plus. From a heuristic viewpoint, as the number of variables is increased in an aerial photo volume table model, its interpretation becomes more of a mathematical abstraction rather than one easily used in the field. As a consequence to the above behaviors, an upper limit of 5 independent terms was placed on the final model.

The "optimum" tradeoff between maximization of R^2, a minimum of multicollinearity effects. and having C_p as close to the number of variables as possible, resulted in the following five term model:

$$\text{Volume} = b_0 + b_1 \text{ X CC} + b_2 \text{ X CD} + b_3 \text{ X CDCC} + b_4 \text{ X HCC2} + b_5 \text{ X AB2C}.$$

In view of the subjective criteria previously mentioned, an examination of the HCC2 term is desirable and follows. Height times canopy closure squared can be thought of as analogous to stocking squared times height. The resultant dependent term, if HCC2 is treated separately, would be feet^5, in effect a nonsense parameter. Furthermore, a term that is easily comparable with stocking, CDCC, already exists, implying redundancy. When considering the analogy with tree diameter squared multiplied times height, the slope coefficient for the AB2C term, comparable to individual tree volume, was not significantly different from zero. Upon replacing the HCC2 term with HT, two things occurred. First, the AB2C term had the highest probability of being significantly different from zero among all five term models containing it. Second, inclusion of the simple height term drastically simplifies field interpretation and application.

Some curious aspects regarding the use of APVT's will be considered before discussing the final model in detail. A majority of previous work in the development of APVT's incorporated either field measured height, or ancillary mappings of site index or height as the "vertical" component of table construction. While this is useful in expiditing data collection, a high risk of bias introduction exists and should be compensated for, whether by completely enumerating height in the field or relating field and photo measurements by post-photointerpretative ground verification. Mead and Smith (1979b) restricted themselves to strict field measurements of height as did Avery (1958) and Avery and Myhre (1959). Use of actual field height estimates eliminates additional variation due to photo interpretation, consequently forcing an increase in the precision of the model. It is probably not appropriate to refer to these as true aerial photo volume tables.

Crown diameter as an input to an APVT has either been ignored or measured in the field. Avery (1958) measured crown diameter in the field on the three largest trees in a study stand. Avery and Myhre (1959) discarded crown diameter as having too small an overall contribution to the overall model. No mention was made of average crown diameter or the range of crown diameters within a stand. The possibility that the three largest crown diameters biased the relative contribution of crown diameter towards a smaller value was not examined. Mead and Smith (1979b) did not even consider this in their analysis, presenting only vague reasons for this.

Height estimation, as done for this study, is intermediate in its desirability in terms of enumerating the actual variability we would expect in photographic measurements of height. The "semi-simulated" form of height estimated by Schumacher and Coile (1960) in their site index based height equation is likely to be intermediate in variability between field measured and photo measured height.

Some measure of variability in height estimation is desirable even if for the reason that it can be adjusted for in a means whereby unbiasedness is preserved insofar as possible, for example using covariate analysis.

While the use of either APVT is useful in and of itself, an additional source of prediction error exists due to the quality of the photography used. Scale, tonal, textural, and processing errors can contribute to an excessive error of prediction. Finally, the human factor of training and experience comes into play in contributing to the overall volume estimation error in a multiplicative or even exponential manner.

Several remarks concerning the interpretative quality of the aerial photography used in this study also need to preface detailed discussion of the final model. In each of the cases just discussed, large scale photos, usually at a scale of 1:8,000, were used. The nominal scale of the photography used in this study was 1:20,000. The actual scale varied from 1:20,800 to 1:21,000. The interpretative quality of the photography can be described as exceptionally poor. Tonal, and textural contrast was almost nonexistent. Color quality was poor, dominated by a bizarre orange caste. As a consequence to the poor photo interpretative quality of the available coverage, a considerable reduction in the proportion of variation "explained" by the model (R^2) occurred. The final value of the coefficient of determination was around 0.5.

Models for both volume and the log of volume were chosen (Table 6). The coefficient of determination from these models, given the photographic quality problems, compares favorably with previous work done in southern forests. Avery and Myhre (1959) obtained a coefficient of multiple correlation of about 0.77. This compares well with the multiple correlation for the volume model, .6871, and the log volume model, .6298. The difference between the models in this study and previous models is attributable to the photographic quality, scale, and interpretative problems discussed earlier.

In both models, HT, CDCC, and AB2C were the slope terms which were significantly different from zero, at the .05 level. They can be

Table 6. Aerial photo volume table equations: results for the final models.

Equation form:

$$\text{Volume, } \log(\text{volume}) = b_0 + b_1CD + b_2CC + b_3HT + b_4CDCC + b_5AB2C$$

Regression Statistics Summary

Dependent Variables	$F_{(5,141)}$	Prob(F)	Mean Sq. Error
Volume (cords)	25.221	.0001	50.2311
ln(Volume)	18.546	.0001	0.1810

Dependent Variables	R^2	Coefficient of Variation	PRESS mean
Volume	.47212	36.8043	-0.0029
ln(Volume)	.39677	14.9216	-0.0016

Dependent Variables	PRESS Standard Error	PRESS Variance
Volume	.62343	57.1341
ln(Volume)	.03631	0.1938

Note: see Table 4 for variable definitions

Table 6. (cont.) Aerial photo volume table: final equation results.

Regression Coefficients

Coefficient	Volume	log(Volume)
b_0	12.20876	-0.243064
b_1	-3.13304	-0.010958
b_2	0.20380	0.027639
b_3	0.02111	-0.001572
b_4	0.338764	0.025407
b_5	2.5076×10^{-5}	6.3033×10^{-7}

visualized as height, the stocking surrogate, and the individual tree volume surrogate, respectively. Canopy closure and crown diameter were retained in the final model due to their contribution to the overall models' explanation of variance in both the volume and log(volume) models. It was also done to facilitate comparisons with the expected canopy closure and expected crown diameter alternatives.

The relative degree of imprecision in the final model candidates is reflected in the prediction residual sum of squares (PRESS) statistics. This can be done by comparing the variance of PRESS to mean square error (MSE). The closer the variance of PRESS is to MSE, the implication is one of relative precision in a model. Bias can be qualitatively evaluated by examining the mean of PRESS. The closer the mean of PRESS is to zero, the implication is one of relative unbiasedness in a model. In the case of the final APVT models, both models were relatively imprecise and relative unbiasedness.

Given the photographic quality problems, and time restraints on this study, the model retained for use with the indirect regression portion of the spectral analysis was the volume model. A volume table for the log of volume model was developed for presentation in table form (Table 7).

"Expected" models

Estimates of expected canopy closure (ECC) given crown diameter, and expected crown diameter (ECD) given canopy closure, were made for each study stand using equations (7) and (6), respectively (Tables C1 and C2, Appendix C, present regression analyses for both APVT alternative formulations).

The expected canopy closure model was only a slight improvement over the original model, in terms of R^2. The expected crown diameter model was substantially better than the original model, also in terms of R^2. ECD's relationship to stems per acre and consequently volume's relationship to the number of stems per acre is the likely factor in driving up the precision of the model. However, for the reasons described with respect to previous work on APVT's, this model and the one using ECC cannot be considered as actual aerial photo volume table equations. Nevertheless, the presumption of crown diameter average in a stand, based on the number

Table 7. Aerial photo volume table for Polk County, Texas, using the log volume model.

Crown Diameter = 10 feet

Canopy Closure (percent)	Height (feet)				
	50	60	70	80	90
	(Volume in cords per acre)				
10	2.42	2.41	2.40	2.39	2.38
20	5.27	5.26	5.26	5.27	5.25
30	8.11	8.12	8.12	8.12	8.13
40	10.96	10.97	10.98	10.99	11.00
50	13.81	13.83	13.84	13.86	13.88
60	16.66	16.68	16.71	16.73	16.75
70	19.51	19.54	19.57	19.59	19.62
80	22.36	22.39	22.43	22.46	22.50
90	25.21	25.25	25.29	25.33	25.37
100	28.05	28.10	28.15	28.20	28.24

Crown Diameter = 15 feet

Canopy Closure (percent)	Height (feet)				
	50	60	70	80	90
	(Volume in cords per acre)				
10	3.67	3.67	3.67	3.67	3.67
20	7.83	7.84	7.86	7.87	7.88
30	11.99	12.02	12.04	12.07	12.10
40	16.15	16.19	16.23	16.27	16.31
50	20.30	20.36	20.42	20.47	20.53
60	24.46	24.53	24.60	24.67	24.74
70	28.62	28.71	28.79	28.87	28.96
80	32.78	32.88	32.98	33.07	33.17
90	36.94	37.05	37.16	37.27	37.39
100	41.10	41.22	41.35	41.48	41.60

Table 7. (cont.) Aerial photo volume table for Polk County, Texas.

Crown Diameter = 20 feet					
Canopy Closure (percent)	Height (feet)				
	50	60	70	80	90
	(Volume in cords per acre)				
10	4.94	4.95	4.96	4.97	4.98
20	10.43	10.46	10.50	10.53	10.57
30	15.91	15.97	16.03	16.09	16.15
40	21.39	21.48	21.56	21.65	21.73
50	26.88	26.99	27.10	27.21	27.32
60	32.36	32.50	32.63	32.77	32.90
70	37.85	38.00	38.17	38.33	38.49
80	43.32	43.52	43.71	43.88	44.07
90	48.81	49.02	49.24	49.45	49.66
100	54.29	54.53	54.77	55.00	55.24

Crown Diameter = 25 feet					
Canopy Closure (percent)	Height (feet)				
	50	60	70	80	90
	(Volume in cords per acre)				
10	6.23	6.25	6.28	6.30	6.32
20	13.05	13.12	13.18	13.24	13.30
30	19.88	19.98	20.08	20.19	20.29
40	26.71	26.85	26.99	27.13	27.27
50	33.53	33.71	33.89	34.07	34.25
60	40.35	40.58	40.80	41.02	41.24
70	47.18	47.44	47.70	47.96	48.22
80	54.00	54.30	54.60	54.90	55.20
90	60.83	61.17	61.51	61.85	62.18
100	67.65	68.03	68.41	68.79	69.17

of stems per acre is still a valid one. The average expected crown diameter is larger than measured crown diameter, because as canopy closure tends toward complete closure, individual crowns tend towards ever greater degrees of overlapping with neighboring trees.

In the case of forest stands under intensive management, assuming accuracy of canopy closure measurements, once a stand has been treated, significant radial crown growth occurs before radial bole growth. The tree moves to exploit the newly available canopy space left vacant by harvested trees. The speculative nature of these assertions will need corroboration.

In summary, the development of the APVT for Polk County, Texas, used the basic outline of Pope (1962). The final models compared well with previous work done in southern forests, specifically Avery (1958) and Avery and Myhre (1959). Two alternatives evaluating expected crown diameter and expected canopy closure indicated that canopy closure is the more reliable of the photographically derived measurements from this study. This combined with the precision of the ECD model reinforced Avery and Myhre's assertion that photo measurements must be reliable before aerial photo volume table models will approach conventional volume tables in statistical precision, accuracy, and unbiasedness. An additional indication is that canopy closure can be measured accurately on photos of substantially smaller scale than can crown diameter.

Canopy closure relationships

The hypothesis posed for canopy closure analysis was that in managed forest stands, differential levels of stocking could be estimated using canopy closure.

Initial selection of stand parameters included height, stems per acre, and their crossproduct. Selection of these variables is based on the following considerations. As height increases, an increase in crown diameter and a corresponding increase in canopy closure would be expected. An increase in the number of stems per acre will result in an increase in canopy closure. The crossproduct was included to examine whether or not it showed greater sensitivity to canopy closure rather than the individual stand paremters themselves. Basal area was not used in the first iteration.

A regression predicting canopy closure from height, stems per acre, and their crossproduct (HT X ST) was developed (Table 8). The estimated equation appears to be modestly biased and imprecise, based on the PRESS, prediction residual sum of squares, statistics. One explanation for the low coefficient of determination can be addressed by recalling the discussion concerning the expected canopy closure and crown diameter APVT models. To reiterate, after forest management activities occur within a stand, substantial radial crown growth occurs before appreciable diameter accumulation happens. The tree moves to occupy space vacated by harvested trees.

The crossproduct between height and stems per acre is naturally correlated with the inputs to it to a high degree. As a consequence, multicollinearity is a factor in this model.

Prediction of stand parameters based on canopy closure was the second step in this section. Basal area, average tree diameter and crown diameter were included in this iteration. Each of these stand parameters was regressed against canopy closure as a dependent variable (Table 9).

The closest correspondence can be seen in basal area, which is confounded with tree size and height in a similar manner as is canopy closure. The interaction of height times stems per acre can be considered as similar to height and stems per acre separately. This interaction,

Table 8. Canopy closure regression summary for the model using height, stems per acre, and their crossproduct.

Equation form:

$$CC = b_0 + b_1HT + b_2ST + b_3ST \times HT,$$

where: CC = canopy closure,
ST = stems per acre, and
HT = average stand height.

Regression Statistics Summary Listing

$F_{(3,143 \text{ d.f.})} = 12.971$ Prob(F) = .0001

$R^2 = .2139$ Coefficient of Variation = 11.151

PRESS mean = -0.1018

PRESS variance = 73.2383 Mean Square Error = 67.681

Regression Coefficient Estimates

Coefficient	Estimate	t statistic	Prob(t)
b_0	31.0247	3.109	.0023
b_1	0.4641	3.594	.0004
b_2	0.2126	2.871	.0047
b_3	-0.0021	-2.190	.0030

Table 9. Regression estimation of stand parameters using canopy closure as the independent variable.

Equation form:

$$\text{Parameter} = b_0 + b_1 CC,$$

where: CC = canopy closure.

Intercept Coefficients

Parameter	b_0	t statistic	Prob(t)
Stems/acre	-44.0803	-1.184	.2383
Height	58.6660	9.136	.0001
Stems X Height	-4335.1700	-1.447	.1501
DBH	8.7790	9.303	.0001
Crown Diameter	11.9390	10.401	.0001
Basal Area	-17.3070	-1.153	.2509

Slope Coefficients

Parameter	b_1	t statistic	Prob(t)
Stems/acre	2.3201	4.633	.0001
Height	0.2948	3.413	.0008
Stems X Height	197.9910	4.913	.0001
DBH	0.0105	0.824	.4115
Crown Diameter	0.0144	0.934	.3521
Basal Area	1.0617	5.256	.0001

Parameter	R^2
Stems/acre	.1290
Height	.0744
Stems X Height	.1427
DBH	.0047
Crown Diameter	.0059
Basal Area	.1650

Note: DBH denotes average tree diameter.

although not additive, appears to be only slightly more sensitive to canopy closure and consequently redundant.

Dbh and crown diameter apparently do not influence canopy closure. The study area consists of managed forest stands, having silvicultural activities occuring on them at some point in time. As a consequence, the radial crown growth phenomenon, discussed in the previous section, can be applied here as well.

The relative contribution of individual stand parameters to canopy closure was examined through a quadratic regression alternative to analysis of variance, termed a response surface analysis (Myers 1976, see Table C3, Appendix C, for the results of this analysis). Each stand parameter was regressed against canopy closure, linear and quadratic contributions to each model extracted, and overall significances examined.

Stems per acre, height, and basal area are again the field measured variables showing appreciable differences across levels of canopy closure. The relatively low, respective, coefficients of determination are likely due to the accelerated radial crown growth phenomenon discussed earlier.

The multivariate test, canonical correlations (Table C4, Appendix C), corroborated the previous two analyses. In both cases, an extremely significant, on the order of 10^{-6}, test statistics for testing for zero correlation were found. A relatively low canonical R^2 was also found. For stems per acre, height, and their crossproduct, the standardized canonical coefficient vector represented a contrast between the two untransformed variables and their crossproduct. It alludes to the point, discussed earlier, concerning the redundancy of height times stems. The interpretation of canonical vectors involves a substantial degree of subjectivity, so the coefficient vectors should be considered under a caveat emptor.

The largest canonical coefficients, or loadings, are on stems per acre, height, Dbh, and with the opposite sign, basal area for the second canonical analysis (also on Table C4, Appendix C). Three points need emphasis here: first, the most influential variables appear to be stems per acre, Dbh, and height; second, height follows Dbh in loading magnitude, and third, basal area appears to contrast stems per acre and Dbh, which is analogous to stocking as represented by basal area.

Recapitulating, a relationship appears to exist between canopy closure and stems per acre, height, and basal area. Basal area appears to be most closely associated with canopy closure, however basal area is confounded among stems per acre and individual tree size in a similar fashion as canopy closure. This indicates that the two stand parameters cannot be remotely used to infer stocking and volume on the ground. Disparate stocking situations can result in the same value for basal area and canopy closure. Crown diameter tends to be biased towards smaller size values, as canopies tend towards complete closure. The subjective nature of canopy closure measurements off of aerial photography using visual interpretation, in addition to its high degree of uniformity over various stocking situations, biasing towards higher values because of potential confusion of understory vegetation as overstory, renders it marginally useful in drawing inferences concerning ground stand parameters.

Spectral analysis

Correlations

Coefficients of simple correlation between spectral signature, including raw data and vegetation indices, and the field and aerial data were obtained (Table 10). On the basis of simple correlation, crown diameter, volume, height, basal area, and tree diameter have a predominant number of significant correlations. In light of the significant correlations among Landsat bands (see page 68), and among the stand variables, simple correlations, although significant on the surface, are only useful in pairwise comparisons. The majority of regression models used in this analysis consisted of pairwise analyses.

Partial correlation coefficients, correlations between two variables adjusting for all other variables in a data set, give an indication of the association between variables in a multivariate sense (Table 11). Given the coarse resolution of Landsat data and its inherent noise, and the additional variation in the data due to sensor system and stmospheric noise, highly significant partial correlations were not expected. Any significance level of .20 or smaller was deemed to be signficant in the context of this study.

A number of candidates appeared, even so far as to be considered as significant in the "classical" statistical sense. The best candidates for crown diameter were: band 4 (B4), band 6 (B6), band 7 (B7), brightness (B), greenness (G), the difference-difference (DD), the difference-7 (D7), and the difference-6 (D6). The raw data could be considered as a function of the amount of vegetation being viewed by the sensor, and what portion is actually tree crowns. A similar argument can be posed for the three difference indices (D7, D6, DD), and greenness. Brightness, which would be expected to be negatively associated with crown diameter having more of the soil component intercepted by foliage, has a significant positive association. One potential, and somewhat speculative, explanation can be examined by considering the foliage density within a crown. In larger crowns, the relative foliage density, as is being viewed by the sensor, is smaller due to the larger crown volume available,

Table 10. Simple correlations between spectral data and stand parameters, with the significance probability of the correlation in the row immediately below the correlation coefficient.

Vegetation Index	Stand Parameter			
	CD	CC	Vol.	Height
B4	-.48654	.07156	-.52188	-.70482
	.0001	.5308	.0001	.0001
B5	-.45785	.07412	-.50589	-.68941
	.0001	.5162	.0001	.0001
B6	-.40318	.03968	-.48507	-.68820
	.0001	.7284	.0001	.0001
B7	-.08843	-.04339	-.29214	-.28012
	.4383	.7042	.0090	.0100
R7	.46500	-.08349	.48473	.66214
	.0001	.4649	.0001	.0001
R6	.43306	-.07715	.48154	.63302
	.0001	.4992	.0001	.0001
ND6	.44698	-.09136	.48215	.65383
	.0001	.4233	.0001	.0001
ND7	.48200	-.09765	.48884	.68570
	.0001	.3919	.0001	.0001
D7	.49051	-.12235	.39878	.62168
	.0001	.2828	.0003	.0001
D6	-.01788	-.07561	-.13723	-.10760
	.8757	.5078	.2278	.3452
DD	.03933	-.07257	-.14920	-.09087
	.7377	.5252	.1894	.4258
G	.41604	-.12461	.29252	.48634
	.0002	.2739	.0089	.0001
B	-.43012	.05258	-.50474	-.66039
	.0001	.6453	.0001	.0001
GB	.48628	-.09477	.49487	.69143
	.0001	.4061	.0001	.0001

Note: See page 29 for vegetation index definitions.

Table 10 (cont.)

Vegetation Index	Stand Parameter		
	Stems	BA	DBH
B4	-.06849 .5486	-.42512 .0001	-.51154 .0001
B5	-.04656 .6837	-.40957 .0001	-.52528 .0001
B6	-.10320 .3654	-.39378 .0003	-.40201 .0001
B7	-.14316 .2082	-.24217 .0322	.09466 .4076
R7	.03303 .7726	.39369 .0003	.54561 .0001
R6	.02120 .8529	.38395 .0005	.56036 .0001
ND6	.00696 .9515	.37787 .0006	.57337 .0001
ND7	.01643 .8857	.38854 .0004	.56401 .0001
D7	-.05048 .6586	.31228 .0051	.56672 .0001
D6	-.19383 .0870	-.14278 .2094	.16430 .1502
DD	-.15193 .1813	-.13196 .2463	.09178 .4211
G	-.09594 .4003	.22131 .0500	.52012 .0001
B	-.08488 .4570	-.40977 .0001	-.45967 .0001
GB	.02469 .8290	.39833 .0003	.56246 .0001

Note: CC = canopy closure, CD = crown diameter, BA = basal area, DBH = average tree diameter.

Table 11. Partial correlations between spectral data and stand parameters, with the significance probability of the correlation in the row immediately below the correlation coefficient.

Vegetation Index	Stand Parameter			
	CD	CC	Vol.	Height
B4	.22568	-.08137	-.07515	-.06100
	.0470	.4788	.5132	.5957
B5	.13052	-.01651	-.05710	-.00747
	.2547	.8859	.6195	.9482
B6	.19280	-.08971	-.03631	.08850
	.0908	.4441	.5723	.4410
B7	.20279	-.09339	-.06557	-.09010
	.0750	.4161	.5684	.4327
R7	-.02788	-.01855	.03430	.01361
	.8086	.8719	.7656	.9059
R6	-.04573	-.00946	.06815	.01708
	.8086	.9345	.5533	.8820
ND6	-.03485	-.03807	.05423	.04687
	.7620	.7372	.6484	.6836
ND7	-.00507	-.05717	.01291	.06002
	.9649	.6191	.9107	.6017
D7	.14591	-.09418	-.03797	.10534
	.2024	.4121	.7414	.3587
D6	.14404	-.10155	-.000002	.12232
	.2083	.3764	.9999	.2860
DD	.18284	-.09338	-.04247	.09778
	.1091	.4161	.7120	.3944
G	.15051	-.09338	-.02238	.10753
	.1884	.4161	.8458	.3487
B	.20168	-.09398	-.05652	.06862
	.0766	.4131	.6231	.5505
GB	-.01583	-.05136	.02524	.04916
	.8906	.6552	.8329	.6691

Note: See page 29 for vegetation index definitions, and CD = crown diameter, CC = canopy closure, BA = basal area, and DBH = tree diameter.

Table 11 (cont.)

