5-1988

A New Approach to Forest Site Quality Modeling

David L. Verbyla
Utah State University

A NEW APPROACH TO FOREST SITE QUALITY MODELING

by

David L. Verbyla

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Forest Resources

Approved:

Major Professor	Committee Member
Committee Member	Committee Member
Committee Member	Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

1988

ACKNOWLEDGEMENTS

I want to thank the members of my committee, Drs. Richard Fisher, John Hanks, Dave Roberts, H. Charles Romesburg, and Neil West for their advise and their hours of review of this dissertation. This study was partially funded by the USDA Forest Service.

I thank Richard Fisher, Paul Mohai, H. Charles Romesburg for providing me with the oppurtunity to teach various courses.

Various people hired me for computer work during my stay at Utah State University. I thank Dennis Austin, Fred Baker, Steve Daniels, Bill Gartner, George Hart, Jim Long, Paul Mohai, H. Charles Romesburg for this support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Multiple regression and discriminant analysis procedures are commonly used to develop forest site quality models. When they contain many independent variables relative to sample size, these models may be subject to predicton bias. Fit statistics such as R^2 in regression and classification tables in discriminant analysis show the apparent model accuracy but this may be a biased estimate of the model's actual accuracy. Sample splitting methods such as cross-validation and the bootstrap can be used to get an unbiased actual accuracy estimate.

A discriminant procedure called classification tree analysis uses cross-validation to build the classifier with the greatest estimated actual accuracy. Because cross-validation is used in model development, the model is less likely to be over-fit with insignificant variables when compared with stepwise linear discriminant analysis.

Classification tree analysis and linear discriminant analysis were used to develop models that discriminate prime vs. nonprime ponderosa pine (*Pinus ponderosa*) sites. Prime sites are defined as having site index_{25} greater than 7.6 meters; nonprime sites have site index_{25} less than 7.6 meters. Forest habitat type, percent sand content, and soil pH were incorporated in both models. The cross-valiation estimate of classification tree actual accuracy was 88 percent. A random bootstrap estimate of the linear discriminant function actual accuracy was 80 percent.

A multiple regression model developed with random plots revealed little useful information and was biased when applied to prime site plots. The conventional regression approach using random plots may be misleading if one is interested in identifying relatively rare prime sites.

Forest habitat types within the ponderosa pine series in southern Utah were examined as site quality indicators. The site index range within any one habitat type was broad. However, the best ponderosa pine sites consistently occurred in only *Pinus ponderosa*/*Quercus gambelii*, and *Pinus ponderosa*/*Symphoricarpos oreophilus* habitat types; or in habitat types within the *Pseudotsuga menziesii* or *Abies concolor* series. Therefore forest habitat type when used with other site variables may be useful in predicting prime sites.

The effect of aspect at the upper elevational limit of ponderosa pine was examined by comparing mean site index and mean initial 10 year diameter increment on southerly and northerly slopes from two cinder cones. Southerly aspects on both cinder cones had greater mean diameter increment. Southerly aspects on the highest elevation cinder cone had the greatest mean site index. There was no significant difference in mean site index on the lower elevation cinder cone. Optimal aspect for height and diameter growth may differ due to 1)the effect of density on diameter increment; and/or 2)available soil water limiting height growth during the spring and ambient temperature/solar radiation limiting diameter growth in late summer. Optimal aspect for forest production is not constant but varies with tree species, elevation, latitude, and other factors affecting site microclimate.

CHAPTER I

GENERAL INTRODUCTION

One objective in site quality research is to predict tree growth potential based on site properties. The typical approach is to measure soil, topographic and vegetation variables from randomly selected plots. Multiple regression procedures are then employed to develop a model based on some subset of the potential predictor variables measured. Hundreds of studies have been published using this approach (see reviews by Carmean (1975), Hagglund (1981), Grey (1983)). This dissertation demonstrates weaknesses in this conventional approach and offers a more rational and simple approach for developing site quality models. Chapter 2 demonstrates that models developed by some statistical procedures may have prediction bias. Models may appear to predict well with the sample cases used for model development. However, models with prediction bias will predict with less than expected accuracy when applied to new sample cases. Cross-validation and the bootstrap are recommended to researchers as methods to estimate prediction bias.

Stepwise regression and discriminant analysis procedures may have prediction bias because they tend to include too many predictor variables in the model (Flack and Chang (1987), Freedman (1983), Lovell (1983)). A new method , classification and regression trees (Breiman et al. 1984), uses cross-validation during model development to minimize model overfitting with too many predictor variables. This method is discussed in chapter 3.