Vegetation Index	Stand Parameter		
	Stems	BA	DBH
B4	-.04154 .7180	-.07491 .5145	-.11842 .3018
B5	.04689 .6835	-.02060 .8579	-.17728 .1205
B6	-.11039 .3360	-.04058 .7243	.15841 .1660
B7	-.12858 .2619	-.05566 .6284	.18441 .1060
R7	-.06448 .5449	.02223 .8468	.24868 .0281
R6	-.07143 .5343	.03751 .7444	.28873 .0105
ND6	-.11017 .3269	.01092 .9244	.32079 .0042
ND7	-.12039 .2830	-.01436 .9007	.30820 .0060
D7	-.17273 .1305	-.04947 .6671	.31581 .0049
D6	-.18391 .1070	-.03605 .7540	.35566 .0014
DD	-.14170 .2159	-.04489 .6963	.23964 .0346
G	-.16670 .1447	-.03876 .7362	.31848 .0045
B	-.07527 .5125	-.04594 .6896	.05995 .6021
GB	-.11456 .3179	.00090 .9994	.32193 .0040

allowing a larger soil reflectance component to be seen.

Canopy closure had no candidates, among partial correlations, which qualified as significant. A likely cause for this is understory vegetation masking itself as overstory vegetation. This will be referred to as the "disguise" effect.

Volume and height had no candidates significantly different from zero among partial correlations. Apparently associations with spectral signature, in a multivariate sense, cannot be drawn. Numerous possibilities exist, however, among the pairwise simple correlations. Stems per acre has a few, marginal, candidates qualifying as significant, including: greenness, difference-7, and difference-6. All three reflect the increased number of crowns visible as number of stems increase.

Basal area is too mixed among stems per acre and individual tree size to show any significance in partial correlations. This is a reflection of the "disguise" effect. Tree diameter, with the exception of band 4 (B4) and brightness (B), is significantly associated with all vegetation indices. Either this is due to a strong relationship between crown and tree diameter, or is an anomaly of this individual data set.

In summary, a number of vegetation indices, including the raw Landsat bands, are significantly associated with the stand parameters of interest in this study, with useful associations being defined as a significance probability of 0.2 or smaller. As will be shown in a later discussion, volume, when fitted to all four bands in a multiple regression, showed a significant association with spectral data. Either the relationship of volume to all four bands is through an indirect interaction of other stand parameters, or is unique to this data set.

An overriding theme, which will be repeatedly emphasized throughout the remainder of this study, is that corroboration of these results on other data sets, from other locations and other imagery, must preface application of these results to operational forest management.

Clustering/classification approach

This approach utilized the stand data, clustering them into groups based on natural clumpings within the data itself. Each group derived from the clustering was used as a category around which classification and separation analyses were done using stand concomitant spectral data.

The methodology behind clustering will be briefly outlined first. A data point is compared with the other points in the data set. A criteria of "similarity" is developed for the data points within the data set. If the data point is closest to a certain other point, in terms of the measure of similarity, the two points are grouped into a cluster. The next data point is then compared with the cluster just developed, and the remaining data points. The grouping process is repeated until the resultant clusters are no longer sufficiently similar to group, or an arbitrary stopping point is reached. This approach is known as an agglomerative one. A reverse argument, splitting the entire data set, which is considered as a single cluster, on the basis of a measure of "dissimilarity" is also possible. This approach is called a divisive one. Discussions of the development and uses of clustering can be found in Johnson and Wichern (1982), and Mardia et al. (1979).

The algorithm used in this clustering, the FASTCLUS procedure on SAS (SAS 1982), develops clusters by sorting on the nearest centroid using Euclidean distances. The development of this algorithm is discussed in Anderberg (1973).

Since a considerable amount of data was lost due to a gap between the Landsat scenes obtained for this study, the actual classification was done on the reduced data set of 79 stands. Two cluster types were extracted: first, based on the field data of interest in this study, average Dbh, volume in cords per acre, stems per acre, and basal area in square feet per acre, referred to as cluster type 2, and second, on the principal component loadings of the field data variables listed above, referred to as cluster type 1. Principal component analysis rotates the axes of the raw data, extracting uncorrelated linear combinations from the pnput data (Tables C5 and C6, Appendix C). The actual field data means from the reduced data set, extracted after the loss of the field data in the gap between Landsat scenes, were compiled for both cluster type 1 (Table 12), and cluster type 2 (Table 13). These means provided an indication of the stand conditions found within each cluster derived from the complete data and summarized for the reduced data set.

An approximate stand description, based on the cluster means for the reduced data set follows. For cluster type 1, clusters 1 and 10 represent heavily stocked small poletimber stands, differentiated by height. A definition for adequate stocking which will be used in this study can be defined as that combination of average tree diameter and number of stems per acre such that the resultant basal area will be 80 square feet per acre. Cluster number 3 represents a slightly understocked large pole timber stand. Cluster 9 represents a more severe degree of understocking versus cluster 3. Clusters 6 and 7 represent differential stocking in larger saw timber stands. Poletimber is defined as trees with a diameter of between 5.0 and 10.9 inches, and saw timber as trees of 11.0 inches in diameter or greater. The net breakdown based on the reduced data set is one of relative stocking, by relative size, from understocked to either adequately or overstocked.

Cluster type 2 breaks down into two relative size classes: small saw timber and small poletimber. Further cluster division is based on stocking in an almost linear trend. The small saw timber, in order of increasing stocking, are cluster numbers 5, 8, 3, 9, and 6, and the small poletimber order is cluster numbers 10, 1, and 7.

Since each of the input stand variables are correlated with one another, cluster type 1, the principal component loading scores, is a better indication of how stands group among themselves. This is because principal components are uncorrelated among one another, and clustering statistics are reliable for interpretation whereas correlated clusters are not reliable (Johnson and Wichern 1982, Mardia et al., Anderberg 1973).

Each vegetation index, the four raw Landsat bands, and the first principal component extracted from the raw Landsat data, was used in the classification analysis separately, by cluster type. A brief outline of the principles of classification follows.

Suppose there are a series of categories into which the observations from a data set containing the variables of interest are to be classified. The purpose is to find a rule, based upon knowledge of the given data set,

Table 12. Stand parameter means for cluster type 1, by cluster number.

Cluster Number	Means of Stand Parameters			
	Volume	DBH	Height	Stems
1	40.22	7.17	69.14	535.45
3	15.95	8.09	71.48	165.77
4	21.13	11.66	83.80	80.89
6	27.91	9.77	87.88	157.06
7	14.24	9.17	80.86	104.24
8	50.76	11.46	98.64	178.28
9	9.76	8.69	71.00	85.68
10	30.38	7.76	78.85	319.65

Table 13. Stand parameter means for cluster type 2, by cluster number.

Cluster Number	Means of Stand Parameters			
	Volume	DBH	Height	Stems
1	30.38	7.76	78.85	319.65
3	17.88	9.27	79.31	120.22
5	7.11	9.20	74.11	47.44
6	33.12	9.42	82.93	203.14
7	40.22	7.17	69.14	535.45
8	13.47	9.47	78.05	90.77
9	21.93	8.80	79.23	163.57
10	15.99	7.23	64.71	222.01

Note: Volume is in cords per acre, DBH is average tree diameter in inches, average height in feet, and stems is the number of stems per acre.

to place these observations into the categories of interest. An observation is assigned to a category if the probability of its membership in that category is greater than the probability of its membership in any other category. The general development of classification, based upon the likelihood ratio principle, is, with a multivariate normal distributional assumption, an analog and extension of the conventional t-test and analysis of variance. For more complete introductions to the development of classification, the reader is referred to Johnson and Wichern (1982), Mardia et al. (1979), Lachenbruch (1975, 1979), and Kshirsagar (1972).

Classification, using a multivariate normal distributional assumption, develops discriminant rules dependent upon the homogeneity of covariance matrices. To this end, the application of a test of covariance matrix homogeneity was performed using the asymptotic approximation of the likelihood ratio test for it (Mardia et al. 1979, Muirhead 1982). This test uses the matural logarithms of the determinants of covariance matrices. If the matrix is scalar, 1 by 1, the determinant is simply the log of variance, however the test is still valid. Each cluster type was examined by vegetation index, and the natural log of its variance was determined (Table C7, Appendix C, for cluster type 1, and Table C8, Appendix C, for cluster type 2).

If covariance matrices proved homogeneous, a linear classification rule was developed. If the matrices proved heterogeneous, a quadratic classification rule was developed. Johnson and Wichern (1982) present a development and definition of linear versus quadratic rules.

For cluster type 1, principal component loadings, the only indices variances not significantly different from one another were the normalized differences, ND7 and ND6. The nature of these indices is one forcing the computed value into a range between zero and one. As a consequence, by their properties, variances will be forced into an even smaller range. The test statistic will then be forced away from the critical value needed to reject the null hypothesis of homogeneity.

Cluster type 2, based on the raw data, had a larger number of linear classification rules versus type 1. Those indices having quadratic rules had homogeneity test statistic values with critical probability values of between .05 and .01. This degree of overlapping was not surprising, because the clusters were oriented in an almost linear trend, as discussed in the section on cluster means. Inspection of the spectral means for each respective cluster type from the reduced data set will further illustrate the relative degree of separation in cluster type 1, and overlapping in cluster type 2. (Table C9, Appendix C) .

Classification was then performed using the spectral data for each cluster type, using a linear or quadratic classifier, depending on the results of the covariance homogeneity test. The results of the classification were summarized as the total number of stands classified into a cluster type category (Table 14 for cluster type 1, and Table 15 for type 2). The original cluster means, from the reduced data set (see Tables 12 and 13) for the cluster categories into which a majority of stands were classified, were used as an indication of the nature of stand category into which classification happened. Means of the reclassified categories were not taken due to time constraints, and because clustering results are highly dependent for small sample sizes such as with this study. Additional samples and/or data sets need to be taken in order to verify that the clusters obtained are stable at larger sample sizes. A number of clusters in the study data set contain a single observation that would imply that they are outliers and should be discarded. Those stands which could be considered as outliers are the heavily stocked small pole timber ones. If the analyst were to obtain a larger sample size, it is likely that stands of this composition would be adequately represented. Finally, for the results about to be discussed to be applicable, using operational forest management categorizations is preferable to clustering categories, which are data set dependent.

Utilizing the original means as an index of how stands are clumping together, a number of patterns emerge. With cluster type 1, the majority of stands appear to be driven into four of the original clusters: 3, 6, 7, and 9. The majority of stands in cluster type 2 appear to be driven into clusters 8 and 9, and in some cases into number 5 as well. Cluster type 2 appears to be aligned along the stocking, and as a consequence, volume axis. Diameter (Dbh), and height appear to be uniform across clusters. The distinctions among cluster type 1 stands are most apparent along both the height and stocking axes. Clusters 3 and 9 appear to be heavily and lightly stocked poletimber, respectively. Clusters 6 and 7 appear to be

Table 14. Classification results for cluster type 1. The number of stands classified into a particular cluster are presented by vegetation index.

Cluster Number	Vegetation Index					
	D7	D6	DD	R7	R6	OC
1	0	0	0	0	0	1
3	29	38	41	41	40	18
4	0	3	0	0	0	5
6	29	8	14	28	26	17
7	8	9	12	6	8	17
8	0	0	0	0	0	2
9	13	21	12	4	5	18
10	0	0	0	0	0	1

Cluster Number	Vegetation Index					
	ND7	ND6	G	B	GB	OC
1	0	0	0	0	0	1
3	30	25	26	43	39	18
4	0	0	0	0	0	5
6	27	26	8	28	34	17
7	10	10	25	4	1	17
8	0	0	0	2	0	2
9	12	18	25	2	5	18
10	0	0	0	0	0	1

Cluster Number	Vegetation Index		
	BANDS	PRIN1	OC
1	0	0	1
3	21	40	18
4	1	0	5
6	30	30	17
7	4	4	17
8	3	0	2
9	20	5	18
10	0	0	1

Note: OC denotes the original stand count.

Table 15. Classification results for cluster type 2. The number of stands classified into a particular cluster are presented by vegetation index.

Cluster Number	Vegetation Index					
	D7	D6	DD	R7	R6	OC
1	0	0	0	0	0	1
3	0	0	34	0	0	19
5	0	2	2	0	0	6
6	0	0	0	0	0	5
7	0	0	0	0	0	1
8	32	31	29	20	22	22
9	47	46	14	28	34	24
10	0	0	0	0	0	1

Cluster Number	Vegetation Index					
	ND7	ND6	G	B	GB	OC
1	0	0	0	0	0	1
3	0	0	0	0	0	19
5	0	0	0	22	30	6
6	0	0	0	0	0	5
7	0	0	0	0	0	1
8	27	24	31	25	32	22
9	52	55	48	32	17	24
10	0	0	0	0	0	1

Cluster Number	Vegetation Index		
	BANDS	PRIN1	OC
1	0	0	1
3	33	0	19
5	14	0	6
6	3	0	5
7	0	0	1
8	17	45	22
9	12	34	24
10	0	0	1

Note: OC denotes the original stand count.

heavily and lightly stocked sawtimber, respectively. The main point to bear in mind here is that classification did not force each of the stands into a single category, as would be expected in a forest too spectrally homogenous to be effectively classified.

The relative degree of separation between clustering categories is another useful index of how stands are distributed in spectral space. The Mahalanobis distance, D^2, is a squared distance, weighted by the covariance matrix, commonly used to measure the separation between categories in classification analysis. It is defined as follows, using matrix notation:

$$D^2 = (m_1 - m_2)'S^{-1}(m_1 - m_2), \text{ where:} \qquad (12)$$

m_1, m_2 are the category means of interest, and the covariance matrix inverse is S^{-1}.

Table C14, Appendix C, exhibits the Mahalanobis distances between classes, by cluster type and cluster number, along with the probability of a greater D^2. The five following vegetation index combinations were used fro the sake of brevity, due to time constraints on this study, to illustrate these distances: (1), the four Landsat bands (B4, B5, B6, B7), (2), G, B, GB, (3), R7 and R6, (4), ND7 and ND6, and (5), D7, D6, and DD.

A consistent pattern of cluster separation occured within both cluster types (Table 16). A greater degree of separation is attained among cluster type 1 versus cluster type 2. The primary variable that influences separation is stocking with emphasis on height and volume in a joint sense. The probability of a greater D^2, since it is distributed as an F statistic multiplied by the appropriate constant (Mardia et al. 1979), is an indication of the significance of the pairwise separation between cluster cateogries. Good separation does NOT imply good classification, however. Nevertheless, the overall indication is that timber types in southeastern forests can be separated using Landsat multispectral scanning data, for which the D^2 statistics are given as reinforcement.

To summarize, the results of the approach using clustering followed by classification, two main points will be reiterated. First, pairwise cluster separation was shown for a substantial proportion of the clusters

Table 16. Significantly separated cluster pairs using D^2, the Mahalanobis distance. Significance is at the .05 level.

(cluster type 1)				
Vegetation Index Combinations				
B4-B7	G,B,GB	R7,R6	ND7,ND6	D7,D6,DD
1,3	1,3	---	---	1,3
1,4	1,4	1,4	1,4	1,4
1,6	1,6	1,6	1,6	1,6
1,7	1,7	1,7	1,7	1,7
1,8	1,8	1,8	1,8	1,8
1,9	1,9	1,9	1,9	1,9
1,10	---	1,10	1,10	1,10
3,4	3,4	3,4	3,4	3,4
3,6	3,6	3,6	3,6	3,6
3,7	3,7	3,7	3,7	3,7
4,6	4,6	4,6	4,6	4,6
4,7	4,7	4,7	4,7	4,7
4,9	4,9	4,9	4,9	4,9
6,7	6,7	6,7	6,7	6,7
7,9	7,9	7,9	7,9	7,9
8,9	8,9	8,9	8,9	8,9
(cluster type 2)				
1,3	1,3	1,3	1,3	1,3
1,5	1,5	1,5	1,5	1,5
1,6	1,6	1,6	1,6	1,6
1,8	1,8	1,8	1,8	1,8
1,9	1,9	1,9	1,9	1,9
5,9	5,9	5,9	5,9	5,9
---	6,8	---	---	---
6,9	6,9	6,9	6,9	6,9
7,8	7,8	7,8	7,8	7,8
7,9	7,9	7,9	7,9	7,9

available, to be significant at the .05 level. Second, classification did not force the data into a single category, as would be expected in a spectrally homogeneous forested scene. The implication is that, using spectral data, separation of forest stands into mensurationally based categories is possible.

Direct regression approach

The direct regression approach used volume, in cords per acre, and directly fitted it against spectral signature (Table 17 and Table C15, Appendix C, present statistics for volume estimation).

The best candidate among single vegetation indices, excluding the principal component scores and raw Landsat bands, was the Kauth and Thomas brightness index (Figure 3 presents a scatterplot of volume versus the brightness index). The coefficient of determination, .2548, is a good value in terms of Landsat data, due to the coarseness of resolution and its averaging effect across a picture element. It confirms the simple correlation coefficient between the two (see Table 10). The brightness index was intended as a relative index of the amount of soil visible on an image. The interpretation of the negative slope can be viewed as an inverse relationship between the amount of canopy foliage, and consequent amount of volume, intercepting the soil reflectance component. The "disguise" effect, defined earlier, probably was the main factor in not having a higher coefficient of determination. The relative precision and bias of the model, as reflected in the PRESS summary (Table C15, Appendix C), was one of the best among all single vegetation indices.

The simple ratios, R7 and R6, and normalized differences, ND7 and ND6, being designed along similar lines as brightness, can be considered as proportionality forms of indices. As the soil component increases, the simple ratios decrease in value, and a similar pattern of behavior occuring with the normalized differences. Their lower coefficients of determination, less precise and more biased PRESS summaries, when compared with brightness, are likely due to the "disguise" effect. A similar argument can be made with the greenness index, designed to indicate the relative amount of vegetation on a scene. The behavior of the greenness-brightness ratio (GB) can be attributed to the disguise effect and B.

Table 17. Direct volume regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C15, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.2376	23.995	.0001
R6	.2319	23.245	.0001
ND7	.2390	24.178	.0001
ND6	.2325	23.332	.0001
G	.0856	7.205	.0089
B	.2548	26.323	.0001
GB	.2485	25.429	.0001
D7	.1590	14.560	.0003
D6	.0188	1.478	.2278
DD	.0223	1.758	.1894
PRIN1	.2474	25.308	.0001
PRIN1-PRIN4	.2919	7.627 (1)	.0001
B4-B7	.2919	7.627 (1)	.0001

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f..

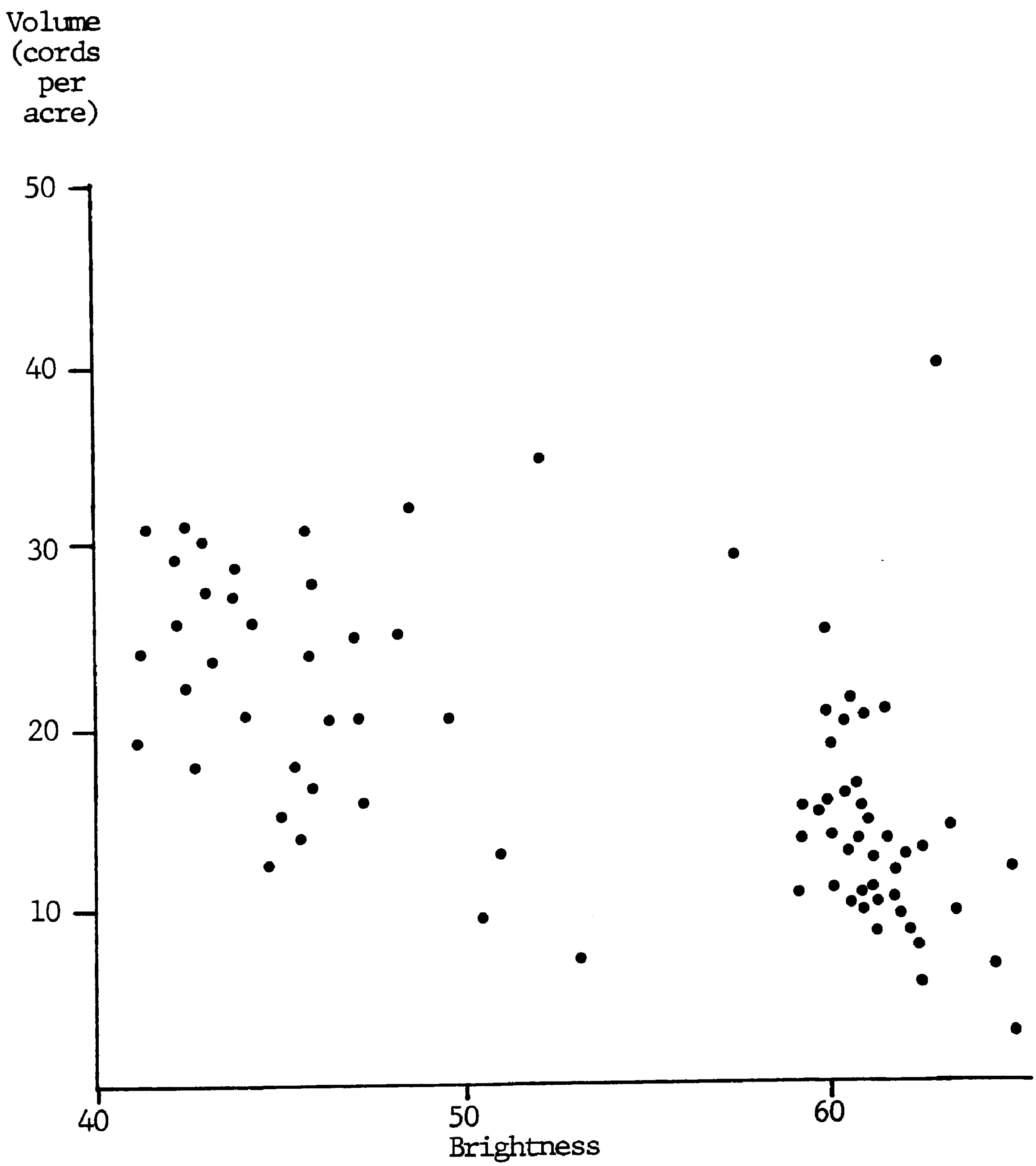


Figure 3. Scatterplot of volume versus the Kauth and Thomas brightness index.

The difference-6 index (D6) has low relative performance, likely due to the "disguise" effect and to the location of band 6 on a typical vegetation reflectance curve. The majority of stands in the study area consists of a mixture of pine and hardwoods. The differential reflectance between pine and hardwood in the individual stand, combined with the "disguise" effect induce additional variation not "explained" by the regression model. The difference-7 (D7) is intermediate in behavior which is also likely due to differential reflectance and the "disguise" effect.

Principal components were extracted from the correlation matrix of the raw Landsat data (Table 18). The first component, accounting for over 85 percent of the sample variation, can be thought of as the data set brightness component. The remaining components can be thought of as, respectively, an infrared index, a red-greed contrast, and a green-infrared contrast.

The first principal component is intermediate in behavior and was not ranked as "best" among single indices because of relatively large bias and imprecision from the PRESS summary, and its lower coefficient of determination. All four components, while having the largest R^2 and relatively good precision and bias characteristics, are subjectively incorrect in interpretation. The purpose of principal component analysis is reduction of a problems dimensionality through finding linear, orthogonal, uncorrelated combinations of the original data while keeping as much of the original sample variation as possible in as few linear combinations as possible. The latter components, accounting for little of the original variation of the sample, can induce unpredictable effects in estimation of new independent observations. Furthermore, the use of a particular equation, either using the original components or transforming back to the original raw variables, is a largely judgemental decision.

Finally, the use of all four bands is inappropriate due to the multicollinearity effects resulting from the strong correlation among the bands (Table 19). The high degree of correlation among each of the bands can be attributed to their proximity to each other on the electromagnetic spectrum. The response in one band will almost always affect the other bands.

Table 18. Principal components of the raw spectral data.

Component Number	Eigenvalue	Cumulative Proportion
1	3.4006	.850160
2	0.5765	.994284
3	0.0177	.998697
4	0.0052	1.000000

Landsat Band	Eigenvectors			
	1	2	3	4
4	0.5161	-0.3903	-0.5291	0.5490
5	0.5223	-0.3261	0.7871	0.0358
6	0.5408	0.0345	-0.3090	-0.7816
7	0.4104	0.8603	0.0707	0.2939

Note: The coefficients of the eigenvectors form a linear combination of the raw data. For example, let T = the new variable created by the first component, then,

$$T = 0.5161 \times \text{Band } 4 + 0.5223 \times \text{Band } 5 + 0.5408 \times \text{Band } 6 + 0.4104 \times \text{Band } 7.$$

Table 19. Partial correlations among the raw Landsat bands. The significance probability of the correlation is in the row immediately below the correlation coefficient.

Landsat Band	Landsat Band			
	4	5	6	7
4	1.0000 .0000	0.9471 .0001	0.7027 .0001	0.5391 .0001
5		1.0000 .0000	0.6473 .0001	0.4571 .0001
6			1.0000 .0000	0.9464 .0001
7				1.0000 .0000

To summarize, a simple linear regression formulation shows a significant association between timber volume in cords per acre and spectral signature, as represented by Landsat digital data. The verification analysis, to be discussed later, predicted volume for a set of stands reserved from the original analysis. The assertions discussed here will be reinforced in that section.

One point to bear in mind for this section is that the measure of volume used in this study was a merchantable one. Merchantable volume is that portion of total volume expected to be actually used at the saw or paper mill. Total volume, or biomass, measures are likely to be better associated with spectral signature.

Indirect regression approach

The emphasis in this section was to estimate canopy closure, crown diameter, and average stand height from spectral data, and use these estimates as inputs for calculating volume using the aerial photo volume table equation developed in the first analytical section.

The same procedure was used to estimate volume in the direct approach as was used in this section for canopy closure, crown diameter, and stand height. The discussion will be restricted to the best five crown diameter, two canopy closure, and five height regression models. This was done for the sake of brevity due to the time restrictions on this study. The indirect contribution to the verification analysis used, in addition to the five direct regression models deemed best, the fifty combinations of canopy closure, crown diameter, and height defined above. The 2,202 possible volume estimation combinations proved too cumbersome for complete discussion in this study due to time considerations. The key points raised here are adequately demonstrated with the fifty five volume estimation methods selected.

The results for crown diameter show the best five vegetation index candidates to be: all four bands, D7, GB, ND7, and R7 (Table 20 and Table C16, Appendix C, present crown diameter regression statistics). If an analyst were interested in sidestepping multicollinearity effects, the difference-7 (D7) index would be the preferable one (Figure 4 presents a

Table 20. Crown diameter regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C16, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.2162	21.243	.0001
R6	.1875	17.774	.0001
ND7	.2323	23.303	.0001
ND6	.1998	19.225	.0001
G	.1686	15.617	.0002
B	.1850	17.849	.0001
GB	.2365	23.848	.0001
D7	.2406	24.396	.0001
D6	.0003	0.025	.8757
DD	.0015	0.119	.7307
PRIN1	.1632	15.109	.0002
PRIN1-PRIN4	.2984	7.534 (1)	.0001
B4-B7	.2984	7.534 (1)	.0001

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f..

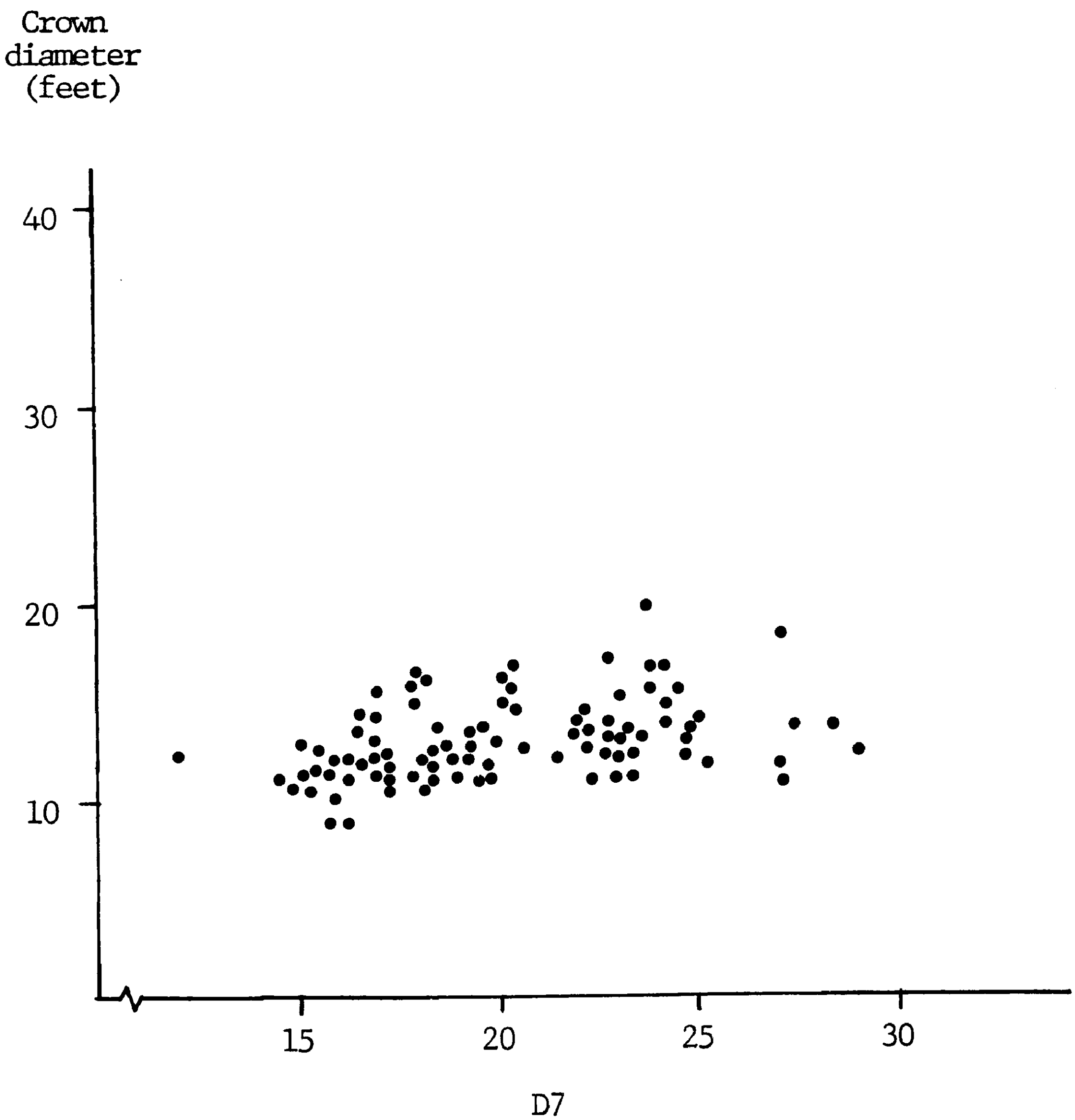


Figure 4. Scatterplot of crown diameter versus the difference-7 index (D7).

scatterplot of D7 versus crown diameter).

The D7, ND7, and R7 indices are apparently similar behaved for this stand parameter. As the amount of foliage and infrared reflectance component increases, crown diameter increases. The source of the low coefficient of determination is both an extension of the "disguise" effect and the coarse resolution of the Landsat sensor. There is no certainty as to whether or not crowns being viewed by the sensor are in the understory or the overstory. Clumps of trees can be viewed as a single overly large crown, or understory vegetation can inflate the average crown size. There is also no certainty as to whether or not, in near completely closed canopies, the average crown size is small as a result of numerous small trees, or large resulting from few, large, trees. This effect will be referred to as crown size confounding.

The greenness-brightness ratio is likely a combination of components equally contributed by greenness and brightness rather than a predominance of one over another. All four individual indices selected are likely to be more reliable in estimating crown diameter when understory vegetation is readily distinguishable from overstory, specifically in the autumn when it is completely cured.

Another source of variation that is possible in decreasing the coefficient of determination is due to tectural differences in an image. Stands having widely disparate height classes such as a thorough mixture of dominant through overtopped trees can be viewed through a sensor as a uniform stand at either a large crown diameter if more dominant trees exist, or small crown diameter if more codominant through overtopped trees are in the stand.

Canopy closure, being subjective in nature due to the experience and judgement of the photointerpreter, was not expected to have a good fit, primarily due to the "disguise" effect. The two candidates selected for use in the verification analysis were the four Landsat bands (B4, B5, B6, B7), and the greenness index (Table 21 and Table C17, Appendix C, present regression statistics for canopy closure). Greenness is the single index closest to being significant in terms of regression statistics (Figure 5 presents a scatterplot of greenness versus canopy closure). The low coefficient of determination is due to the narrow range of values

Table 21. Canopy closure regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C17, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.0070	0.541	.4645
R6	.0060	0.461	.4992
ND7	.0095	0.741	.3919
ND6	.0083	0.648	.4233
G	.0155	1.214	.2739
B	.0028	0.213	.6453
GB	.0090	0.698	.4061
D7	.0150	1.170	.2828
D6	.0057	0.443	.5078
DD	.0053	0.408	.5250
PRIN1	.0018	0.143	.7067
PRIN1-PRIN4	.0174	0.328 (1)	.8581
B4-B7	.0174	0.328 (1)	.8581

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f.

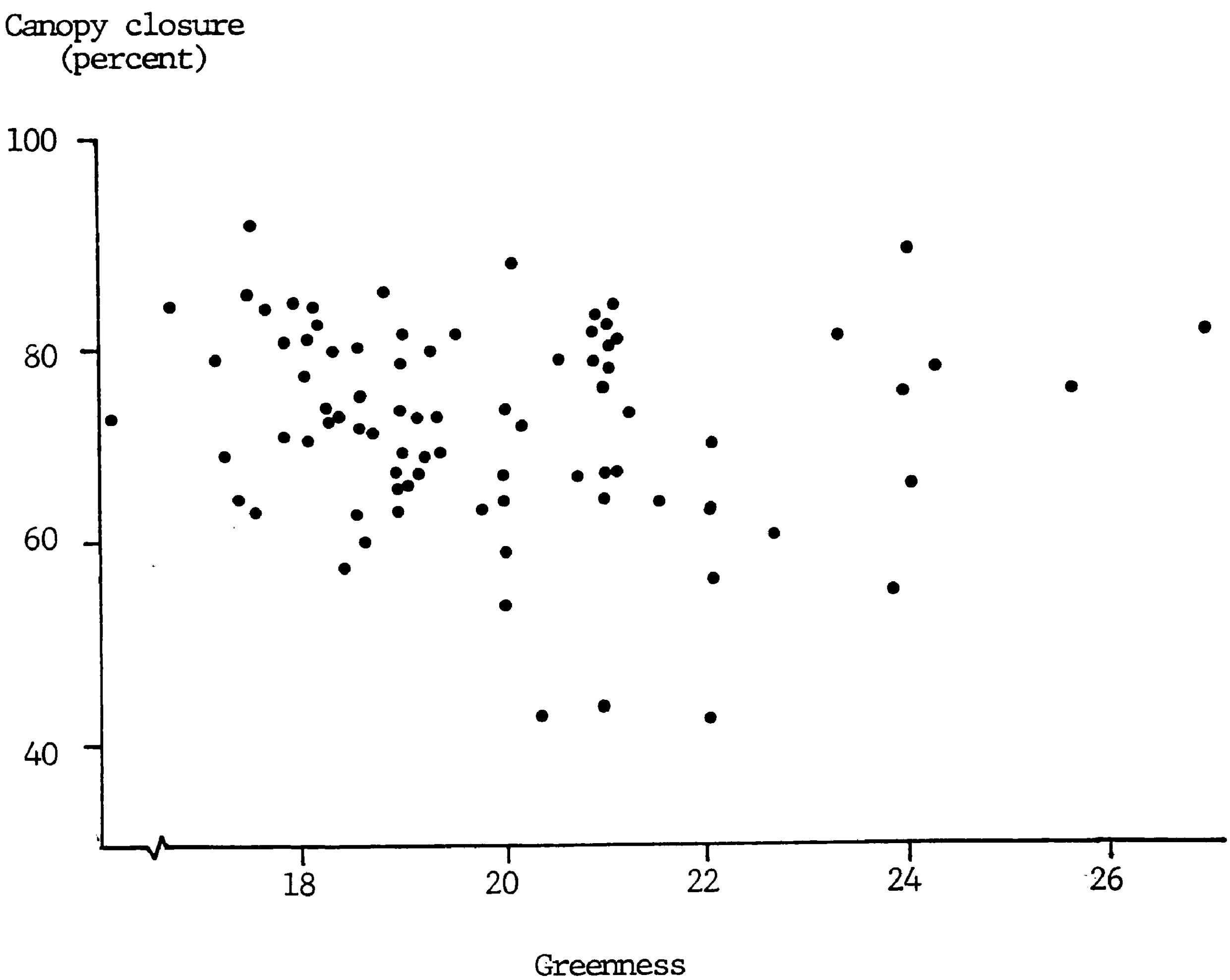


Figure 5. Scatterplot of canopy closure versus the Kauth and Thomas greenness index.

observed in the reduced data set and the "disguise" effect. In all of the regressions fitted to canopy closure, none were shown to be different from zero at the .10 level of significance.

The crown size confounding effect also comes into play here in that inflated crown size tends to inflate the canopy closure estimate.

The degree of fit for all equations estimating height was unexpectedly good (Table 22 and Table C18, Appendix C, present regression statistics for height). The candidates selected for use in verification analysis were: the four Landsat bands (B4-B7), the greenness-brightness ratio, normalized difference-7, simple band 7 to band 5 ratio, and brightness. The best single vegetation index is the greenness-brightness ratio (Figure 6 presents a scatterplot of greenness-brightness versus height). The possible reasons for the good degree of fit about to be discussed are speculative in nature, and this should be borne in mind.

Taller trees will have a larger relative volume of crown foliage versus shorter trees, assuming that crown ratio, the proportion of tree actually supporting foliage, is constant. A larger infrared reflectance component would then be seen by the sensor. Changes in crown geometry as trees age could also affect the bidirectional reflectance properties of the canopy. Older, taller, trees are flattened into an ellipsoidal shape, presenting more foliage to the sensor. Younger, shorter, trees, being more conical in shape, present relatively less foliage to the sensor. Differential crown geometries across species could also affect reflectance properties of the canopy, and possibly bias the height estimate. Textural differences could also be responsible for the good height estimates. Taller trees project more shadow within a picture element, covering the ground and understory and suppressing the soil component, thereby enhancing the infrared and greenness components to reflectance and the greenness-brightness ratio, respectively. Finally, physiological differences between young and old trees could also affect their reflectance properties, changing their tonal characteristics.

The three aerial photo volume table inputs, estimated from multispectral imagery was an intermediate step in the preparation for the verification analysis. Comparisons of direct, indirect, and actual field volumes were done for a set of stands reserved from initial analysis.

Table 22. Height regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C18, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.4384	60.115	.0001
R6	.4007	51.487	.0001
ND7	.4702	68.334	.0001
ND6	.4275	57.496	.0001
G	.2365	23.854	.0001
B	.4361	59.552	.0001
GB	.4781	70.553	.0001
D7	.3865	48.508	.0001
D6	.0116	0.902	.3452
DD	.0083	0.641	.4528
PRIN1	.4068	52.801	.0001
PRIN1-PRIN4	.5085	19.136 (1)	.0001
B4-B7	.5085	19.136 (1)	.0001

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f..

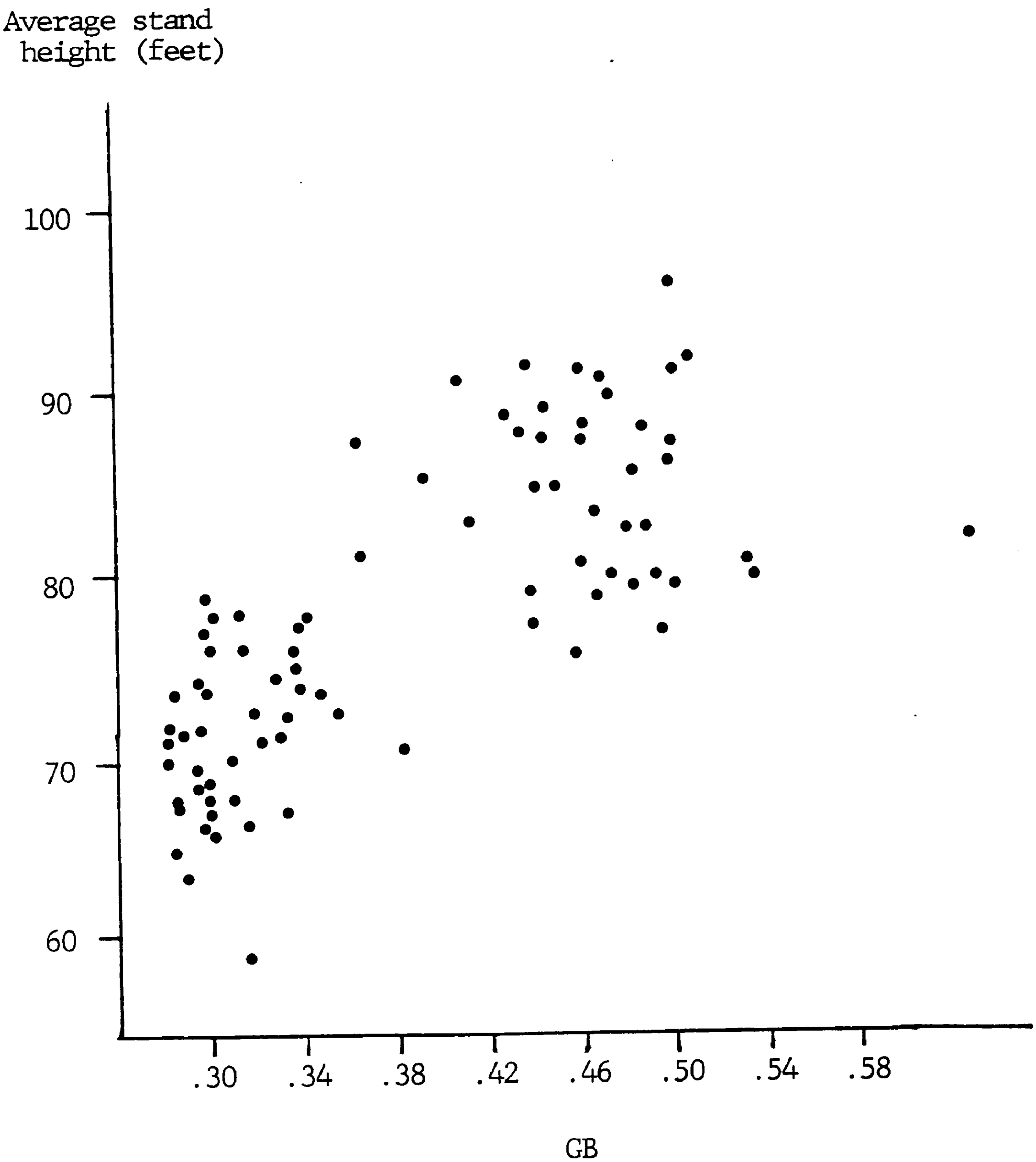


Figure 6. Scatterplot of average stand height versus the greenness-brightness ratio (GB).

This was done to obtain an indication of how a spectrally derived volume estimator would work when operationally developed.

Verification analysis

Twenty two stands were reserved from analysis for the purpose of predicting volume based on the derived spectral regression models developed in the previous sections. Fifty five volume estimation methods, five direct and fifty indirect, were used to estimate volume (these are defined in Table 23).

The comparative approach consisted of two steps. First, each volume estimation was compared, in a pairwise manner, with the actual field volume in a one way analysis of variance, with each estimation technique considered as a treatment. The second step compared direct versus indirect methods, also in a pairwise fashion, regressing the second against the first. The coefficient of determination was used as an index of comparison for the two methods. The General Linear Models procedure (GLM) on SAS was used for each step in this section (SAS 1982).

The general pattern of means of each estimation method is one of underestimation with direct patterns, and overestimation with indirect methods (Table 24). A notable exception is that of method VA2, the direct method using brightness, which was nearly equal to the actual field volume.

The analysis of variance summary (Table 25) for each estimation method showed, in all cases, that estimated volume was not significantly different from actual field volume at the .10 level. At the .05 level, VA3, VA4, VA5, V212, V213, V214, V221, V222, V223, V224, V322, V323, V324, V422, V423, V424, V522, V523, and V524 were significantly different from field volume. With the exception of V212, V213, and V214, the methods of indirect estimation which were significantly different used the second estimation method for canopy closure, greenness, which in all cases is subject to the "disguise" and crown size confounding effects. Another possible reason for this significance is the doubling effect to variation encountered in the indirect method, using two methods of estimation that have sampling and experimental errors associated with them. The analysis of variance results also reflect same estimation behavior.

Table 23. Volume estimation method definitions for verification analysis.

Direct Estimation Methods	
VACT = Mean Volume of the 22 reserved stands	
VA1 = All four Landsat bands regression	(B4-B7)
VA2 = Brightness regression	(B)
VA3 = Greenness-brightness ratio regression	(GB)
VA4 = Normalized difference-7 regression	(ND7)
VA5 = Band 7/Band 5 ratio regression	(R7)
Indirect Estimation Methods	
CD1 = All four bands regression	(B4-B7)
CD2 = Difference-7 regression	(D7)
CD3 = Greenness-brightness ratio regression	(GB)
CD4 = Normalized difference-7 regression	(ND7)
CD5 = Band 7/Band 5 ratio regression	(R7)
CC1 = All four bands regression	(B4-B7)
CC2 = Greenness regression	(G)
HT1 = All four regression	(B4-B7)
HT2 = Greenness-brightness ratio regression	(GB)
HT3 = Normalized difference-7 regression	(ND7)
HT4 = Band 7/Band 5 ratio regression	(R7)
HT5 = Brightness regression	(B)

Note: The APVT equation used to estimate volume is:

$$V_{ijk} = 12.20876 - 3.13304CD_i + 0.2037978CC_j + 0.02115364HT_k + 0.338764(CD_i \times CC_j) + 2.5076 \times 10^{**-6}(CC_j \times CD_i^2 \times HT_k).$$

Each subscript refers to an estimation method used for the inputs to the APVT equation. For instance, V111 refers to the combination of CD1, CC1, and HT1, with this order of variables used for all combinations of APVT inputs.