A classification tree discriminating prime vs. nonprime ponderosa pine sites in soutern Utah is presented in chapter 4. Since silvicultural decisions often involve only the best sites, models that discriminate prime sites may be more useful than models developed to predict site index over a random sample of sites. Relationships that occur over a random sample of sites may not be important on prime sites. A multiple regression model developed with randomly selected plots is compared with models developed to discriminate prime sites.

Chapter 5 examines the potential of habitat types as site quality indicators. Past studies have shown mean site index to be significantly different among habitat types from different series. However, few studies have examined within-series habitat types as site quality indicators.

Most site quality studies include aspect in regression models as cosine transformation such that the northeast aspect is optimal. It is generally believed that northerly aspects in the northern hemisphere have greatest forest production because of greater available soil moisture and more favorable temperatures. However, the optimal aspect may vary with season and therefore may differ for spring height growth versus summer diameter growth. Chapter 6 examines whether mean diameter and height growth are the greatest on the same aspect.

The following chapters were written for submission as journal articles and therefore can be read independently. Because of this, some material is redundant among chapters.

CHAPTER II

POTENTIAL PREDICTION BIAS IN REGRESSION AND DISCRIMINANT ANALYSIS

INTRODUCTION

Many site quality studies have used multivariate statistical procedures to develop models that appear accurate. Goodness-of-fit statistics such as coefficient of determination (R^2) in regression or classification tables in discriminant function analysis are typically reported as an assessment of model accuracy. The purpose of this chapter is to show that this approach may be misleading due to prediction bias. I will demonstrate why prediction bias can occur in common statistical models and how prediction bias can be estimated.

MULTIPLE REGRESSION

A multiple regression model can always be perfectly fitted through N data points by using N-1 independent variables in the model. Therefore a multiple regression model containing many independent variables relative to the sample size will always have a good fit through the data. However, such a model may be subject to <u>prediction bias</u>; that is the apparent predictive ability based on data used to fit the model may be much greater than the model's actual predictive ability with independent data (Neter and Wasserman 1974). This can lead to erroneous conclusions about the biological significance of the independent variables in the model.

For example, Corns and Pluth (1984) reported an increase in R^2 value from .58 to .91 with the addition of vegetation variables to a white spruce (Picea glauca) site index regression model. Based on this increase in R^2 value the authors concluded: "Results of this study indicate that vegetational attributes used in addition to soil and site properties as independent variables in tree growth predictions can account for significant amounts of the variability in western Alberta lodgepole pine and white spruce MAI (mean annual increment) and SI (site index)." This may be true. However, the white spruce nine-independent variable regression model was based on only 30 stands. Because of the large number of independent variables relative to the sample size , such a model may be subject to positive prediction bias.

To illustrate this prediction bias potential, stepwise regression was run on a data set of thirty cases to develop a nine-independent variable model. All variables were uniformly independently randomly distributed within two times the standard deviation for each variable reported by Corns and Pluth (1984) (Table 2.1). This resulted in a highly significant equation ($P<.01$) with an R^2 value of .74 . The R^2 value for the equation increased from .55 to .74 with the inclusion of the random "vegetational" variables into the model. This could lead to the erroneous conclusion that vegetational variables are important as predictor variables, when actually they were only random "noise" being added to an over-fitted model.

TABLE 2.1. Comparison of site index regression models developed with measured[a] and random[b] data sets

Independent variable	Coefficient Real data	Random
Elevation (m)	-0.004**	-0.007*
Slope angle(%)		
Slope aspect	0.92	
Log thickness organic horizon		
Hue B horizon	0.58	0.90**
Value B horizon		
Croma B horizon		-0.45
Drainage class	2.32*	-0.65
Log hydraulic conductivity (cm/day)		
Stone volume (%)		
1/log litter cover	9.67*	-4.59**
Canopy cover (%)		0.07
Deadfall cover (%)	-1.11**	-0.21
<u>Ledum groenlandicum</u> cover (%)	-1.64**	-0.75**
<u>Rosa acicularis</u> cover (%)	-1.97*	
<u>Calamagrostis canadensis</u> cover (%)		0.39
<u>Cornus canadensis</u> cover (%)	0.77**	
Regeneration density (stems/ha)		
R^2	0.91	0.74
F -value	22.75**	6.38**

[a]Table 4, p. 20 (Corns and Pluth 1984).
[b]Random data generated with VAX-11 FORTRAN-77 V3.0 RAN function; data generated within mean $\pm$ 2 times standard deviation of each variable as reported by Corns and Pluth (1984).
*P < 0.05 **P < 0.01

R^2 can be computed as follows:

$$\text{Unadjusted } R^2 = 1 - SSE/SST \quad [2.1]$$

Where SSE is the sum of squares error and SST is the sum of squares total. For a given response set, SST remains constant and SSE can never decrease. Therefore, the R^2 value often increases when independent variables are added to the model. The adjusted R^2 corrects for this bias (Zar 1974) and should be reported instead of R^2. The adjusted R^2 can be computed as follows:

$$\text{Adjusted R2} = 1 - [\ (n-1)/(n-p-1)]SSE/SST \quad [2.2]$$

Where n is the number of sample cases and p is the number of predictor variables in the model.