Note: Tables 17 and C15, Appendix C, present regression statistics for direct volume regressions.

Table 24. Mean of field volume estimates (cords per acre), and direct and indirect spectral volume estimation methods.

Actual mean of the reserved stands

23.645 cords per acre

Volume Estimation Method	Mean	Volume Estimation Method	Mean
Direct Methods			
VA1	20.553	VA2	23.787
VA3	20.781	VA4	20.693
VA5	20.544		
Indirect Methods			
V111	33.294	V112	33.175
V113	33.141	V114	33.029
V115	33.421	V121	36.267
V122	36.143	V123	36.108
V124	35.990	V125	36.399
V211	31.690	V212	31.589
V213	31.559	V214	31.458
V215	31.794	V221	34.494
V222	34.390	V223	34.358
V224	34.252	V225	34.602
V311	32.272	V312	32.166
V313	32.137	V314	32.029
V315	32.416	V321	35.134
V322	35.023	V323	34.992
V324	34.880	V325	35.283
V411	33.108	V412	32.997
V413	32.967	V414	32.855
V415	33.264	V421	36.059
V422	35.944	V423	35.913
V424	35.976	V425	36.211
V511	32.108	V512	32.006
V513	31.979	V514	31.875
V515	32.256	V521	34.954
V522	34.848	V523	34.819
V524	34.711	V525	35.108

Note: See table 23 for estimation method definitions.

Table 25. Hypothesis test results comparing field volume and estimates derived from spectral estimators.

Estimate Method	$F_{(1,20)}$	Prob(F)	R^2
		Direct Methods	
VA1	2.72	.1148	.11969
VA2	1.52	.2323	.07051
VA3	3.75	.0671	.15783
VA4	3.86	.0636	.16168
VA5	4.04	.0581	.16803
		Indirect Methods	
V111	1.06	.3148	.05047
V112	1.56	.2267	.07217
V113	1.59	.2217	.07369
V114	1.62	.2178	.07491
V115	0.68	.4209	.03267
V121	1.36	.2578	.06534
V122	1.88	.1856	.08588
V123	1.92	.1815	.08744
V124	1.95	.1782	.08871
V125	0.95	.3410	.04542
V211	2.75	.1130	.12079
V212	3.68	.0695	.15534
V213	3.79	.0657	.15394
V214	3.93	.0613	.16423
V215	2.19	.1548	.09854
V221	3.23	.0875	.13898
V222	4.12	.0555	.17087
V223	4.22	.0531	.17438
V224	4.34	.0504	.17819
V225	2.73	.1144	.11992
V311	1.97	.1754	.08981
V312	2.56	.1255	.11337
V313	2.57	.1243	.11402
V314	2.64	.1199	.11661
V315	1.34	.2602	.06296
V321	2.38	.1386	.10635
V322	2.98	.0999	.12953
V323	2.99	.0992	.13008
V324	3.06	.0957	.13262
V325	1.71	.2052	.07898
V411	2.01	.1718	.09125
V412	2.60	.1229	.11486

Table 25 (cont.)

Estimate Method	$F_{(1,20)}$	Prob(F)	R^2
	Indirect Methods (cont.)		
V413	2.61	.1221	.11592
V414	2.67	.1179	.11776
V415	1.35	.2593	.06315
V421	2.41	.1365	.10742
V422	3.00	.0985	.13056
V423	3.01	.0981	.13087
V424	3.08	.0948	.13327
V425	1.71	.2059	.07874
V511	2.06	.1670	.09324
V512	2.66	.1188	.11723
V513	2.66	.1184	.11749
V514	2.71	.1152	.11944
V515	1.35	.2586	.06331
V521	2.47	.1314	.11009
V522	3.08	.0947	.13332
V523	3.08	.0947	.13343
V524	3.13	.0923	.13519
V525	1.73	.2031	.07968

Let VACT = field volume, and V(i) = estimated volume, then the null hypothesis for this test can be written as:

H_0: VACT = V(i), versus H_A: VACT $\neq$ V(i).

Note: See Table 23 for definitions of spectral volume estimation methods.

The comparison between direct and indirect estimation methods, using the coefficient of determination as a comparative index and derived from simple linear regressions, shows that in all cases the correspondence was very high, indicating potential operational applicability (Table 26).

To recapitulate the results of this section, a significant association between spectral signature and timber volume, and those stand parameters used as inputs to an aerial photo volume table, exists. The direct estimation method regressed volume against a vegetation index. The indirect method estimated canopy closure, crown diameter, and average stand height as inputs to an APVT. Prediction of volume on a reserved set of stands using a number of estimation techniques, both direct and indirect, was shown not be significantly different from the actual field volume at the .05 level. Those methods, significant at the .10 level, are apparently at the extreme ends of underestimation for direct methods, and overestimation for indirect methods.

Several points must be remphasized at this time. Volume, in cords per acre, was a merchantable measure of volume. A more appropriate measure for use with remote sensing devices is total tree volume such as cubic feet per acre. Height was estimated using a site index based equation developed by Schumacher and Coile (1960), from the available site index information obtained from the cooperating landowner. A stringent assumption of successful site tree location was made, effecting the use of the Schumacher and Coile equation. Canopy closure and crown diameter were measured on the only aerial photography available for the area at the time desired. The interpretative quality of the photography was poor, dominated by an orange caste, and with almost nonexistent tone, texture, and contrast. The scale of the photography was at the extreme limit for effective photo interpretation of canopy closure and crown diameter. An inflated sampling and experimental error introduced additional variation into the analysis. The need for corroboration of these results and that the context of this study was to demonstrate the existence of quantitative relationships between forest stand parameters and spectral signature was the goal, and not the development of operational techniques, must be kept in mind.

Table 26. Comparison between the direct and indirect volume estimation methods, using the coefficient of determination as a comparative index.

Indirect Method	Direct Method				
	VA1	VA2	VA3	VA4	VA5
V111	.92546	.91494	.90327	.88425	.89674
V112	.94529	.92872	.94491	.92837	.93945
V113	.94752	.93879	.95257	.93949	.94547
V114	.93927	.93401	.95116	.93716	.94070
V115	.91949	.96756	.85350	.84798	.86934
V121	.91224	.92347	.92268	.90552	.90991
V122	.91423	.91977	.95654	.94195	.94327
V123	.91215	.92641	.96206	.95122	.95225
V124	.89036	.91284	.96114	.94910	.94913
V125	.91321	.97761	.89270	.89245	.90387
V211	.94284	.89334	.96299	.94328	.94578
V212	.95381	.83725	.98626	.97043	.97044
V213	.93491	.87904	.99739	.98222	.98212
V214	.91212	.86422	.99433	.98162	.98157
V215	.97387	.97947	.94879	.95055	.95835
V221	.90975	.87678	.96750	.94912	.94828
V222	.87949	.83943	.97936	.96499	.96506
V223	.87375	.83936	.98352	.97348	.97394
V224	.84418	.81787	.98254	.97106	.97339
V225	.93891	.95889	.97544	.98170	.98319
V311	.98595	.96604	.94326	.93848	.94959
V312	.97404	.96326	.94938	.95357	.95878
V313	.96512	.96827	.94139	.94969	.95419
V314	.95064	.96208	.93736	.94881	.95208
V315	.92229	.99506	.80961	.81294	.83205
V321	.96770	.96428	.96907	.97304	.97473
V322	.94329	.94571	.97143	.98508	.98145
V323	.93236	.94889	.96293	.97630	.97679
V324	.91266	.93719	.96209	.97634	.97506
V325	.91531	.99676	.85095	.97095	.87671
V411	.98133	.97286	.93969	.93923	.93935
V412	.96496	.96733	.94304	.95129	.95568
V413	.95391	.97099	.93279	.94526	.94896
V414	.93819	.96409	.92793	.94401	.94474
V415	.90951	.99738	.79392	.80598	.82197
V421	.96409	.96769	.96441	.96945	.97325
V422	.93204	.94286	.96313	.97660	.97708
V423	.91291	.94842	.95246	.97022	.97035
V424	.89838	.93596	.94898	.96983	.96828
V425	.90041	.99627	.83572	.85917	.86417

Table 26 (cont.)

Indirect Method	Direct Method				
	VA1	VA2	VA3	VA4	VA5
V511	.97259	.97243	.93517	.93714	.94394
V512	.94934	.96112	.93767	.94942	.95070
V513	.93967	.96409	.92635	.94281	.94342
V514	.91546	.95176	.92108	.94165	.93898
V515	.89472	.99581	.77876	.77970	.80373
V521	.94460	.96103	.96201	.96954	.97081
V522	.90883	.93484	.95949	.97629	.97482
V523	.89473	.93408	.97485	.96935	.96760
V524	.86814	.91597	.94407	.96898	.96551
V525	.88317	.98962	.82492	.95371	.85459

Note: See Table 23 for volume estimation method definitions.

Other stand parameters

In addition to volume, canopy closure, crown diameter, and average stand height, additional stand parameters are of interest to the forest manager. The primary ones are stems per acre, basal area, and average tree diameter (Dbh). These parameters will be examined in a univariate and multivariate sense. Canopy closure, crown diameter, and height will also be examined in a multivariate sense. The multivariate technique used here is canonical correlations, a technique that finds relationships between two sets of variables by finding pairs of linear combinations of these variables which are maximally correlated.

The univariate technique, regression, was done in the same manner as was done for volume, canopy closure, crown diameter, and height. The "best" candidates among the vegetation indices for tree diameter (Dbh) were shown to be: all four Landsat bands, the normalized difference-6, the difference-7, and the normalized difference-7 (Table 27 and Table C19, Appendix C, present regression statistics for tree diameter). A strong relationship between crown diameter and Dbh is the most likely reason for the regression coefficients being significantly different from zero (Figure 7 illustrates the relationship between spectral signature and Dbh with a scatterplot of Dbh versus the normalized difference-6 index).

Alternative explanations for the strong relationship between Dbh and spectral signature, in terms of Landsat data, are speculative in nature. Differential crown geometry and shape presented to the sensor as the tree ages is one possibility. The increase in Dbh as height increases is the association effecting this. Another possibility is that this data set is an anomalous one, with these model fittings unique to this data set alone.

The best candidates for basal area are: all four Landsat bands, the simple band 7 to band 5 ratio (R7), brightness, and the greenness-brightness ratio (Figure 8 illustrates the best index, R7, with a scatterplot of basal area versus R7. Table 28 and Table C20, Appendix C, present regression statistics for basal area). Basal area is confounded among stems per acre and individual tree size in a similar fashion as is canopy closure, discussed in the canopy closure and APVT sections. The

Table 27. Average Dbh regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C19, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.2987	32.792	.0001
R6	.3140	35.245	.0001
ND7	.3181	35.921	.0001
ND6	.3290	37.750	.0001
G	.2705	28.566	.0001
B	.2088	20.323	.0001
GB	.3184	35.969	.0001
D7	.3212	36.430	.0001
D6	.0267	2.112	.1502
DD	.0084	0.654	.4211
PRIN1	.1856	17.551	.0001
PRIN1-PRIN4	.3871	11.687 (1)	.0001
B4-B7	.3871	11.687 (1)	.0001

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f.

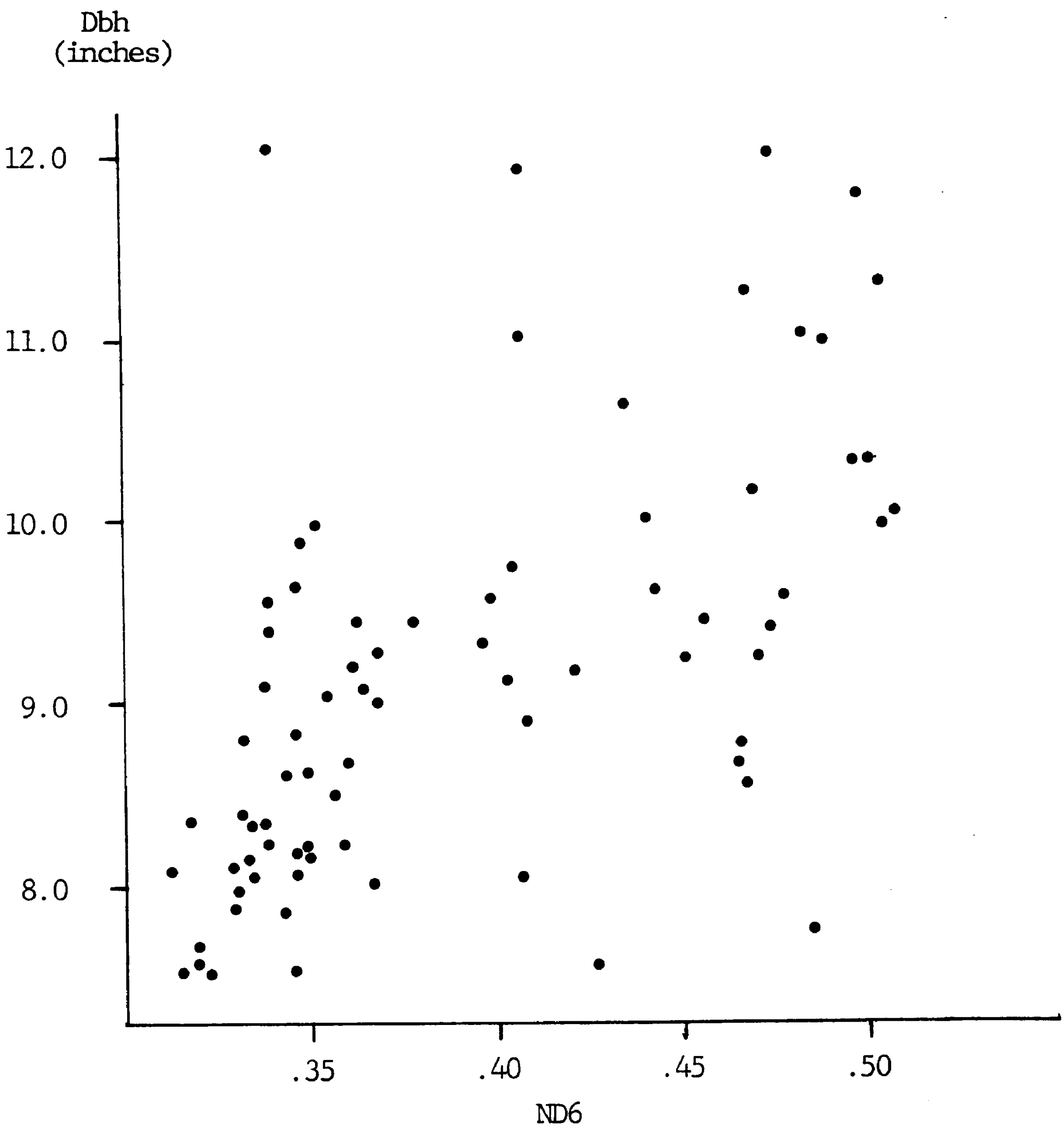


Figure 7. Scatterplot of average tree diameter (Dbh) versus the normalized difference-6 index (ND6),

Table 28. Basal area regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C20, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.1850	14.123	.0003
R6	.1471	13.314	.0005
ND7	.1510	13.691	.0004
ND6	.1435	12.905	.0006
G	.0490	3.965	.0500
B	.1681	15.557	.0002
GB	.1591	14.654	.0003
D7	.0975	8.320	.0051
D6	.0204	1.602	.2094
DD	.0174	1.365	.2463
PRIN1	.1635	15.051	.0002
PRIN1-PRIN4	.1927	4.415 (1)	.0030
B4-B7	.1927	4.415 (1)	.0030

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f.

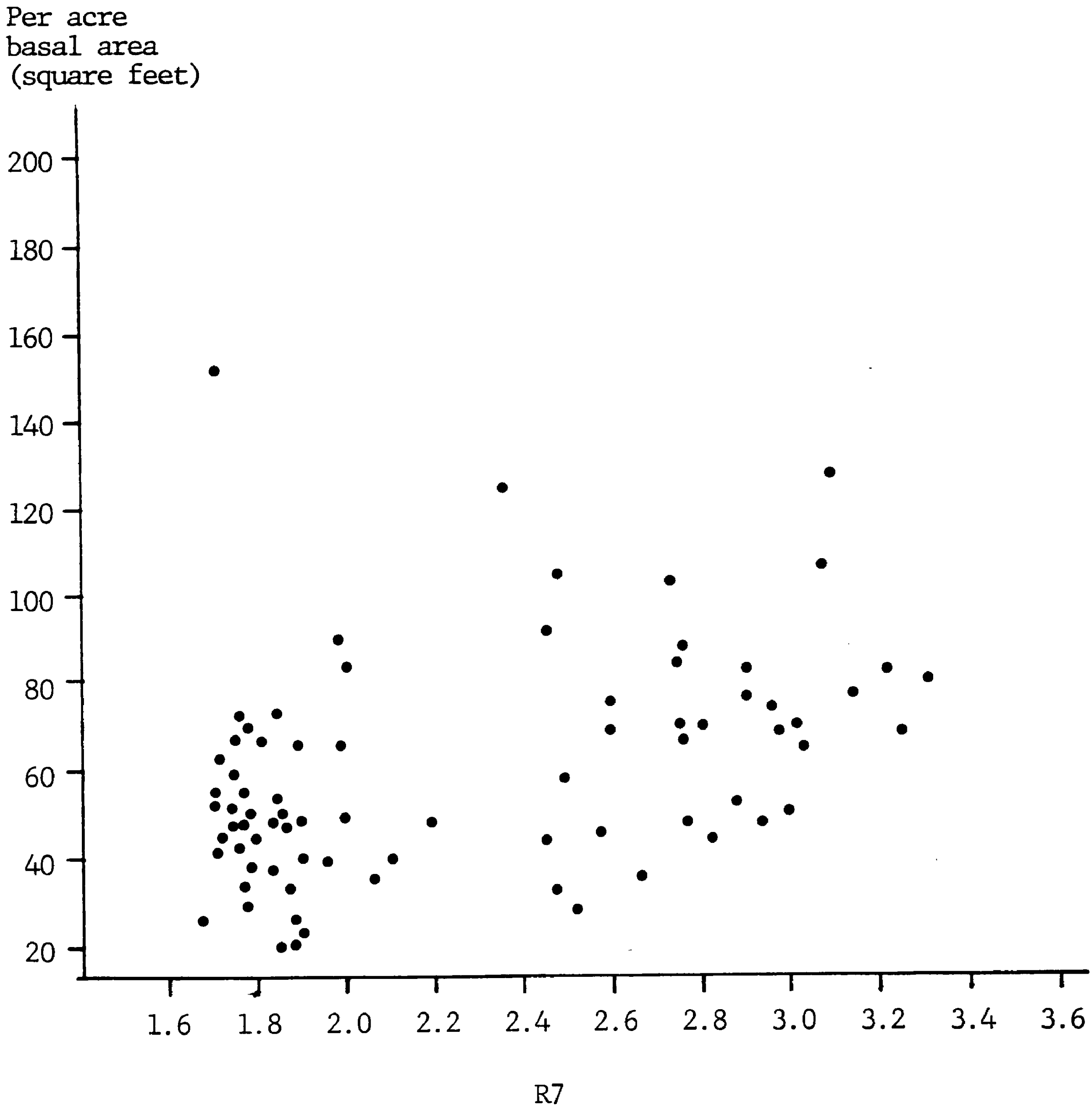


Figure 8. Scatterplot of per acre basal area versus the Landsat band 7 divided by band 5 ratio (R7).

"disguise" effect can also play a role, inflating basal area estimates through the sensor seeing, in effect, a "false canopy" component of reflectance which is actually understory vegetation. Physiological differences between larger, older, trees versus smaller, younger, trees could alleviate the confounding effect of basal area in that the stocking in a stand could, for instance, be separated among poletimber and saw timber.

Stems per acre has no usable candidates (Table 29 and Table C21, Appendix C, present stems per acre regression statistics). It is probably not a useful parameter on Landsat due to its resolution, and the radial crown growth phenomenon discussed in the canopy closure results (Figure 9 illustrates this with a scatterplot of stems per acre versus the D6 index).

The canonical correlation analysis reinforced the regression results in multivariate space, providing additional insight to these relationships. Two sets of ground parameters were used in describing the canonical relationships: (1), stems per acre, basal area, and average tree diameter; and (2), canopy closure, crown diameter, and height (Table C10, Appendix C presents the results for variable set 1, and Table C12 the results for set 2). The most "influential" variables were also determined and examined (Tables C11 and C13 for variable sets 1 and 2, respectively).

The pattern of behavior with stems per acre, basal area, and Dbh, is highly significant, at the .01 level. The significant situations arise when the most "influential" variables are either Dbh or basal area in a single sense, or both in a joint sense. Influence of variables is determined by the magnitude of covariance matrix eigenvector coefficients. Those canonical correlations which cannot be shown to be significantly different from zero are those where stems per acre is the singly or jointly most influential variable.

A similar pattern of behavior occurs when canopy closure is the singly or jointly most variable of set number 2. When this occurs, the canonical correlation is not significantly different from zero. In the case of crown diameter and height being the singly or jointly most influential variables, canonical correlations are significantly different from zero.

In both cases, the vegetation indices which were not different

Table 29. Stems per acre regression from spectral data: highlights of summary statistics. Complete statistics are given in Table C21, Appendix C.

Vegetation Index	R^2	$F_{(1,77)}$	Prob(F)
R7	.0011	0.084	.7726
R6	.0004	0.035	.8529
ND7	.0003	0.021	.8857
ND6	.000048	0.004	.9515
G	.0092	0.715	.4003
B	.0072	0.519	.4570
GB	.0006	0.047	.8290
D7	.0025	0.197	.6586
D6	.0376	3.006	.0870
DD	.0231	1.819	.1813
PRIN1	.0089	0.694	.4076
PRIN1-PRIN4	.0473	0.919 (1)	.4574
B4-B7	.0473	0.919 (1)	.4574

(1) Note: The F statistics for these two regressions are with 4 and 74 d.f..

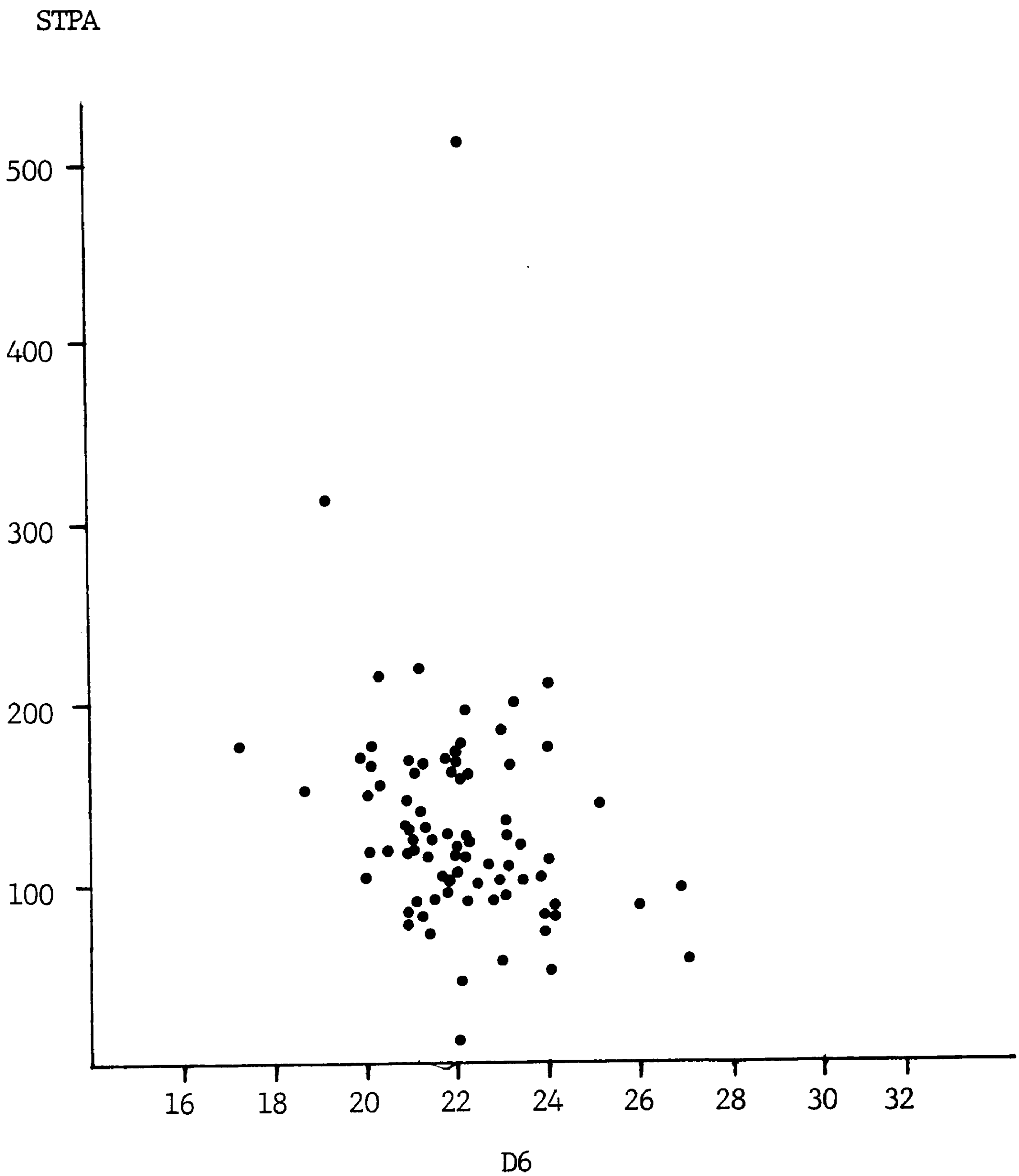


Figure 9. Scatterplot of stems per acre (STPA) versus the difference-6 index (D6).

from zero were the difference-6 and difference-difference indices.

Likely explanations for the low correlations are the reflectance differences between pine and hardwood, and the changes in imagery due to the fact of the summer of 1980 being an extraordinarily dry one. Band 6 is located at the point where a vegetation reflectance curve approaches the infrared plateau. Extreme drought could depress the curve for those species that are more susceptable to drought, introducing additional variation into their reflectance characteristics. The "disguise" effect is another reason for the lack of correlation in the D6 and DD indices, since these are the canonical correlations most influenced by canopy closure and stems per acre. The radial crown growth phenomenon is an alternative reason for the lack of correlation with stems per acre. The detail of both of these reasons have been outlined in the previous sections

To summarize, Dbh and basal area have a significant association with spectral signature. When all three of these parameters are reduced to a canonical correlation linear combination, significant correlations were found when canopy closure or stems per acre were not the most influential variables, in terms of eigenvector loading. The implication is that, provided corroboration of these results, basal area and Dbh can be estimated using spectral data, in addition to volume, crown diameter, and average stand height. A fairly thorough enumeration of forest stands is thus possible using spectral data.

Chapter VII.

SUMMARY DISCUSSION

This section presents a synopsis of results obtained for each sub-objective of this study, a general synthesis assembling each of the sub-objectives in terms of the overall objective, implications of these results, and recommendations for further analysis.

Subobjective synopsis

Aerial photo volume tables

The aerial photo volume tables and equations developed for this study (Tables 8 and 9) are statistically comparable with previous work done in southeastern forests (Avery 1958, Avery and Myhre 1959). In all cases, the coefficient ofmultiple correlation ranged from 0.68 to 0.77, with the differences accountable to photographic quality and interpretative differences. The final model for this study was easier to interpret than previous work in terms of the regression coefficients. Individual tree stocking was accounted for in the model by a term directly analogous to basal area (CDCC). Volume for individual trees was accounted directly for by the use of a term comparable with tree diameter squared multiplied by height (AB2C). A majority of APVTs in previous work used equations which were selected on the basis of regression slopes and either not using mensurational characteristics or using them only minimally. These models developed into little more than theoretical abstractions. The use of photographically measured variables, such as canopy closure, crown diameter, and parallactic height, is largely subjective in nature and will introduce bias and variable selection that will not be easily interpreted in terms of mensurational variables within a stand. The models used in this study represent an intermediate step in rectifying the subjectivity inherent in APVT development. Its regression slope coefficients are more easily interpreted in terms of stand characteristics.

Using stems per acre to construct expected crown diameter and canopy closure models, it was shown that canopy closure was more reliable of the two photo measurements taken and that smaller scale photography can be used to measure canopy closure effectively than crown diameter. Crown diameter was consistently underestimated, because in forested stands crowns tend toward greater overlapping as canopies become more completely closed. On occasions, the largest crowns measured within a photo plot were overestimated due to confusion of a group of trees with a single crown. The expected crown diameter model (Table C1, Appendix C) showed that if an accurate crown diameter estimate is available, and not biased towards smaller values due to crown overlap, the coefficient of determination nearly doubles.

Canopy closure

Canopy closure, which is confounded among stems per acre and tree diameter, cannot provide a reliable estimate of site occupancy. The likely explanation is that when forest management activities occur, such as thinning, an individual tree undergoes a period of radial crown growth, before substantial diameter accumulation occurs, referred to as the radial crown growth phenomenon. Due to this effect, trees in different stands can have the same value of canopy closure for disparate stocking conditions as well as crown diameter/Dbh ratios. As a consequence, a single estimated value for canopy closure can be the result of a range of differing stand stocking conditions.

Spectral analysis: clustering/classification approach

The raw stand data were shown to cluster into five relatively distinct groups based upon stems per acre, volume, height, and Dbh (Tables C5 and C6, Appendix C). Classification, using the spectral signatures associated with individual stands, did not force all observed stands into a single category as would be expected in a relatively homogenous population. Significant separation among stand cluster types was shown for a number of combinations of Landsat vegetation indices, as indicated by the Mahalanobis D^2 statistic (Table C14, Appendix C). For at least 50 percent of the vegetation indices used, a significant difference in covariance matrices was shown to exist, reinforcing the separation assertion. It was shown that forest stand types based on field measured mensurational parameters could be separated using spectral signature on a multispectral scanning device, Landsat, and that classification into categories using field measured and spectral data can be done.

Spectral analysis: direct regression approach

Most of the vegetation indices examined have a statistically significant, at the .05 level, association with timber stand volume using a simple linear regression model. The best model for stand volume estimation was the Kauth and Thomas brightness index (B). Individual, older, trees have a larger crown volume, but a relatively lower foliage density presented to the sensor. As a consequence, a larger soil background component is visible to the sensor, for which the brightness index was developed. In the case of the other indices, the "disguise" effect, understory vegetation appearing as overstory vegetation, degraded the overall performance of the regression models.

Spectral analysis: indirect regression approach

This analysis was preparatory to the verification analysis, consisting of estimation of crown diameter, canopy closure, and average stand height in separate regression analyses. The three most useful vegetation indices for estimating crown diameter were the difference-7 (D7), the greenness-brightness ratio (GB), and the normalized difference 7 (ND7), respectively accounting for 24.06, 23.65, and 23.23 percent of the sample variation. Each index responds to a relative increase in the infrared component, implying a relative increase in crown diameter. The "disguise" effect impacts estimation in that understory vegetation viewed near a tree canopy can force overestimation of average crown diameter across a picture element. Canopy closure had no usable candidates, due to the strong influence of the "disguise" effect.

Height was predicted unexpectedly well by the regression models posed for it. Four explanations can account for this good degree of fit. First, as height increases over time, there is within the crown more foliage contributing to the infrared reflectance component viewed by the sensor. Second, as a tree matures, crown geometry changes, presenting different profiles to the sensor. Third, physiological changes within an aging tree can contribute to differential spectral response across stand age. Fourth, stand composition can affect spectral response. If a stand has many dominant trees, additional shadow effects can be a significant contributor to a pixel's response. Smoother textured stands can have different spectral responses than rougher ones.

Spectral analysis: verification analysis

Direct and indirect methods were used to estimate volumes for a set of stands reserved from initial analysis. No significant difference was found between spectrally estimated and actual field volumes, at the .05 level. Underestimation occurred using direct methods, and overestimation occurred when using indirect methods. The primary source for the underestimation using direct methods was a combination of the "disguise" effect which drove up the estimate, and the dampening effect of the contributions from bands 4 and 5, appearing as unvegetated ground. The primary source of overestimation using indirect methods was the inflated values estimated for canopy closure due to the "disguise" effect.

Spectral analysis: other stand parameters

Additional stand parameters were examined in an effort to improve enumeration of forest stands using remotely sensed data. The stand parameters of interest were average tree diameter, per acre basal area, and the number of stems per acre. Average tree diameter had a number of useful candidates including all four Landsat bands, the normalized difference-6, difference-7 and normalized difference-7 indices. Each of these is due to the strong relationhip between crown diameter and Dbh. The "disguise" effect affects Dbh estimation by inflating an individual trees apparent crown diameter, if appearing as large crowns. The estimate can also be driven down if tree crown diameter is underestimated in a smooth textured image. Basal area in confounded in a similar manner as canopy closure, among stems per acre and individual tree diameter, presenting to the sensor a value of basal area which can result from a disparate range of stocking values. The "disguise" effect is, to a lesser extent than with canopy closure, responsible for the relatively low degree of fit for the models posed here. Understory vegetation appearing as false elements in the canopy can inflate a basal area estimate. Stems per acre apparently cannot be estimated using remote sensing data.

Overall objective summary

The objective of this study was to answer the question: Is there a quantifiable relationship between spectral signature, as represented on the Landsat satellite, and timber stand volume or those stand parameters that can be used to estimate volume through an aerial photo volume table approach?

The answer is a qualified yes.

A number of approaches used in this study have shown this relation. First, clustering of stands, based on the inherent characteristics within the stand, showed that relatively distinct categories can be distinguished with the field data. Separation analysis of these categories, using the Mahalanobis distance with stand concomitant spectral data, is significant in a large number of cases, at the .05 level (Table C14, Appendix C). Classification analysis of these stands showed the separation to be an effective one. Second, a regression analysis between volume and spectral signature showed a significant relationship between the two for two of the three APVT inputs, crown diameter and height, and for volume itself for a substantial number of vegetation indices. The lack in the third APVT input, canopy closure, is accounted for by the "disguise" effect, as is discussed in Chapter VI, which is an unavoidable consequence when using Landsat data. Volumes estimated for each direct and indirect method are not significantly different from actual field volumes in a set of stands reserved from initial analysis, and the two methods corresponded well with one another.

Enumeration of forest stand conditions is possible using Landsat data with a statistically significant level of accuracy. Over regional areas, such as forest survey regions, the amount of information that is extractable with Landsat offsets the inherent noisiness in it. The parameters showing the most promise are height, average tree diameter, basal area, crown diameter, and volume. The correlations between spectral signature and each of these stand parameters can be attributed to the interrelationships among stand parameters, specifically the spectral components that contribute to the estimation of one are highly influential in estimating another. For example, the regression models predicting crown diameter and average tree diameter are related by the strong association between the two stand parameters. Height and its relation to crown diameter and volume, can be related using a similar argument as with crown diameter and average tree diameter. Basal area can be derived from the relationship of spectral signature and crown diameter. A limitation exists on estimating basal area, stems per acre, and crown diameter using Landsat imagery because of the "disguise" effect. Understory vegetation being interpreted as overstory vegetation which can inflate or can deflate estimates, depending on stocking conditions within a stand.

In general, it was shown that volume estimation, stand enumeration, and classification of forest stands using field derived mapping criteria can be done using spectral signature from remotely sensed data.

A final point must be remphasized at this time. The purpose of this study was not to develop operational methods for use by forest managers, but to examine whether or not the relationships sought in this study existed. These relations were, with the qualifications discussed in Chapter VI, shown to exist. Implications of this study will now be discussed

Implications

Two areas of application can be directly affected by the results of this study. The first area is in the design and implementation of forest survey and the second is the development of geographic information systems. Throughout both areas, the potential for using the results of this study with alternative remote sensing devices will be discussed.

Forest survey applications

The results of this study can have a significant impact in the area of forest survey or inventory. Improving forest survey design is one major area of application of these results. Stratification of the forest population can now be performed using spectrally derived estimates of stand parameters, as well as by forest types.

A hypothetical example helps explain how this might be accomplished. First, satellite imagery of the area to be inventoried is obtained from the appropriate source. Classification into descriptive categories, i.e. pine, hardwood, and mixed, is then performed using conventional routines for classification. Following that, a set of stand parameter estimation equations based on spectral signature is used in the entire area or in its component forest types, depending on inventory objectives. Then, this "refined classification" is entered into a geographic information system (GIS), where, for instance, ownership, soils, and topography are merged with the "refined classification" into a single data file. Next, field plots are assigned to each classified category according to conventional rules of stratified sampling allocation, described by Cochran (1977). The total number of field plots is then held in reserve until a pilot survey is performed on a subarea of the inventory area. Preliminary summary statistics for the pilot study area are then computed, along with confidence levels.

Comparison of these confidence levels with those desired for the total sample is made and then confidence levels for the pilot study area are then obtained, based on spectral data and GIS information. If these new confidence levels are within desired limits for accuracy, precision, and bias, then no further field plots need be taken for the inventory area. If greater summary statistic precision is desired, then the allocation procedure is repeated to determine the number needed for the refinment in accuracy and precision desired. If the total number of plots is less than that estimated previously for complete sampling on the inventory area, then the appropriate subsample, equal to the number of plots required to refine the accuracy and precision of summary statistics, is drawn. The remaining plots are either discarded or held in reserve for special administrative purposes. In summary, the total number of field plots required for a regional survey, when using remote sensing and GIS information as inputs to the sampling design process, should be less than those required for conventional surveys.

The cost of remotely sensed data, especially from spaceborne sensors is currently prohibitive for most forest management operations. Alternate remote sensing devices exist which could mitigate this cost barrier.

Alternative sensors and GIS potentials

A potential for alternative applications of the results of this study concerns the direct video digitizing of aerial photography. Color or color infrared aerial photography can be digitized using three filters: red, green, and blue as an example. The number of gray shades desired can be obtained, and then processed through software using the results of this study. In addition to single perspective imagery, digitization using the stereoscopic information on photography can incorporate the actual canopy geometry into the image analysis process. Two images are separately recorded, capturing the effective area in a photograph, registered and merged into a single data file. The combined data files from stereo pairs can then be assembled into a mosaic over the area of interest.

Incorporation if stereoscopy can greatly enhance volume estimation, which would remove the subjectivity inherent in human photo interpretation measurement of stand parameters. A hypothetical example of the resolution possible follows using the parameters of the color infrared portion of the National High Altitude Photography (NHAP) program. The nominal representative fraction of this photography is 1:58,000. Suppose an analyst wanted to digitize this photography into a 512 by 512 pixel format. The size of an individual picture element would be 29 meters on a side which is comparable with Thematic Mapper resolution. Now further suppose that the analyst wanted to digitize a subsample of this photo coverage to a 1024 by 1024 pixel basis. The individual pixel would be 14.5 meters on a side, small enough to partition individual patches of even aged timber within the typical East Texas forest. The element of steroscopy would provide an additional three channels of data, and a fourth channel of information incorporating the actual stereoscopic element from both images. The reflectance properties of an area could then be sampled from two vantages. The total number of data channels available on an area would then be seven. Another example of commonly used photographic scales in forestry is U.S.D.A., Forest Service 1:15,840 color film for stand mapping. Digitization to a 512 by 512 pixel basis would result in a picture element size of 7 meters, small enough to detect the average crown diameter within a stand.

Two consequent image processing technique are needed. First, the incorporation of stereoscopy into the image processing analysis would require image registration using photogrammetric stereo model theory. Second, a decision making process of how stereo images are to be digitized needs developing. A systematic digitization procedure must be in place before the digitization of large numbers of photographs can be done, lest the sidelap portions of photos seriously distort the total results of a project.

The use of aerial multispectral scanners is another alternative to the use of spaceborne imaging systems. A high degree of flexibility is possible using these systems. For example, flight planning for a region of interest can be done in such a fashion as to incorporate the element of stereo into the actual image data output, by overlapping flight scanning swaths. Combined with a large number of channels, for example ultraviolet to thermal infrared, stereoscopic capability can facilitate a more complete enumeration of the forested scene of interest. Image processing can follow the same precepts as photo digitization, including those aspects needing theoretical and organizational development before actual imaging can occur. An advantage of aerial multispectral scanners over photographic digitization is that the data is actual spectral data, whereas photograph digitization records a "pseudochannel" of data, recorded under visible light.

In either case, the relationships between forest stand parameters must be evaluated for each particular imaging system and technique. An essential precursor to evaluation of any imaging system, including image processing, is a field evaluation of the radiometric interrelationships among forest stand parameters, not done in this study due to cost limits.

Suppose an analyst were interested in working in an area where the acquisition of satellite imagery would be the most useful. Two sensors, one in orbit and the other scheduled for launch in the fall of 1985, offer a substantial improvement in resolution, repetition of coverage, and number of channels over the Landsat Multispectral Scanner (MSS). The Thematic Mapper (TM) aboard Landsats 4 and 5, with its seven channels and thirty meter resolution, offers the capability to evaluate the even aged patches of timber, typical within an East Texas forest. The number of data combinations possible with TM data exceeds that for Landsat MSS by three orders of magnitude. Individually, Landsats 4 and 5 repeat coverage of an area every 16 days versus 18 days with older Landsats.

The SPOT satellite, due for launching in the fall of 1985, is another satellite meriting consideration for the application of the results of this study. It has the capability for 10 meter resolution in its panchromatic mode, and 20 meter resolution in its multispectral mode. In addition to the fine level of resolution, the sensor has the capability to pivot off of nadir, allowing multiple views of an area. This genuine stereo capability can allow for penetration of a forest stand by sampling the reflectance properties from multiple vantages. The net effect is to double, treble, and even quadruple the number of channels available to an analyst, over and above the three obtainable from one pass.

Each spaceborne sensor has data which is formatted in a unique style for which an image processing system must be compatible. This drawback to spaceborne sensors could be circumvented using the alternative sensors just discussed.

In all of the cases discussed, the results of this study are applicable. The airborne level approach, airborne scanners and the digitization of aerial photography, need the necessary stereo modelling theory development in addition to calibration of these results for their application to alternative sensors. However, the flexibility available in selecting coverage of an area of interest offsets the extra development step needed for their application and should free that analyst from the rigid schedule associated with spaceborne imagery, and its consequent gamble on image quality, cloud cover, and contrast.

This argument must not be construed as advocacy for discontinuing the use of satellite imagery. A study involving large regional areas, such as East Texas, using imagery derived from the alternatives just discussed, will encounter prohibitive costs, especially when special mission planning is required. The cost of imagery obtained aerially versus from space must be estimated before an inventory can proceed. In small areas, for instance single counties or small blocks of counties, the airborne approach would prove more cost effective, and over larger areas such as entire forest survey regions, using spaceborne imagery would prove best.

Incorporation of these results into the rapidly expanding technology of geographic information systems can have substantial impact as well. Forest stand parameter enumeration using the results of this study, or refined developments from them, when incorporated into the timber layer of a forest management GIS, can allow "in-house" examination of stand conditions before field surveys take place. By using added layers of spatial information available on a GIS, the design of both intensive inventories and extensive surveys can be refined to a degree heretofore unrealized. Ownership, topography, soils, previous timber surveys, and other resource data can be incorporated within the design scheme. A multivariate stratification design is thus possible with the many layers of a GIS. When combined with estimates of stand volume, and other stand parameters, available from remotely sensed information, field plot allocation can proceed on the basis of seeking the appropriate subsample, rather than a complete survey enumeration.

Graphical presentation of regional summary statistics could be done using the spatial portion of the GIS data base. Conventional tabular summaries could be extracted from the attribute portion of the data base. Combined regional summaries could be presented graphically and in tabular form using the spatial and attribute data files simultaneously. Forest management scenarios can be performed within the data base as well. Finally, the updating capability of forest stand parameter information possible using remote sensing data can, when combined in a GIS, provide the forest manager with a flexibility in decision making unavailable in the past.

A last point concerns the cost analysis of whether or not a particular remote sensing system should be used, whether or not alternative survey/inventory designs should be used, and whether or not the data should be included in a GIS. Developing and testing of the most cost efficient combination of sensors, data bases, and survey designs is an area of research essential to the application of the results of this study.

The use of GIS technology, in concert with an appropriate combination of sensors, is the most promising approach for improved survey design. In small regional areas, such as single counties or small blocks of counties, airborne sensing devices such as aerial multispectral scanners or photo digitization are likely to be the cost effective approach to acquiring the remotely sensed contribution to the geographical data base. For larger areas such as survey regions, satellite imagery of proper timing and qualtiy can contribute the remotely sensed part of the GIS data base.

In summary, the results of this study point towards two areas of potential applicability: (1) utility for design and implementation of regional forest surveys, and (2) as an input to a GIS data base, including forest survey applications.

The major areas of research for future analysts include: (1) an analysis of the costs and benefits of using remote sensing as a major input to forest survey,(2), adaptation of these results to other sensors, and (3), the potential incorporation of these results into the development and use of operational geographic information systems. The demonstration of a significant relationship between forest stand parameters and spectral signature has shown the way for evaluation and use of remote sensing technology in the quantification of forest stands not previously available to the forest researcher.

Analytical recommendations

The model derived in the APVT analysis, which was in good agreement with previous work in southeastern forests, can be improved. First, measurements of crown diameter and canopy closure should be done in a more objective manner. Crown diameter can either be measured in the field or analyzed using digitized aerial photography. Canopy closure can be measured on photography using a systematic method, for example a fine dot grid, or evaluated by digitizing aerial photography. Height, measured in an indirect fashion in this study, should be either field measured, photo measured or estimated in an automated manner. Alternative sensing methods for the determination of crown diameter, height, and potentially canopy closure are currently being investigated, including as an example the pulsed airborne laser profiler. The stereoscopic model characteristics available in the airborne multispectral scanner, camera digitization of aerial photography, and the SPOT satellite can be used to estimate height without field measurements, apart from calibrating field verification. When combined with crown diameter and canopy closure measurements that are available from photographic measurements, a direct volume estimate, using an APVT, is possible without the need for human interpretation and its consequent potential for bias introduction. When these volume estimates are compiled over a regional area, field plots can be drawn using a subsampling approach rather than a stratified design. The total number of field measurements required for implementation of the survey could be substantially reduced over conventional techniques. A reduction of even ten percent, on a regional area, can effect considerable cost savings.