Forward inclusion stepwise procedures are often used to avoid the prediction bias potential associated with many model predictor variables relative to sample size. In stepwise regression, independent variables are entered into the model, one by one, on the basis of some statistical criteria (usually the largest F-value). The objective is to isolate a subset of available predictor variables that will yield the "best" model with relatively few independent variables.

However, with stepwise regression there may exist many possible combinations of independent variables that could conceivably make up the final model. Because of this, the usual F-statistics and R^2 value generated from the stepwise regression procedure are biased (Rencher and Pun 1978, Berk 1978, Diehr and Hoflin 1974, Pope and Webster 1972, Draper et al. 1971). This can also lead to prediction bias if many predictor variables are tested by the stepwise procedure. For example, Page (1976) used stepwise multiple regression on 103 independent variables to predict site index of balsam fir (Abies balsamea) and black spruce (Picea mariana in Newfoundland. The stepwise procedure selected the "best" subset of eight independent variables for each regression equation. Because there were so many different combinations of predictor variables there is a high chance that one of the combinations will fit the sample data well but predict poorly when tested on new data. Table 2.2 compares results from Page's stepwise regression models with stepwise regression of 103 uniform random variables. All predictor and response variables used in the stepwise regressions were independent random uniform integers varying from 1 to 100. The R^2 values from the stepwise regression of random numbers were higher than the actual site index equations for half of the models. Because of the high potential of prediction bias in these models, predictor variables selected by the stepwise procedure can be biologically insignificant.

TABLE 2.2. Comparison of R^2 values from regression models developed with measured[a] and random[b] data sets

Regression model	R^2 values Real data	Random data
Avalon Peninsula:		
Fir---well drained sites (48 plots)	0.70	0.74
Fir---poorly drained sites (27 plots)	0.94	0.88
Spruce---well drained sites (25 plots)	0.91	0.93
Spruce---poorly drained sites (50 plots)	0.71	0.61
Western Newfoundland:		
Fir---well drained sites (59 plots)	0.60	0.59
Fir---poorly drained sites (16 plots)	0.94	0.99
Spruce---well drained sites (34 plots)	0.86	0.87
Spruce---poorly drained sites (41 plots)	0.77	0.73

[a]Table 2 p. 136 (Page 1976).
[b]Random data generated with VAX-11 FORTRAN-77 V3.0 RAN function; data for all variables randomly uniformly distributed between 1 and 100.

DISCRIMINANT ANALYSIS

One objective of discriminant analysis is to weight and combine independent variables in a linear function which predicts class membership. Although the mathematics of discriminant analysis differs from regression, prediction bias can still result due to a large number of independent variables relative to sample size and/or many combinations of independent variables possible with stepwise discriminant analysis. For example, Tom and Miller (1979,1980) reported a 97 percent classification accuracy in predicting site index class by using discriminant analysis on LANDSAT-1 data and ancillary map variables. However, the model contained 19 independent variables and was based on a sample size of only 37 plots. To illustrate the potential prediction bias of such a model, discriminant analysis was run on 19 independent random uniform variables from a sample of 37 cases. This resulted in a classification accuracy of 94.6 percent (Table 2.3). Thus it seems likely that even though the discriminant functions predicted sample site index class well, the predictive power of the functions when applied to new data would be lower due to prediction bias.

TABLE 2.3. Discriminant analysis classification results of random[a] data

		Predicted group membership								
		Group number								
	Actual membership	1	2	3	4	5	6	7	8	9
Group 1	1	1	0	0	0	0	0	0	0	0
Group 2	9	0	7	1	0	0	0	1	0	0
Group 3	3	0	0	3	0	0	0	0	0	0
Group 4	8	0	0	0	8	0	0	0	0	0
Group 5	3	0	0	0	0	3	0	0	0	0
Group 6	3	0	0	0	0	0	3	0	0	0
Group 7	6	0	0	0	0	0	0	6	0	0
Group 8	1	0	0	0	0	0	0	0	1	0
Group 9	3	0	0	0	0	0	0	0	0	3

Overall classification accuracy = 94.6 percent
(35 out of 37 plots correctly classified)

[a]Random numbers generated with VAX-11 FORTRAN-77 V3.0 RAN function; numbers were random uniform integers between 1 and 100. Group membership was randomly assigned independent of predictor variables.