The actual development of an aerial volume table equation should consider as many potential dependent terms as possible. Some examples of this include: logarithmic independent terms, higher order independent terms such as cubic ones, and even nonlinear regression.

The use of canopy closure is, however, marginally useful with remote sensing devices, as an estimating factor for volume and other forest stand parameters. Its confounded nature, among stems per acre and individual tree size, introduces additional variability not needed in the analysis. Development of an index of stocking which is not confounded as canopy closure and basal area, or subject to extreme outliers as is stems per acre, is a useful direction for research with remote sensing devices.

With respect to the spectral signature portion of the analysis, several suggestions are in order. Higher order terms can be tried in regressing spectral signature against forest stand parameters. Forest stand parameters could be transformed, using logarithms for example, before fitting them to spectral signature. Various multivariate transformations, not attempted in this study due to time constraints, can be tried. Factor analysis techniques could be used to extract linear combinations from the raw data, as an alternative to principal component analysis or Gram-Schmidt orthogonalization (Rao 1973). Combinations of vegetation indices can be used on the independent side of regression models. An illustrative example of this might be band 7 to band 5 ratio and the band 6 to band 4 ratio.

Classification of stands for field based parameter maps can be done using operational forest management mapping categories rather than clustering based on the inherent "clumpings" that occur within a given data set. Other classification techniques can be tried for separating stands, including those not explicitly requiring a multivariate normal distributional assumption, such as the nearest neighbor and parallelpiped techniques (Mardia et al. 1979, Johnson and Wichern 1982, American Society of Photogrammetry 1983). Incorporation of the information from a GIS data base can be used as ancillary information prior to and during classification analysis.

Other sensing platforms can be used to evaluate the results of this study, after applying the appropriate calibrations for a given sensor. Those covered in the implication section include the Thematic Mapper (TM), SPOT, airborne multispectral scanners, and the digitization of aerial photography. Sensing in other regions of the electromagnetic spectrum should also be explored, such as with the thermal infrared, passive, and active microwave (radar). The thermal infrared region is sensed by band 7 on the Thematic Mapper device. Active and passive microwave have been shown to be useful in analysis of soils properties, however their use, and that of the thermal infrared, has only been tested for use in forests in a cursory manner (American Society of Photogrammetry 1983).

Corroboration of these results is the essential first step before any attempt at field applications. Since a small sample size in a limited regional area was used in this study, replication beyond the study area is essential before these results can be refined for operational use. The results of this study, nevertheless, opens avenues for future research which should be explored in depth, due to their considerable potential.

Chapter VIII.

CONCLUSIONS

Landsat Multispectral Scanner (MSS) imagery can be used to estimate timber volume, and other forest stand parameters, such as basal area, average stand height, and average tree diameter, with a statistically significant level of accuracy. The low cost of Landsat, relative to current forest survey techniques, would allow estimation of forest stand conditions over large areas at substantial savings.

If remotely sensed data were incorporated into the design and implementation of forest survey, more detailed mapping of forest stands could be done. Timber stand conditions could be used as mapping criteria in addition to forest types. This could be done over and above the current forest survey mapping categories of forest, nonforest, and two water categories.

The repetitive coverage of the Landsat satellite enables analysis of changes at a greater frequency than is currently done. For example, change detection can be done at a frequency greater than the five year midcycle updates done for the U.S.D.A., Forest Service regional surveys, and would be at little, if any, increasing in costs. The decrease in the number of field plots required for a particular survey could be reduced substantially, further reducing costs.

Newer sensors, such as the Thematic Mapper, can allow mapping at even finer detail than with the Multispectral Scanner. The potential cost savings with these finer resolution sensors can result in both finer detailed mappings and further reduction in the number of field plots required for the survey. The savings could easily offset the additional costs associated with acquisition of data from these sensors.

The use of alternative sensing systems such as aerial multispectral scanners and digitized aerial photography offer an economical alternative to satellite systems. The results of this research are also applicable to these systems. The use of aerial photograph digitizing will reduce the error associated with human photo interpretation, and reduce image processing cost. High altitude photos could be computer evaluated, performing both the image processing and forest survey stratification simultaneously, further reducing costs.

Incorporating geographic information systems with the results of this study can facilitate the long term storage of survey data in a far cheaper and less perishable form than with paper map storage. It can also contribute to the forest survey process by allowing graphical presentation of summary statistics, and incorporating spatial data into the survey design process which can further reduce the number of plots required for field implementation of the survey.

The potential of using remote sensing techniques to evaluate and analyze forest stands promises substantial cost savings in forest survey and inventory. Future research should be in two directions: (1), examining and adapting the results of this study to other remote sensing devices such as computer analysis of digitized aerial photographs, and (2), analyzing the costs and benefits of performing a forest survey based on remote sensing. The potential economies offered by using remote sensing techniques in forest inventory and survey is sufficient justification for further work in this area of research.

REFERENCES

Aaronson, A.J., and L.L. Davis, 1979. An evaluation of relationships between vegetative indices, soil moisture, and wheat yields. Technical Memorandum Number 9, Crop Condition Assessment Division, U.S.D.A./F.A.S., Houston, Texas.

Aggers, L.W., and E.B. Kelley, 1976. Douglas County cover condition mapping and forest volume inventory. Proceedings of the American Society of Photogrammetry, Fall Convention. Seattle, Washington. pp. 218-229.

Aldrich, Robert C., 1967. Stratifying photo plots into volume classes by crown closure comparator. U.S.D.A., Forest Service Research Note PSW-151.

__________________, 1979. Remote sensing of wildland resources: A state of the art review. U.S.D.A., Forest Service General Technical Report RM-71. 56 p.

Aldrich, Robert C., and Nancy X. Norick, 1969. Stratifying stand volume on non-stereo photos. U.S.D.A., Forest Service Research Paper PSW-51. 14 p.

American Society of Photogrammetry, 1983. Manual of Remote Sensing (second edition). American Society of Photogrammetry, Falls Church, Virginia. 2440 p.

Anderberg, M.R., 1973. Cluster Analysis for Applications. Academic Press, New York, New York. 359 p.

Avery, T.E., 1958. Composite aerial volume table for southern pines and hardwoods. Journal of Forestry, 56(10), pp. 741-745.

__________, 1979. Forester's Guide to Aerial Photo Interpretation. U.S. Department of Agriculture, Agriculture Handbook 308. 41 p.

Avery, T.E., and David Myhre, 1959. Composite aerial volume table for southern Arkansas. U.S.D.A., Forest Service, Southern Forest Experiment Station Occasional Paper 172. 9 p.

Boyd, William E., 1984. Determining rengeland bursh canopy densities with Landsat MSS data. Ph.D. Dissertation, Texas A&M University, College Station, Texas. 74 p.

Cochran, W.G., 1977. Sampling Techniques, third edition. John Wiley and Sons, New York. 428 p.

Colwell, R.N., and S.J. Titus, 1976. Sam Houston National Forest inventory and development of a survey planning model. Final Report NASA Contract 9-1445Z. Space Science Laboratory, University of California, Berkeley.

Dilworth, J.R., 1977. Log Scaling and Timber Cruising. O.S.U. Bookstore, Corvallis, Oregon. 471 p.

Finley, Robert J., 1979. Landsat analysis of the Texas coastal zone. Bureau of Economic Geology Report of Investigations No. 93. The University of Texas at Austin. 71 p.

Fleming, Michael D., and Roger M. Hoffer, 1979. Machine processing of Landsat MSS and DMA topographic data for forest cover type mapping. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, Indiana. pp. 377-391

Harding, Roger A., and Robert B. Scott, 1978. Forest inventory with Landsat. Phase II Washington Forest Productivity Study. Division of Technical Services, Department of Natural Resources, Olympia, Washington. 221 p.

Hay, C.M., Kuretz, C.A., Odenweller, J.B., Schefner, E.J., and B. Wood, 1979. Development of AI procedures for dealing with the effects of episodal events on crop temporal spectral response. AgRISTARS, SR-B9-00434, Contract NAS 9-14565.

Heath, G.R., 1974. ERTS data tested for forestry applications. Photogrammetric Engineering. 40(11), pp. 1087-1091.

Heller, R.C., (technical coordinator) 1975. Evaluation of ERTS-1 data for forest and rangeland survey. U.S.D.A., Forest Service Research Paper PSW-112. 67 p.

Helms, J.R., and W.A. Shain, 1981. Problems related to the use of remote sensing for inventory and mapping of lower coastal plain forests. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, Indiana. pp. 540-543.

Hocking, R.R., 1976. The analysis and selection of variables in a linear regression analysis. Biometrics. 32(1), pp. 1-50.

Hoffer, R.M., and LARS Staff, 1973. Techniques for computer-aided analysis of ERTS-1 data, useful in geologic, forest, and water resource surveys. Proceedings of the Third Earth Resources Technology Satellite-1 symposium. Volume 1, Section A. Goddard Space Flight Center, Washington, D.C.. pp. 1687-1708.

Hoffer, R.M., and Staff, 1975a. Natural resource mapping in mountainous terrain by computer analysis of ERTS-1 satellite data. Agricultural Experimental Station Research Bulletin 919. Purdue University, West Lafayette, Indiana. 124 p.

—————————, 1975b. Computer-aided analysis of SKYLAB multispectral scanner data in mountainous terrain for land-use, forestry, water resource, and geologic applications. LARS Technical Report 121275. Purdue University, West Lafayette, Indiana.

Hoffer, R.M., Fleming, M.D., Bartolucci, L.A., Davis, S.M., and R.F. Nelson, 1979. Digital processing of Landsat MSS and topographic data to improve capabilities for computerized mapping of forest cover types. LARS Technical Report 011579. Purdue University, West Lafayette, Indiana. 159 p.

Husch, B., Miller, C.I., and T.W. Beers, 1972. Forest Mensuration, second edition. Ronald Press, New York. 410 p.

Johnson, R.A., and D.W. Wichern, 1982. Applied Multivariate Statistical Methods. Prentice-Hall, Englewood Cliffs, New Jersey. 594 p.

Kauth, R.J., and G.S. Thomas, 1976. The "Tasseled Cap"- A graphic description of spectral temporal development of agricultural crops as seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, Indiana. pp. 4B41-4B51.

Kshirsagar, A.M., 1972. Multivariate Analysis. Marcel Dekker, Inc., New York. 534 p.

Lachenbruch, P.A., 1975. Discriminant Analysis. Hafner, New York. 128 p.

—————————, 1979. Discriminant analysis. Biometrics. 35(1), pp. 69-85.

Landgrebe, D.A., Hoffer, R.M., Goodrick, F.E., and Staff, 1972. An early analysis of ERTS-1 data. Proceedings of the NASA Symposium on Preliminary Results of ERTS Data Analysis. Goddard Space Flight Center, Washington D.C..

Lautenschlager, L.F., and C.R. Perry, 1981. Comparison of vegetation indices based on satellite acquired data. Proceedings of the Section on Survey Research Methods of the American Statistical Association, August 10-13, 1981, Detroit, Michigan. pp. 77-82.

Mardia, K.V., Kent, J.T., and J.M. Bibby, 1979. Multivariate Analysis. Academic Press, New York. 521 p.

Mazade, A.V. (editor), 1981. Ten Ecosystem Study: Final Report. Lockheed Engineering and Management Services, Inc., Houston, Texas. 83 p.

Mead, R.A., Driscoll, R.S., and J.A. Smith, 1979a. Effects of tree distribution and canopy cover on classification of ponderosa pine forest from Landsat-1 data. U.S.D.A., Forest Service Research Note RM-375. 2 p.

Mead, R.A., and James L. Smith, 1979 b. Development of an aerial volume table for trees infested with southern pine beetles. Report 80-1 School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg. 33 p.

Misra, P.N., Wheeler, S.G., and R.E. Oliver, 1977a. Kauth and Thomas brightness and greenness axes. IBM personal communication, Contract NAS 9-14530, RES 23-46.

Misra, P.N., and S.G. Wheeler, 1977b. Landsat data from agricultural sites-crop signature analysis. Proceedings of the 11th International Symposium on Remote Sensing of the Environment, ERIM, University of Michigan, Ann Arbor.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. John Wiley and Sons., New York. 673 p.

Myers, R.H., 1976. Response Surface Methodology. Virginia Polytechnic Institute and State University, Blacksburg. 246 p.

Nelson, R.F., and R.M. Hoffer, 1979. Computer-aided processing of Landsat MSS data for classification of forestlands. LARS Technical Report 102679. Purdue University, West Lafayette, Indiana. 94 p.

Neter, John, Wasserman, William, and M.H. Kutner, 1983. Applied Linear Regression Models. Richard D. Irwin, Inc., Homewood, Illinois. 547p.

Oregon State Dept. of Forestry, 1978. Douglas County forest condition mapping and forest inventory project. Final Report. Pacific Northwest Regional Commission, Salem, Oregon.

Perry, C.R., and L.F. Lautenschlager, 1984. Functional equivalence of spectral vegetation indices. Remote Sensing of Environment. 14(2), pp. 169-182.

Pope, Robert B., 1962. Constructing aerial photo volume tables. U.S.D.A., Forest Service Research Paper PNW-49. 25 p.

Rao, C.R., 1973. Linear Statistical Inference and its Applications, second edition. John Wiley and Sons, New York. 522 p.

Richardson, A.J., and C.L. Wiegand, 1977. Distinguishing vegetation from soil background information. Photogrammetric Engineering and Remote Sensing. 43(11), pp. 1541-1552.

SAS Institute, Inc., 1982. SAS Users Guide: Statistics, 1982 edition SAS Institute, Inc., Cary, North Carolina. 584 p.

Schumacher, F.X., and T.S. Coile, 1960. <u>Growth and Yields in Natural Stands of Southern Pines</u>. T.S. Coile, Inc., Durham, North Carolina. 115 p.

Strahler, A.H., Logan, T.L., and N.A. Bryant, 1978. Improving forest cover classification accuracy from Landsat by incorporating topographic information. Proceedings of the 12th International Symposium on Remote Sensing of the Environment. ERIM, University of Michigan, Ann Arbor.

Strahler, A.H., Franklin, J., Woodcock, C.E., and T.L. Logan, 1981. FOCIS: A forest classification and inventory system using Landsat and digital terrain data. U.S.D.A., Forest Service, State and Private Forestry Applications Program Report NFAP-255.

Texas Natural Resources Information System, 1979. Digital processing of Landsat data for land cover/land use mapping on the Texas coast. Support Activities Report 012. TNRIS Task Force, Austin, Texas. 30p.

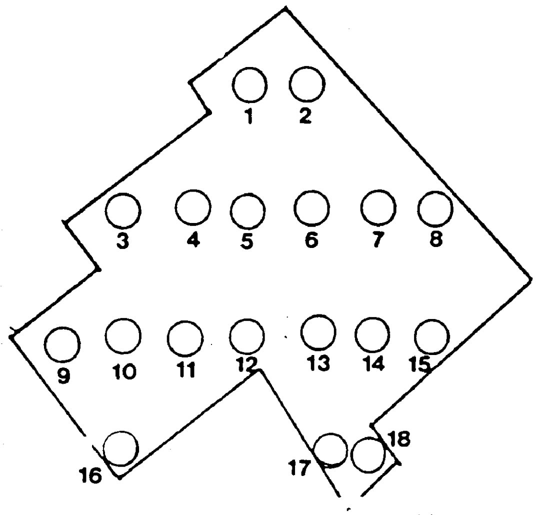
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 8(1), pp. 127-150.

Wheeler, S.G., Misra, P.N., and A.Q. Holmes, 1976. Linear dimensionality of Landsat data with implications for classification. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, Indiana.

Williams, Darrel L., 1976. A canopy related stratification of a southern pine forest using Landsat digital data. Report Number X-923-76-188, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland. 10 p.

APPENDIX A.

AN EXAMPLE OF THE FIELD DATA AND STAND MAPS USED IN THIS STUDY.



Data used from this stand map

Size Class	Volume Cords/Acre	Stems per Acre	Basal Area per Acre
5-6	1.69	41.32	8.33
7-10	6.38	49.82	21.67
11-12	5.34	19.42	15.56
13-18	12.67	22.30	28.89
19-24	4.71	4.69	10.00
25+	0.38	0.23	1.67

Site Index = 94

Figure A1. An example of stand maps and field data used in this study.

APPENDIX B.

DETAILS OF THE TEST FOR SAMPLING ADEQUACY DONE IN THE STUDY AREA.

A total of 448 stands within the study area were considered as potential candidates for selection and analysis. Since the actual selection procedure was not done in a randomized fashion, but was based on photographic location, a total of 146 stands, an almost one third sample, were chosen for initial analysis.

An initial design criterion was to avoid pine plantations, their young age and consequently low volumes. It was desired to test whether or not the stands selected were representative of the overall population of candidate stands.

Three steps were necessary for this test. First, selection of an appropriate set of variables on which to test was to be done was performed. Basal area for both pine and hardwood was the final choice. Its confounded nature, among stems per acre and individual tree size, covers the entire range of stand situations encountered, without extreme outliers which can occur with other stand parameters such as stems per acre.

Second, a test for equality of covariance matrices was done. The typical two sample test of the equality of means is predicated on the equality of variances or covariance matrices. When a situation arises where unequal covariance matrices exist, the mean comparison test statistics have undetermined and vague distributional porperties known as the Behrens-Fisher problem (Rao 1973). Third, and finally, the test for equality of mean vectors was done.

The test for covariance matrix equality used in this analysis was the same, asymptotic, approximation of the likelihood ratio test used in the clustering/classification analysis. The test treated the study stands and what will be called the remainder stands as classification categories.

The frequency of study and remainder stands (Table B1), the results of the chi square test for covariance matrix homogeneity, and the results of the stand classification in the form of a confusion matrix are presented (Table B2). The test for homogeneity does not reject the null hypothesis that the cavariance matrices are equal, at a large level of significance. The vast majority of stands were classified into type one, the remainder category, indicating that as a whole, all stands selected for analysis are drawn from the same population.

Table B1. Frequency of study stands with respect to the available candidates, and the remainder stands.

Stand Type	Frequency	Prior Probability
1	302	.67410714
2	146	.32589286
Total	448	1.00000000

Note: Stand type 1 is the remainder category, and stand type 2 the stands selected for analysis.

Table B2. Test of covariance matrix homogeneity and classification results between the study stands and the remainder stands.

Test of covariance matrix homogeneity

Chi Square	d.f.	Prob(Chi^2)
1.25555	3	.7397

Classification results

Stand Type	1	2	Total
1	297	5	302
2	142	4	146
Total	439	9	448

Since covariance matrices are equal, interpretation of the mean tests are clearcut. Two tests on the means of basal area were done. First, pine and hardwood basal area were jointly compared in a multivariate analysis of variance using stand type as treatments. Stand type is defined as: (1), the stands that were not selected for analysis, and (2), those selected for analysis. The second test compared pine and hardwood basal areas separately in a univariate analysis of variance.

The results of both multivariate and univariate analyses show a significant difference between the study and remainder stands (Table B3). When examining the individual basal area analyses, the explanation for the difference emerges. Pine basal area is significantly different between stand types whereas hardwood basal area is not. The selection of study stands was designed to use natural stands wherever possible. This forced plantation stands into the remainder category, causing the significant difference. Industrial forest management in East Texas is proceeding in a direction emphasizes conversion from naturally occuring mixed stands towards pine plantations, further contributing to the difference. It was deemed that the study stands were for practical purposes sufficiently representative of the area to proceed with the analysis.

Table B3. Hypothesis results testing for equality of mean basal area values for between the study and remainder stands.

Multivariate results
Hotelling-Lawley trace = .02285756
Pillai's trace = .02234676
Wilk's criterion (U) = .97735324
Approximate F (2,445 d.f.) = 5.09
Prob(F) = .0065

Univariate comparisons

Variable	Mean Square Factor	Mean Square Error	$F_{(1,446)}$	Prob(F)	R^2
Pine BA	6639.78	706.922	9.39	.0023	.02063
Hdwd BA	25.99	264.579	0.10	.7541	.00022

APPENDIX C.

MISCELLANEOUS TABLES REFERRED TO FROM THE MAIN TEXT.

Table C1. Expected crown diameter APVT models.

Dependent Variable	$F_{(5,141 \text{ d.f.})}$	Prob(F)	Mean Square Error
Volume	62.038	.0001	29.373
ln(Volume)	95.789	.0001	0.068

Dependent Variable	R^2	Coefficient of Variation	PRESS mean
Volume	.6875	27.674	0.0372
ln(Volume)	.7726	9.162	0.0029

Dependent Variable	PRESS standard error	PRESS variance
Volume	.44199	27.7175
ln(Volume)	.02225	0.0728

Regression coefficient estimates

Coefficient	Volume	ln(Volume)
b_0	-68.4212	1.3591
b_1	1.0218	0.0124
b_2	0.0509	-0.0259
b_3	1.6106	-0.0900
b_4	-0.0475	2.3469×10^{-5}
b_5	3.3875×10^{-6}	8.8060×10^{-8}

Note: The model form is the same as with the final APVT with expected crown diameter substituted for crown diameter. See Tables 4 and 6 for variable definitions and the final APVT models, respectively.

Table C2. Expected canopy closure APVT models.

Dependent Variable	$F_{(5,141 \text{ d.f.})}$	Prob(F)	Mean Square Error
Volume	22.107	.0001	53.341
ln(Volume)	16.325	.0001	0.190

Dependent Variable	R^2	Coefficient of Variation	PRESS mean
Volume	.4394	37.184	0.3191
ln(Volume)	.3667	15.289	0.0228

Dependent Variable	PRESS standard error	PRESS variance
Volume	0.63670	59.6011
ln(Volume)	0.03702	0.2015

Regression coefficient estimates

Coefficient	Volume	ln(Volume)
b_0	171.0470	12.5370
b_1	-16.3850	-1.1360
b_2	-0.9640	-0.0440
b_3	8.2130	0.3600
b_4	-0.1510	0.0052
b_5	-2.7403×10^{-5}	-5.6480×10^{-6}

Note: The model form is the same as with the final APVT with expected canopy closure substituted for canopy closure. See Tables 4 and 6 for variable definitions and the final APVT models, respectively.

Table C3. Response surface of canopy closure on stand parameters

Equation form: $CC = b_0 + b_1 \times \text{Parameter} + b_2 \times \text{Parameter}^2$

Regression coefficient estimates, with 144 d.f..

Parameter	b_0	t statistic	Prob(t)
Stems per acre	317.46	2.23	.C272
Tree diameter	6.45	1.75	.0826
Height	-13.29	-0.55	.5862
Crown diameter	11.56	2.57	.0111
Basal area	102.28	1.77	.0791

Parameter	b_1	t statistic	Prob(t)
Stems per acre	-8.60	-2.06	.0416
Tree diameter	0.08	0.75	.4565
Height	2.47	3.45	.0007
Crown diameter	0.09	0.20	.8453
Basal area	-2.46	-1.50	.1359

Parameter	b_2	t statistic	Prob(t)
Stems per acre	0.08	2.63	.0095
Tree diameter	-0.0005	-0.65	.5137
Height	-0.016	-3.06	.0027
Crown diameter	-0.00009	-0.09	.9309
Basal area	0.0256	2.14	.0341

Table C3 (cont.)

Linear term contribution to overall regression significance

Stand Parameter	$F_{(L)}$	Prob(F)	$R^2_{(L)}$
Stems per acre	22.34	.0001	.1290
Tree diameter	0.68	.4124	.0046
Height	12.23	.0006	.0744
Crown diameter	0.87	.3538	.0060
Basal area	28.31	.0001	.1518

Quadratic contribution to the overall regression significance

Stand Parameter	$F_{(Q)}$	Prob(F)	$R^2_{(Q)}$
Stems per acre	6.91	.0095	.0399
Tree diameter	0.43	.5317	.0030
Height	9.35	.0027	.0564
Crown diameter	0.01	.9309	.0000
Basal area	4.58	.0259	.0341

Note: The F statistics are with 1 and 144 d.f..

Analysis of variance summary of stand parameter significance across levels of canopy closure.

Stand Parameter	$F_{(overall)}$	Prob(F)	$R^2_{(overall)}$
Stems per acre	14.63	.0001	.1688
Tree diameter	0.55	.5769	.0076
Height	10.83	.0001	.1308
Crown diameter	0.44	.6471	.0060
Basal area	16.44	.0001	.1859

Note: The F statistics are with 2 and 144 d.f..

Table C4. Canonical correlations between canopy closure and other stand parameters.

A. Canonical correlations of stems per acre, height, and their crossproduct, with canopy closure.

Canonical Correlation	$F_{(3,143)}$	Prob(F)
.4625033	12.9709	1.53×10^{-7}

Standardized canonical coefficients

	Stems	Height	Stems X Height
Coefficient	2.9698	1.0849	-2.3821
Corr (:, CC)	.3591	.2727	.3778

B. Canonical correlations of stems per acre, tree diameter, crown diameter, and basal area with canopy closure.

Canonical Correlation	$F_{(3,143)}$	Prob(F)
.448881	7.1160	5.79×10^{-6}

Standardized canonical coefficients, listing correlations with canopy closure on the second line.

Stems	Dbh	Height	CD	BA
1.1374	0.4240	0.3881	0.0841	-0.2520
.3591	.0682	.2727	.0773	.4001

Note: CD = crown diameter, and BA = basal area.

Table C5. Cluster analysis summary for the original stand data set: principal component scores.

Principal component analysis

Variable	Eigenvector 1	2	3	4	Proportion #	
DBH	.383	-.618	.536	.427	1	.4999
Height	.592	-.203	-.767	.139	2	.3739
Stems	.290	.724	.143	.608	3	.1001
Volume	.647	.227	.319	-.653	4	.0259

Cluster membership summary

Cluster Number	(component 1 only) Mean	# of stands
1	-2.065	1
2	-5.047	1
3	-0.745	26
4	-0.339	14
5	1.015	3
6	1.329	37
7	-0.439	39
8	3.967	4
9	-1.751	20
10	-0.821	20

Table C6. Cluster analysis summary for the original stand data: raw data.

Cluster Number	Means: Dbh	Height	Stems	Volume	# of stands
1	7.76	78.85	319.65	30.38	1
2	8.80	24.39	31.54	4.29	1
3	9.58	82.75	121.27	20.10	41
4	9.89	105.46	262.61	55.33	1
5	10.55	76.72	43.61	8.22	13
6	9.19	84.75	200.83	31.46	8
7	7.17	69.14	535.45	40.22	1
8	9.89	79.56	87.85	14.45	38
9	9.16	80.70	160.26	29.36	41
10	7.63	71.55	239.34	21.23	2

Note: Volume is in cords per acre, Dbh in inches, height in feet, and stems equals the number of stems per acre.

Table C7. Natural logarithms of cluster type 1 (based on transformed data) covariance matrix determinants. Chi square tests for covariance matrix homogeneity are with five degrees of freedom, except all four bands which is with 50 degrees of freedom.

Cluster Number	Vegetation Indices			
	D7	D6	DD	R7
3	0.9475	-0.2128	2.3441	-3.2660
4	0.9141	1.5650	4.4752	-1.6370
6	2.1620	1.6600	4.9526	-2.2910
7	2.3330	0.6790	2.7936	-1.5610
8	2.9900	1.5910	4.4851	-1.3670
9	1.0360	0.0083	3.2379	-2.7620
Pooled Covariance Matrix	1.7940	0.8769	4.0378	-2.2170
Chi Square Statistic	13.8170	18.7544	24.7370	13.1140
$Prob(Chi^2)$	.0168	.0021	.0001	.0223
Classification Rule Used	Q	Q	Q	Q

Cluster Number	Vegetation Indices			
	R6	ND7	ND6	G
3	-4.0950	-6.3590	-7.3370	-0.1454
4	-2.3110	-5.3640	-6.0540	0.4411
6	-2.7680	-6.1130	-6.5110	1.8270
7	-2.5310	-4.9320	-5.9910	1.1484
8	-2.1830	-5.1840	-5.9150	2.0226
9	-3.4380	-6.0450	-6.8690	-0.0013
Pooled Covariance Matrix	-2.9770	-5.6700	-6.5190	1.0111
Chi Square Statistic	11.4880	10.8410	7.7977	21.8590
$Prob(Chi^2)$	.0425	.0546	.1677	.0006
Classification Rule Used	Q	L	L	Q

Note: Q denotes a quadratic rule, and L a linear rule.

Table C7 (cont.)

Cluster Number	Vegetation Indices B	GB	PRIN1	BANDS
3	2.5933	-6.6880	-0.3550	-2.4790
4	4.3040	-5.5520	1.4461	1.9850
6	2.9040	-6.5780	0.2852	-1.1540
7	4.0570	-5.1700	1.0780	1.1800
8	-11.3400	-5.7200	-2.6510	-0.5450 (*)
9	3.1158	-6.3900	0.2407	1.6510
Pooled Covariance Matrix	3.3990	-5.6980	0.5155	1.7298
Chi Square Statistic	37.3920	13.6610	15.3530	115.0970
Prob(Chi^2)	.0001	.0179	.0090	.0001
Classification Rule used	Q	Q	Q	Q

Note: Q denotes a quadratic rule, and L a linear rule.

(*) Note: Due to a degenerate covariance matrix, only band 4 was used in this calculation.

Table C8. Natural logarithms of cluster type 2 (based on the raw data) covariance matrix determinants. Chi square tests for covariance matrix homogeneity are with four degrees of freedom, except all four bands which is with 40 degrees of freedom.

Cluster Number	Vegetation Indices			
	D7	D6	DD	R7
3	2.2430	0.0845	3.1670	-1.2700
5	1.8510	1.2152	4.3370	-4.1590
6	2.9180	0.0788	3.9260	-1.3130
8	2.4000	0.4880	3.5740	-1.4340
9	2.6750	1.3030	4.4920	-1.2130
Pooled Covariance Matrix	2.4720	0.8379	3.9840	-1.3690
Chi Square Statistic	2.5080	8.5050	10.1230	11.7430
Prob(Chi^2)	.6432	.0747	.0384	.0194
Classification Rule used	L	L	Q	Q

Cluster Number	Vegetation Indices			
	R6	ND7	ND6	G
3	-2.1160	-4.6520	-5.5190	0.7475
5	-4.7050	-7.0130	-7.8700	1.3970
6	-2.2140	-4.7390	-5.7260	1.9710
8	-2.2690	-4.8390	-5.7890	1.1734
9	-2.0020	-4.6090	-5.5020	1.7462
Pooled Covariance Matrix	-2.1920	-4.7620	-5.6890	1.3960
Chi Square Statistic	10.4470	8.9290	8.8760	5.7260
Prob(Chi^2)	.0335	.0629	.0667	.2205
Classification Rule used	Q	L	L	L

Table C8 (cont.)

Cluster Number	Vegetation Indices			
	B	GB	PRIN1	BANDS
3	4.2860	-4.9420	1.3090	0.0997
5	1.7350	-7.5490	-0.7290	-5.7250
6	3.6820	-4.9030	0.5980	2.7830
8	4.0470	-5.0770	1.0650	2.3670
9	4.2929	-4.9020	1.3540	3.3950
Pooled Covariance Matrix	4.1223	-5.0330	1.1594	3.3950
Chi Square Statistic	10.5450	10.1240	8.3800	75.8980
Prob(Chi^2)	.0322	.0384	.0786	.0001
Classification Rule used	Q	Q	L	Q

Note: Q denotes a quadratic rule, and L a linear rule.

Table C9. Means of vegetation indices, including the raw Landsat data and principal component scores, by cluster type and for the entire reduced data set.

Mean Type	Vegetation Index					
	B4	B5	B6	B7	R7	R6
Grand	21.00	17.84	39.88	37.74	2.23	2.31
Cluster Number	Cluster Type 1					
1	26.47	22.76	44.92	39.40	1.73	1.97
3	25.35	21.08	42.73	37.97	1.82	2.04
4	16.79	14.79	38.01	38.30	2.68	2.63
6	14.60	13.11	34.41	35.80	2.76	2.65
7	20.24	17.38	39.77	38.56	2.31	2.34
8	15.25	14.07	36.38	37.45	2.69	2.60
9	25.35	20.87	43.34	38.54	1.87	2.10
10	13.40	11.20	32.53	34.31	3.06	2.90
	Cluster Type 2					
1	13.40	11.20	32.53	34.31	3.06	2.90
3	21.08	17.75	39.52	37.23	2.22	2.31
5	26.33	21.60	45.02	39.98	1.85	2.09
6	19.68	17.11	39.79	39.15	2.37	2.38
7	26.47	22.76	44.92	39.40	1.73	1.97
8	20.44	17.50	40.02	38.42	2.30	2.35
9	20.29	17.34	38.74	36.74	2.25	2.32
10	25.71	22.04	42.93	37.69	1.71	1.95

Table C9 (cont.)

Mean Type	ND7	ND6	G	B	GB	D7
	Vegetation Index					
Grand	0.37	0.39	19.94	54.61	0.38	19.89
Cluster Number	Cluster Type 1					
1	0.27	0.33	18.71	62.23	0.30	16.64
3	0.29	0.34	18.31	59.99	0.31	16.89
4	0.45	0.45	22.29	50.24	0.45	23.51
6	0.46	0.45	20.97	45.42	0.46	22.69
7	0.38	0.40	20.72	54.22	0.39	21.27
8	0.45	0.44	21.93	47.98	0.46	23.38
9	0.30	0.35	19.01	60.43	0.32	17.68
10	0.51	0.49	20.91	42.21	0.50	23.11
	Cluster Type 2					
1	0.51	0.49	20.91	42.21	0.50	23.11
3	0.36	0.39	19.58	54.20	0.37	19.49
5	0.30	0.35	19.79	62.72	0.32	18.38
6	0.40	0.40	21.30	54.31	0.40	22.04
7	0.27	0.33	18.71	63.24	0.30	16.64
8	0.38	0.40	20.67	54.49	0.39	20.92
9	0.37	0.39	19.43	53.04	0.38	19.40
10	0.26	0.32	17.59	60.76	0.29	15.65

Table C9 (cont.)

Mean Type	Vegetation Index					
	D6	DD	PRIN1	PRIN2	PRIN3	PRIN4
Grand	22.04	116.09	0	0	0	0
Cluster Number	Cluster Type 1					
1	22.17	119.41	1.923	-0.017	0.129	-0.049
3	21.64	114.96	1.147	-0.045	0.024	0.018
4	23.22	121.04	-0.881	0.697	-0.042	-0.071
6	21.30	112.70	-2.077	0.125	0.010	-0.003
7	22.38	119.03	-0.010	0.356	0.016	0.312
8	22.30	118.34	-1.421	0.567	0.049	-0.061
9	22.47	117.55	1.729	-0.240	-0.040	-0.016
10	21.33	109.08	-2.865	-0.151	-0.156	-0.004
	Cluster Type 2					
1	21.33	109.08	-2.865	-0.151	-0.156	-0.004
3	21.78	114.68	-0.124	-0.168	-0.017	0.007
5	23.42	122.07	1.871	0.120	-0.062	-0.023
6	22.68	121.10	0.001	0.613	0.031	0.039
7	22.17	119.41	1.923	-0.172	0.129	-0.050
8	22.52	118.93	0.029	0.290	-0.001	-0.003
9	21.39	113.32	-0.410	-0.250	0.019	-0.003
10	20.89	113.50	1.277	-0.641	0.146	-0.003

Table C10. Canonical correlations of vegetation indices versus the set of stems per acre, basal area, and Dbh.

Vegetation Index	Canonical (1) Correlation	F	Prob(F)
R7	.61040	14.8450	1.11×10^{-7}
R6	.61456	15.1724	8.26×10^{-8}
ND7	.62304	15.8614	4.42×10^{-8}
ND6	.62415	15.9547	4.06×10^{-8}
G	.52754	9.6410	.000019
B	MISSING DATA		
GB	.62827	16.3031	2.98×10^{-8}
D7	.59536	13.7271	3.17×10^{-7}
D6	.28106	2.1293	.1036
DD	.20623	1.1105	.3503
PRIN1	.54421	10.5197	7.50×10^{-6}
PRIN1-PRIN4	.22408	1.2688 (2)	.2916
B4-B7	.32770	2.1360 (2)	.0852

(1) This set of canonical correlations are between the appropriate vegetation index, and the linear combination of stems per acre, basal area, and Dbh, for which correlation is maximized.

(2) The F statistics for these correlations are with 3 and 72 degrees of freedom, with all of the remaining test statistics having 3 and 75 d.f..

Note: The hypothesis that the F statistic tests is that the canonical correlation in the first row of the canonical coefficient matrix, and all that follow are equal to zero.

Table C11. Eigenvectors for the set of canonical correlations relating vegetation indices to the set of stems per acre, basal area, and Dbh.

Vegetation Index	Variable	Eigenvectors 1	2	3	Most Influential Variable
R7	Stems	-.00097	.00182	.00554	
	BA	.00439	-.00041	0	Dbh
	Dbh	.04413	-.01500	.21195	
R6	Stems	-.00079	.00382	.05231	
	BA	.00178	-.00026	0	Dbh
	Dbh	.00559	.01516	.21008	
ND7	Stems	-.00101	.00438	.04449	
	BA	.00556	-.01501	.21187	BA
	Dbh	.00176	-.00020	0	
ND6	Stems	-.00084	.00385	.05188	
	BA	.00559	-.01515	.21019	BA
	Dbh	.00173	-.00008	0	
G	Stems	-.00039	.00170	.07089	
	BA	.00360	-.00534	.08648	Dbh
	Dbh	.00468	-.01460	.18210	
GB	Stems	-.00104	.00453	.04320	
	BA	.00179	-.00300	0	Dbh
	Dbh	.00554	-.01496	.21232	
D7	Stems	-.00083	.00377	.05812	
	BA	.00150	.00065	0	Dbh
	Dbh	.00566	-.01526	.20855	
D6	Stems	.00246	-.00950	.15582	
	BA	-.00203	.00735	0	Dbh, Stems
	Dbh	.00498	-.01001	.15030	
DD	Stems	-.00246	.00994	-.14489	
	BA	-.00238	.00731	0	Dbh, Stems
	Dbh	.00483	-.00960	.16087	
PRIN1	Stems	-.00162	.00677	.00922	
	BA	.00100	.00143	-.04121	Dbh
	Dbh	.00560	-.01407	.21234	

Note: See Table C13 for the extraction methods involved in the estimation of the above eigenvectors.

Table C12. Canonical correlations of vegetation indices versus the set of canopy closure, crown diameter, and height.

Vegetation Index	Canonical (1) Correlation	F	Prob(F)
R7	.68305	21.8650	2.82×10^{-10}
R6	.65063	18.3518	5.01×10^{-9}
ND7	.70975	25.3772	1.95×10^{-11}
ND6	.67436	20.8515	6.33×10^{-10}
G	.53621	10.8870	1.16×10^{-5}
B	.67120	20.4878	8.49×10^{-10}
GB	.71495	26.1401	1.12×10^{-11}
D7	.66395	19.7088	1.61×10^{-9}
D6	.13875	0.4842	.6942
DD	.15860	0.6451	.5884
PRIN1	.64597	17.9019	7.35×10^{-9}
PRIN1-PRIN4	.13036	0.4149	.7428
B4-B7	.21080	1.1161	.3483

(1) This set of canonical correlations are between the appropriate vegetation index, and the linear combination of canopy closure, crown diameter, and height, for which correlation is maximized.

(2) The F statistics for these are with 3 and 72 degrees of freedom, with all of the remaining test statistics having 3 and 75 d.f..

Note: The hypothesis that the F statistic tests is that the canonical correlation in the first row of the canonical coefficient matrix, and all that follow are equal to zero.

Table C13. Eigenvectors for the set of canonical correlations relating vegetation indices to the set of canopy closure, crown diameter, and height.

Vegetation Index	Variable	Eigenvector 1	Eigenvector 2	Eigenvector 3	Most Influential Variable
R7	CC	.01098	-.00214	.01134	CD,HT
	CD	-.07577	.00107	.01100	
	HT	.01106	.01051	0	
R6	CC	.00934	-.00211	.01158	HT
	CD	.01098	.01051	0	
	HT	-.07600	.00111	.01076	
ND7	CC	.01098	.00230	.01131	HT
	CD	.01242	.01046	0	
	HT	-.07556	.00123	.01103	
ND6	CC	.00925	-.00229	.00156	CD
	CD	-.07577	.00130	.01077	
	HT	.01252	.01045	0	
G	CC	.02239	-.00325	.00930	HT
	CD	.01811	.01018	0	
	HT	-.07180	.00139	.01278	
B	CC	.00463	-.00169	.01202	HT
	CD	.00761	.01062	0	
	HT	-.07671	.00076	.01025	
GB	CC	.01103	-.00225	.01132	HT
	CD	.01196	.01047	0	
	HT	-.07562	.00118	.01103	
D7	CC	.01825	-.00277	.01012	HT
	CD	.01511	.01034	0	
	HT	-.07364	.00128	.01214	
D6	CC	-.03025	.00520	.01310	HT,CC
	CD	.06177	-.00249	0	
	HT	.03541	.00911	-.00884	
DD	CC	-.05476	.00435	.01342	CD,CC
	CD	.05455	.00504	0	
	HT	.00304	-.00847	.00835	
PRIN1	CC	.00478	-.00156	.01225	All three
	CD	.00664	.01064	0	
	HT	.07692	-.00067	-.00998	
PRIN1-4	CC	.07744	-.02350	-.00895	CC,HT
	CD	.01742	.01027	0	
	HT	.00748	.00338	.01784	

Table C13 (cont.)

Vegetation Index	Variable	Eigenvector 1	2	3	Most Influential Variable
B4-B7	CC	-.06071	.00003	.01681	CC,HT
	CD	.01532	.01037	0	
	HT	.04687	-.00387	.01020	

Note: Let S denote the covariance matrix of the entire set of variables, vegetation indices and the sets of canopy closure, crown diameter, and height, and stems per acre, basal area, and Dbh. Let S_{11} denote the variance of vegetation indices, S_{22} the covariance of stand parameters, and S_{12} the covariance between vegetation indices and stand parameters.

The eigenvectors presented in Tables C11 and C13 are drawn from the matrix determinantal equation:

$\det(S_{21}S_{11}^{-1}S_{12} - l^2S_{22}) = 0$, where: (C1)

det denotes the determinant of the matrix equation,

l^2 are the eigenvalues associated with this expression,

$S_{21} = S_{12}^T$, the transpose of S_{12}, and

S_{11}^{-1} = the inverse of S_{11}.

The eigenvectors of these canonical correlations are associated with the eigenvalues of the above determinantal equation.

Table C14. Mahalanobis distances between clusters, with the significance probability of the distance in the row immediately below the distance value. The distances are given as the distance from a particular cluster, on the left edge, to a corresponding cluster denoted on the top of the table.

B4-B7, with cluster type 1

from #	Distance to cluster number 3	4	6	7	8	9	10
1	0.9216	3.3412	3.6436	2.2490	3.3788	1.2686	4.5222
	.0049	.0000	.0000	.0000	.0037	.0001	.0250
3		2.8500	3.0348	1.8296	2.1928	0.6717	3.7572
		.0431	.0000	.0001	.6500	.2764	.8677
4			1.3011	1.2108	0.8071	2.5792	1.8853
			.0003	.0008	.9477	.0000	.9016
6				1.5360	0.7705	3.0112	1.3811
				.0006	.9890	.0000	.9914
7					1.1381	1.7686	2.5635
					.9662	.0000	.9491
8						2.8257	1.9000
						.0000	.7607
9							3.5807
							.9914

Cluster type 2

from #	Distance to cluster number 3	5	6	7	8	9	10
1	1.8562	2.7953	2.4226	3.3606	2.0863	1.8823	3.2624
	.0000	.0000	.0001	.1499	.0000	.0000	.1701
3		1.3081	1.1462	1.5046	0.6386	0.2841	1.5299
		.4991	.7228	.9905	.2071	.7891	.9901
5			1.5593	1.3394	1.1818	1.5515	1.9207
			.1572	.9691	.0001	.0000	.9158
6				1.6163	0.5778	1.2006	1.9598
				.9350	.1314	.0000	.8910
7					1.4925	1.5291	0.8416
					.0000	.0000	.9512
8						0.7926	1.7834
						.0629	.9873
9							1.4196
							.9942

Table C14 (cont.)

G,B,GB with cluster type 1

Distance to cluster number

from #	3	4	6	7	8	9	10
1	1.3626 .0000	3.4260 .0000	3.6437 .0000	2.3926 .0000	3.4822 .0069	1.4118 .0000	4.5238 .0543
3		3.3503 .0273	3.2034 .0000	1.8392 .0001	3.4486 .6856	0.7737 .2827	3.8597 .9440
4			1.4947 .0009	1.9858 .0000	0.8115 .9853	2.9239 .0000	2.0794 .9517
6				1.7498 .0003	1.1257 .9923	3.0788 .0000	1.3883 .9987
7					2.0117 .9353	1.7795 .0001	2.6565 .9841
8						3.1816 .0000	2.1298 .8348
9							3.6154 .9555

Cluster type 2

Distance to cluster number

from #	3	5	6	7	8	9	10
1	1.8565 .0000	2.8052 .0000	2.4468 .0004	3.4991 .2222	2.0923 .0000	1.8986 .0000	3.2957 .2787
3		1.3242 .6606	1.2052 .8346	1.7769 .9972	0.6513 .3193	0.3585 .7992	1.5913 .9982
5			1.6632 .2104	1.5298 .9882	1.1844 .0004	1.5516 .0000	1.9346 .9718
6				2.0855 .9512	0.7645 .0499	1.3383 .0000	2.1206 .9483
7					1.7016 .0000	1.6930 .0000	0.9829 .0666
8						0.7978 .1190	1.8101 .9977
9							1.4364 .9992

Table C14 (cont.)

R7, R6 with cluster type 1

Distance to cluster number

from #	3	4	6	7	8	9	10
1	0.2877 .5022	2.9008 .0000	3.1435 .0000	1.8182 .0000	2.9227 .0055	0.7492 .0120	4.1241 .0191
3		2.6254 .0299	2.9044 .0000	1.6080 .0001	2.6819 .4956	0.5197 .3077	3.8445 .6829
4			0.8930 .0081	1.4416 .0000	0.7943 .8376	2.3841 .0000	1.2350 .8824
6				1.3623 .0009	0.2238 .9948	2.8708 .0000	1.5493 .9364
7					1.1532 .8713	1.7094 .0000	2.6307 .8278
8						2.6477 .0000	1.6393 .6448
9							3.6120 .7140

Cluster type 2

Distance to cluster number

from #	3	5	6	7	8	9	10
1	1.9445 .0000	2.4614 .0000	2.1477 .0003	2.9245 .1291	1.8779 .0000	1.9801 .0000	3.0888 .1027
3		0.8760 .5824	0.7127 .7712	1.0088 .9752	0.1790 .8304	0.1370 .8837	1.1460 .9682
5			1.5239 .0813	0.7581 .9604	1.0549 .0002	0.9857 .0003	1.0337 .9276
6				1.3343 .8642	0.5669 .0652	0.5768 .0444	1.3203 .8668
7					1.1348 .0000	1.0270 .0000	0.2785 .9811
8						0.1089 .9294	1.2409 .9675
9							1.1323 .9750

Table C14 (cont.)

ND7, ND6 with cluster type 1

Distance to cluster number

from #	3	4	6	7	8	9	10
1	0.3374 .3889	3.1024 .0000	3.3863 .0000	2.0769 .0000	3.1790 .0023	0.7424 .0129	4.1733 .0174
3		2.7656 .0207	3.0629 .0000	1.7719 .0000	2.8541 .4520	0.5042 .3296	3.8369 .6839
4			0.7490 .0315	1.3464 .0000	0.6252 .8959	2.5623 .0000	1.1773 .8925
6				1.3547 .0010	0.2141 .9952	2.9717 .0000	1.5282 .9381
7					1.1653 .8688	1.7996 .0000	2.5229 .8404
8						2.7579 .0000	1.5666 .6697
9							3.5544 .7216

Cluster type 2

Distance to cluster number

from #	3	5	6	7	8	9	10
1	1.8916 .0000	2.3724 .0000	2.1167 .0003	2.8990 .1337	1.8528 .0000	1.9523 .0000	3.0901 .1025
3		0.8318 .6140	0.7320 .7603	1.0669 .9723	0.2432 .7100	0.1520 .8588	1.2123 .9644
5			1.4989 .0880	0.6925 .9668	1.0703 .0002	0.9420 .0005	0.9786 .9348
6				1.4541 .8409	0.5112 .1069	0.5855 .0406	1.4537 .8410
7					1.2282 .0000	1.0843 .0000	0.2939 .9789
8						0.1454 .8778	1.3297 .9628
9							1.1944 .9723

Table C14 (cont.)

D7, D6, DD with cluster type 1

Distance to cluster number

from #	3	4	6	7	8	9	10
1	1.2938 .0000	3.5425 .0000	3.9738 .0000	2.6625 .0000	4.1470 .0002	1.1227 .0004	4.1124 .0499
3		3.0487 .0262	3.2137 .0000	1.9007 .0000	3.3590 .5370	0.4643 .5991	3.8911 .8553
4			1.2269 .0007	1.2492 .0005	1.0305 .8984	2.7832 .0000	2.3004 .8356
6				1.6232 .0003	0.9227 .9815	3.1137 .0000	2.0341 .9375
7					1.5146 .9252	1.6960 .0001	2.8911 .9271
8						3.2026 .0000	2.7808 .4792
9							3.7451 .8688

Cluster type 2

Distance to cluster number

from #	3	5	6	7	8	9	10
1	2.2461 .0000	3.0419 .0000	2.9893 .0000	2.7637 .3030	2.5808 .0000	2.3869 .0000	2.9216 .2554
3		1.2649 .5281	1.2089 .6882	1.4992 .9906	0.6930 .1486	0.3633 .6349	1.0977 .9963
5			1.5058 .1816	0.9449 .9887	1.1706 .0002	1.5379 .0000	1.3553 .9680
6				2.2652 .8418	0.5545 .1580	1.1930 .0000	2.0392 .8790
7					1.7852 .0000	1.8171 .0000	1.2008 .8727
8						0.7909 .0637	1.5969 .9908
9							1.1910 .9966

Table C15. Direct volume regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 17.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	70.043	-0.00300	72.6779
R6	70.567	-0.00037	73.1493
ND7	69.917	0.00010	72.5692
ND6	70.513	0.00164	73.1014
G	84.009	0.01125	87.7007
B	68.465	-0.00198	70.8233
GB	69.043	0.00048	71.7194
D7	77.261	0.00280	80.6160
D6	90.140	0.05731	94.0463
DD	89.825	0.07975	94.7900
PRIN1	69.144	-0.00009	75.1908
PRIN1-PRIN4	67.690	0.08580	75.1098
B4-B7	67.690	0.08580	75.1098

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	-1.835	-0.432	.6669
R6	-12.928	-1.966	.0529
ND7	0.239	0.063	.9502
ND6	-11.993	-1.882	.0636
G	-8.792	-0.862	.3913
B	50.890	7.961	.0001
GB	-3.407	-0.769	.4442
D7	-3.220	-0.559	.5779
D6	36.619	2.443	.0168
DD	39.946	2.445	.0164
PRIN1	18.445	19.715	.0001
PRIN1-PRIN4	18.445	19.920	.0001
B4-B7	40.593	2.842	.0058

Table C15 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	9.091	4.899	.0001
R6	13.581	4.821	.0001
ND7	49.643	4.917	.0001
ND6	78.070	4.829	.0001
G	1.366	2.684	.0089
B	-0.594	-5.131	.0001
GB	58.103	5.046	.0001
D7	1.089	3.816	.0003
D6	-0.825	-1.216	.2278
DD	-0.185	-1.324	.1894
PRIN1	-2.587	-5.301	.0001
PRIN1-PRIN4	-2.587	-5.301	.0001
B4-B7	-2.693	-1.862	.0666
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	1.680	1.370	.1761
B4-B7	-0.118	-0.090	.9283
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	2.720	0.388	.6995
B4-B7	2.910	1.377	.1726
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-20.960	-1.624	.1085
B4-B7	-2.140	-1.405	.1643

Table C16. Crown diameter regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 20.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	2.567	-0.00312	2.6680
R6	2.662	-0.00038	2.7640
ND7	2.515	-0.00224	2.6130
ND6	2.621	-0.00310	2.1789
G	2.724	-0.00139	2.8517
B	2.669	-0.00346	2.7576
GB	2.501	-0.00190	2.6000
D7	2.488	-0.00055	2.6077
D6	3.274	0.00353	3.3966
DD	3.271	0.00800	3.4105
PRIN1	2.741	-0.00348	2.8275
PRIN1-PRIN4	2.422	0.00893	2.8151
B4-B7	2.422	0.00893	2.8151

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	9.110	11.208	.0001
R6	7.435	5.823	.0001
ND7	9.374	12.938	.0001
ND6	7.435	6.051	.0001
G	5.543	3.019	.0034
B	17.948	14.250	.0001
GB	8.738	10.360	.0001
D7	7.731	7.746	.0001
D6	13.211	4.624	.0001
DD	11.693	3.765	.0001
PRIN1	12.763	68.519	.0001
PRIN1-PRIN4	12.763	72.890	.0001
B4-B7	10.834	4.010	.0001

Table C16 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	1.637	4.609	.0001
R6	2.306	4.216	.0001
ND7	9.253	4.827	.0001
ND6	13.666	4.385	.0001
G	0.362	3.952	.0002
B	-0.095	-4.181	.0001
GB	10.703	4.883	.0001
D7	0.253	4.939	.0001
D6	-0.020	-0.157	.8757
DD	0.009	0.345	.7307
PRIN1	-0.394	-3.875	.0001
PRIN1-PRIN4	-0.394	-4.123	.0001
B4-B7	-0.420	-0.157	.8755
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	0.778	3.350	.0013
B4-B7	0.186	0.750	.4572
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	1.494	1.130	.2635
B4-B7	-0.442	-1.110	.2729
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	1.929	0.791	.4317
B4-B7	0.453	1.571	.1205

Table C17. Canopy closure regression from spectral data: remaining summary statistics.
(Note: this is a continuation of Table 21.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	111.765	0.01463	116.4450
R6	111.880	0.02119	116.5420
ND7	111.477	0.01200	116.0910
ND6	111.610	0.01980	116.2420
G	110.802	0.05235	115.7170
B	112.239	-0.00290	116.6180
GB	111.539	0.01189	116.1380
D7	110.865	0.02365	116.6046
D6	111.907	0.01991	115.8880
DD	111.957	0.03366	116.1615
PRIN1	112.342	-0.00284	116.6480
PRIN1-PRIN4	115.070	0.12140	124.9540
B4-B7	115.070	0.12140	124.9540

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	76.650	14.292	.0001
R6	78.368	9.466	.0001
ND7	76.380	15.925	.0001
ND6	79.189	9.876	.0001
G	85.647	7.312	.0001
B	69.064	8.439	.0001
GB	77.404	13.475	.0001
D7	80.163	11.612	.0001
D6	83.889	5.023	.0001
DD	84.381	4.644	.0001
PRIN1	72.806	61.053	.0001
PRIN1-PRIN4	72.806	60.325	.0001
B4-B7	85.864	4.611	.0001

Table C17 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	-1.723	-0.735	.4645
R6	-2.408	-0.679	.4992
ND7	-10.987	-0.861	.3919
ND6	-16.370	-0.805	.4233
G	-0.643	-1.102	.2739
B	0.695	0.462	.6453
GB	-12.227	-0.835	.4061
D7	-0.370	-1.082	.2828
D6	-0.503	-0.665	.5078
DD	-0.100	-0.639	.5250
PRIN1	0.246	0.373	.7101
PRIN1-PRIN4	0.246	0.373	.7101
B4-B7	-0.511	-0.277	.7822
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	-1.610	-1.010	.3175
B4-B7	0.695	0.406	.6861
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	3.074	0.340	.7376
B4-B7	0.413	0.150	.8813
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-3.965	-0.220	.8268
B4-B7	-0.826	-0.416	.6789

Table C18. Height regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 22.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	44.308	-0.02111	45.9480
R6	47.283	-0.02471	49.1322
ND7	41.802	-0.01528	43.3538
ND6	45.171	-0.02230	46.9239
G	60.238	-0.03680	63.1994
B	44.490	-0.00679	46.6050
GB	41.179	-0.01510	42.7630
D7	48.405	-0.02310	50.1577
D6	77.986	0.04823	81.0482
DD	78.248	0.05544	81.3487
PRIN1	46.805	-0.00575	49.0980
PRIN1-PRIN4	40.356	0.00686	45.5570
B4-B7	40.356	0.00686	45.5570

Regression coefficient estimates

Vegetation Index	b_0	t statistics	Prob(t)
R7	52.919	15.670	.0001
R6	40.228	7.474	.0001
ND7	54.782	18.547	.0001
ND6	40.198	7.880	.0001
G	36.482	4.224	.0001
B	117.788	22.860	.0001
GB	50.359	14.720	.0001
D7	47.148	10.335	.0001
D6	91.655	6.524	.0001
DD	90.583	5.964	.0001
PRIN1	78.448	101.920	.0001
PRIN1-PRIN4	78.448	109.761	.0001
B4-B7	87.647	7.947	.0001

Table C18 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	11.445	7.753	.0001
R6	16.545	7.175	.0001
ND7	64.598	7.266	.0001
ND6	98.110	7.583	.0001
G	2.104	4.884	.0001
B	-0.721	-7.177	.0001
GB	74.688	8.398	.0001
D7	1.573	6.965	.0001
D6	-0.599	-0.950	.3452
DD	-0.104	-0.801	.4528
PRIN1	-3.052	-7.266	.0001
PRIN1-PRIN4	-3.052	-7.826	.0001
B4-B7	-1.366	-1.253	.2143
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	3.630	3.810	.0003
B4-B7	0.059	0.059	.9531
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	4.060	0.750	.4557
B4-B7	0.195	0.119	.9054
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-4.773	-0.479	.6333
B4-B7	0.283	0.240	.8110

Table C19. Average Dbh regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 27.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	0.965	-0.00190	1.0035
R6	0.944	-0.00230	0.9837
ND7	0.937	-0.00170	0.9572
ND6	0.924	-0.00190	0.9614
G	1.004	-0.00290	1.0470
B	1.089	-0.00020	1.1379
GB	0.938	-0.00120	0.9748
D7	0.934	-0.00240	0.9684
D6	1.339	0.00770	1.3985
DD	1.365	0.00780	1.4202
PRIN1	1.121	0.00035	1.1727
PRIN1-PRIN4	0.877	0.00610	0.9646
B4-B7	0.877	0.00610	0.9646

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	6.334	12.708	.0001
R6	4.649	6.112	.0001
ND7	6.546	14.790	.0001
ND6	4.685	6.423	.0001
G	3.189	2.860	.0054
B	12.713	15.679	.0001
GB	6.089	11.790	.0001
D7	5.348	8.439	.0001
D6	6.468	3.450	.0007
DD	7.489	3.738	.0004
PRIN1	9.117	76.539	.0001
PRIN1-PRIN4	9.117	86.494	.0001
B4-B7	7.741	4.593	.0001

Table C19 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	1.248	5.726	.0001
R6	1.934	5.937	.0001
ND7	7.018	5.993	.0001
ND6	11.368	6.144	.0001
G	0.297	5.344	.0001
B	-0.066	-4.508	.0001
GB	8.050	5.997	.0001
D7	0.189	6.036	.0001
D6	0.120	1.453	.1502
DD	0.014	0.809	.4211
PRIN1	-0.272	-4.189	.0001
PRIN1-PRIN4	-0.272	-4.734	.0001
B4-B7	-0.193	-1.201	.2334
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	0.577	3.990	.0002
B4-B7	-0.412	-2.573	.0074
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	-1.667	-2.089	.0402
B4-B7	0.552	2.294	.0247
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-2.592	-2.010	.0480
B4-B7	-0.237	-1.366	.1247

Table C20. Basal area regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 28.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	531.900	0.06600	551.2790
R6	536.670	0.01130	555.6480
ND7	534.435	0.01580	554.5240
ND6	539.107	0.02240	558.6580
G	598.628	0.06210	623.6710
B	523.662	-0.00840	542.0950
GB	529.336	0.01420	549.2840
D7	568.076	0.03220	591.7580
D6	616.626	0.11340	642.4830
DD	618.497	0.18180	652.8920
PRIN1	526.538	-0.00790	544.6310
PRIN1-PRIN4	528.778	0.22780	585.8730
B4-B7	528.778	0.22780	585.8730

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	16.544	1.414	.1614
R6	-6.059	-0.334	.7392
ND7	21.542	2.040	.0448
ND6	-3.187	-0.181	.8570
G	5.478	0.201	.8410
B	128.399	7.263	.0001
GB	13.564	1.113	.2692
D7	15.009	0.960	.3398
D6	108.920	2.778	.0069
DD	109.195	2.557	.0125
PRIN1	59.418	23.015	.0001
PRIN1-PRIN4	59.418	22.966	.0001
B4-B7	109.869	2.752	.0074

Table C20 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	19.221	3.758	.0003
R6	28.344	3.649	.0005
ND7	103.837	3.700	.0004
ND6	160.576	3.592	.0006
G	2.704	1.991	.0500
B	-1.263	-3.944	.0002
GB	121.684	3.816	.0003
D7	2.232	2.884	.0051
D6	-2.246	-1.266	.2094
DD	-0.428	-1.168	.2463
PRIN1	-5.466	-3.870	.0002
PRIN1-PRIN4	-5.466	-3.870	.0002
B4-B7	-5.819	-1.474	.1446
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	3.330	0.970	.3348
B4-B7	1.722	0.469	.6405
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	15.960	0.814	.4180
B4-B7	4.373	0.741	.4613
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-37.270	-0.929	.3047
B4-B7	-3.530	-0.829	.4090

Table C21. Stems per acre regression from spectral data: remaining summary statistics. (Note: this is a continuation of Table 29.)

Vegetation Index	Mean Square Error	PRESS mean	PRESS variance
R7	4483.300	0.08900	4668.8800
R6	4486.100	0.12090	4667.2800
ND7	4486.950	0.10280	4667.9200
ND6	4487.940	0.13480	4683.3900
G	4446.850	0.19580	4601.2700
B	4455.830	0.01920	4646.3900
GB	4485.420	0.08100	4667.3800
D7	4476.720	0.12740	4649.6100
D6	4319.537	0.03210	4432.6900
DD	4383.560	0.15770	4539.7800
PRIN1	4448.100	0.00780	4635.6000
PRIN1-PRIN4	4449.030	0.29040	4782.7200
B4-B7	4449.030	0.29040	4782.7200

Regression coefficient estimates

Vegetation Index	b_0	t statistic	Prob(t)
R7	124.372	3.662	.0005
R6	124.322	2.371	.0202
ND7	129.701	4.238	.0001
ND6	130.907	2.575	.0120
G	196.415	2.647	.0098
B	172.113	3.338	.0013
GB	126.414	3.540	.0007
D7	153.145	3.491	.0008
D6	313.408	3.021	.0034
DD	287.008	2.524	.0137
PRIN1	133.978	17.855	.0001
PRIN1-PRIN4	133.978	17.853	.0001
B4-B7	292.054	2.538	.0140

Table C21 (cont.)

Regression coefficients (cont.)			
Vegetation Index	b_1	t statistic	Prob(t)
R7	4.308	0.290	.7726
R6	4.179	0.186	.8529
ND7	11.675	0.144	.8857
ND6	7.876	0.061	.9515
G	-3.131	-0.846	.4003
B	-0.698	-0.748	.4570
GB	20.114	0.217	.8290
D7	-0.963	-0.444	.6586
D6	-8.143	-1.734	.0870
DD	-1.315	-1.349	.1813
PRIN1	-3.410	-0.833	.4076
PRIN1-PRIN4	-3.410	-0.833	.4076
B4-B7	-7.997	-0.698	.4900
Vegetation Index	b_2	t statistic	Prob(t)
PRIN1-PRIN4	-9.780	-0.980	.3281
B4-B7	15.240	1.430	.1560
Vegetation Index	b_3	t statistic	Prob(t)
PRIN1-PRIN4	80.600	1.417	.1607
B4-B7	-4.134	-0.241	.4871
Vegetation Index	b_4	t statistic	Prob(t)
PRIN1-PRIN4	-8.790	-0.084	.9332
B4-B7	-2.140	-1.405	.1643

Table C22. Landsat scene identification.

Quadrangle	Scene Identification Number	Date of Imagery
New Willard NW	30920-16015	9/10/80
New Willard NE	30920-16015	9/10/80
New Willard SE	30920-16015,	9/10/80,
	21707-16084	9/25/79
New Willard SW	30920-16015,	9/10/80,
	21707-16084	9/25/79
Blanchard	21707-16084	9/25/79
Onalaska	21707-16084	9/25/79
Livingston NW	21707-16084	9/25/79
Livingston NE	21707-16084	9/25/79

APPENDIX D.

GLOSSARY OF TERMS USED IN THE TEXT OF THIS STUDY.

This section contains brief definitions for those terms used in this study which may cause the reader confusion. Complete definitions can be found in the Manual of Remote Sensing (American Society of Photogrammetry 1983).

Aerial Photo Volume Table: A table of either individual of stand volumes, obtained through measurements from aerial photographs, with a minimum of ground measurements. The typical photo measurements used are canopy closure, crown diameter, and total tree height.

Ancillary Data: Data used in remote sensing analyses, obtained from auxillary sources, usually not from remote sensing devices.

Asymptotic Porperties: The mathematical properties of a probability density function, occuring at large sample sizes approaching infinity.

Basal Area: The cross sectional area of a tree, measured at diameter breast height, 4.5 feet above the ground.

Bias: A systematic departure from expected statistical behavior in a measured variable.

Canonical Correlations: A multivariate statistical technique that finds linear functions within two sets of data such that the correlation between these functions is maximized.

Canopy Closure: The proportion of a unit area covered by tree crowns which can also be expressed as a percentage.

Classification: A statistical rule developed to assign a data point from a set of observations into that category, for which it is most likely to be a member.

Cluster Analysis: A multivariate statistical technique which forms groupings within a data set based on measures of "similarity", or divides a data set into groups using a measure of "dissimilarity".

Confidence Interval: A range of data values within which a given data observation will fall with a specified probability. (This term is synonymous with confidence levels in the context of this study.)

Confusion Matrix: A matrix presenting, by number of observations, the results of a classification analysis. It includes the number of correctly classified observations on the main diagonal and incorrect classification elsewhere.

Covariance Matrix: The multivariate analog of variances, which are located on the main diagonal, and having covariances elsewhere.

Crown Ratio: The proportion of an individual tree's bole actually supporting foliage.

Digital Terrain Map: Topography digitized on a U.S. Geological Survey 1:250,000 quadrangle in a grid cell format which can be used with remote sensing data.

Geographic Information Systems: Data bases incorporating spatial distribution of information as a component within it.

Heteroscedasticity: The statistical property where, for instance, the residuals in a regression analysis do not have constant variance across the range of the data.

Land Use/Land Cover: A term used by remote sensing specialists to account for the fact that when interpreting images, only land cover is visible, but can be used to draw inferences concerning land use.

Likelihood Ratio Principle: The development of test statistics using the ratio of two probability densities as its basis.

Maximum Likelihood: The development of statistical parameter estimators using the properties of the density functions for which they are to be used.

Mean Vector: The mean of a set of multivariate observations.

Median: The point in a frequency distribution where fifty percent of the observations lie above and below it.

Mosaic: The assembling of aerial photographs, over an area, into a single, composite image by matching of visible common ground points, and effective areas of photographs.

Multicollinearity: The phenomenon, occuring in regression analysis, where independent variables are linear functions of one another.

Multispectral Scanner: A remote sensing device which is capable of simultaneously recording data in several wavelength intervals.

Multitemporal Data: Remotely sensed data, obtained at two or more points in time.

Nearest Neighbor: A classification technique assigning an observation into a descriptive category by comparison with those observations nearest to it, either spatially or spectrally in a pixel by pixel format.

Precision: The degree of spread in an estimate of a random variable.

Principal Components: A multivariate statistical technique which rotates the mathematical axes of a data set. The object is to extract linear, orthogonal combinations of the original data that are uncorrelated and preserve as much of the original sample variation in as few combinations as possible.

Registration: Digital techniques for matching separate images, either spatially, spectrally, and/or geometrically.

Remote Sensing: Evaluation of an object from a distance, with a minimum of or no contact with it.

Residual: The difference between a data observation and its predicted value in a regression analysis.

Silviculture: The art and science of manipulating forest stands. It includes the techniques to harvest, thin, and regenerate forest stands.

Spectral Signature: The response of an object on the ground to solar radiation, viewed by a remote sensing device as reflectance.

Stocking: Measures of the degree to which a forested, or potentially forested, area is occupied by trees.

Training Sets: Areas on a remotely sensed image which have known ground characteristics, and are of known location on the image. They are used as comparative indices around which classification rules are developed.

Vegetation Index: A transformation in digital, remotely sensed, data designed to enhance the relative contrast between vegetation and surrounding areas on an image.