RESAMPLING TO ASSESS PREDICTION BIAS

One obvious way to assess prediction bias is to randomly divide the sample cases into two groups. The first group is then used for model development and the second group is used for model validation. This approach is called 2-fold cross-validation. It has several weaknesses. Since only half the sample cases are used in model development, model degrees of freedom are reduced by half. This will cause a decrease in model statistical significance. Also the estimates of model coefficients will not be as precise and therefore may be unreliable. The estimate of prediction bias may also not be very precise.

These weaknesses can be minimized by dividing the sample cases into many groups. Therefore N-fold cross-validation (where N is the number of sample cases) is often used in model validation. Prediction bias can be assessed with N-fold cross-validation as follows:

1) Delete the _i_th sample case (_i_ initially is 1).
2) Develop the predictive model with the remaining sample cases.
3) Run the model on the excluded case. Model accuracy is estimated as this predicted value minus the actual value of the excluded sample case.
4) Return the excluded sample case and increment _i_ by one.

Continue steps 1) through 4) until all sample cases have been excluded once from model development. The cross-validation estimate of model accuracy is the mean of the accuracy estimates from step 3) . Prediction bias can then be estimated as the original apparent accuracy of the model minus the cross-validation estimate of model accuracy.

Cross-validation has been used by forestry researchers (Frank et al. 1984, Harding et al. 1985). Recently a better resampling procedure called the bootstrap (Efron 1983) has been developed. The bootstrap method is better because it gives a less variable prediction bias estimate than cross-validation (Efron 1982,1983). The bootstrap also seems to be better than cross-validation when a complicated prediction model is used (Gong 1986).

Since the bootstrap (to my knowledge) has not been used by forestry researchers to estimate prediction bias, I will illustrate the procedure. Suppose a forester must decide which of the stands he is managing should receive fertilization. He randomly samples some stands and develops a discriminant function that predicts from two predictor variables fertilizer response or nonresponse. The discriminant function accurately predicts the responsive sample stands, but how accurate will it be in classifying the remaining population of stands?

I simulated this problem with a computer by generating a population of 3000 cases. Half of the population was generated as fertilizer responsive stands and the other half as non-responsive. Five normally distributed predictor variables with means of 10,20,30,40 and 50 and standard deviations of 10 were randomly assigned to each population case. Therefore there was no difference in terms of the predictor variables between the response and nonresponse population groups. A random sample of 30 was then drawn from the population and linear discriminant analysis was conducted. The function correctly classified 20 of the 30 sample cases. Prediction bias of the function was estimated by the following double bootstrap procedure.

1) Select a bootstrap sample (X*) of 10 cases randomly and <u>with replacement</u> from the original sample.
2) Construct a linear discriminant function with the bootstrap sample.
3) Estimate prediction bias (P_{bias1}) as the proportion of bootstrapped cases correctly classified minus the proportion of original sample cases correctly classified by the bootstrap discriminant function.
4) P_{bias1} is slightly biased. This bias can be adjusted for by bootstrapping the first bootstrap sample (Efron 1983). Select the second bootstrap sample (X**) of 10 cases randomly and with replacement from the first bootstrapped sample (X*).
5) Construct a linear discriminant function with X**.
6) Estimate P_{bias2} as the proportion of cases from X** correctly classified minus the proportion of cases from X* correctly classified by the X** discriminant

function.

Steps 1) through 6) are repeated a large (NBOOT=200) number of times. The bootstrapped estimate of prediction bias is as follows:

$$(2*Pbias_1 - Pbias_2)/NBOOT \quad [2.3]$$

The bootstrapped estimate of prediction bias was .18 . The sample apparent accuracy of the function was 66 percent, therefore the bootstrap estimate of actual accuracy is 66 - 18 = 48 percent. Since there was no difference between the population's two response groups in terms of the predictor variables, the function's true accuracy was 50 percent. In most research applications, a model's true accuracy will be unknown (since the population is not measured) and therefore must be estimated.

The bootstrap estimate of regression prediction bias is similar to that of discriminant analysis except that difference between the actual and predicted dependent variable value is bootstrapped rather than discrimination of group membership.

CONCLUSIONS

1) Users of predictive models should realize that there is a cost of prediction bias potential associated with including or examining many predictor variables in model building.
2) Model prediction bias can be estimated without measuring new data. The bootstrap is currently the

best procedure for estimating model prediction bias.

CHAPTER III

CLASSIFICATION TREES: A NEW DISCRIMINATION TOOL

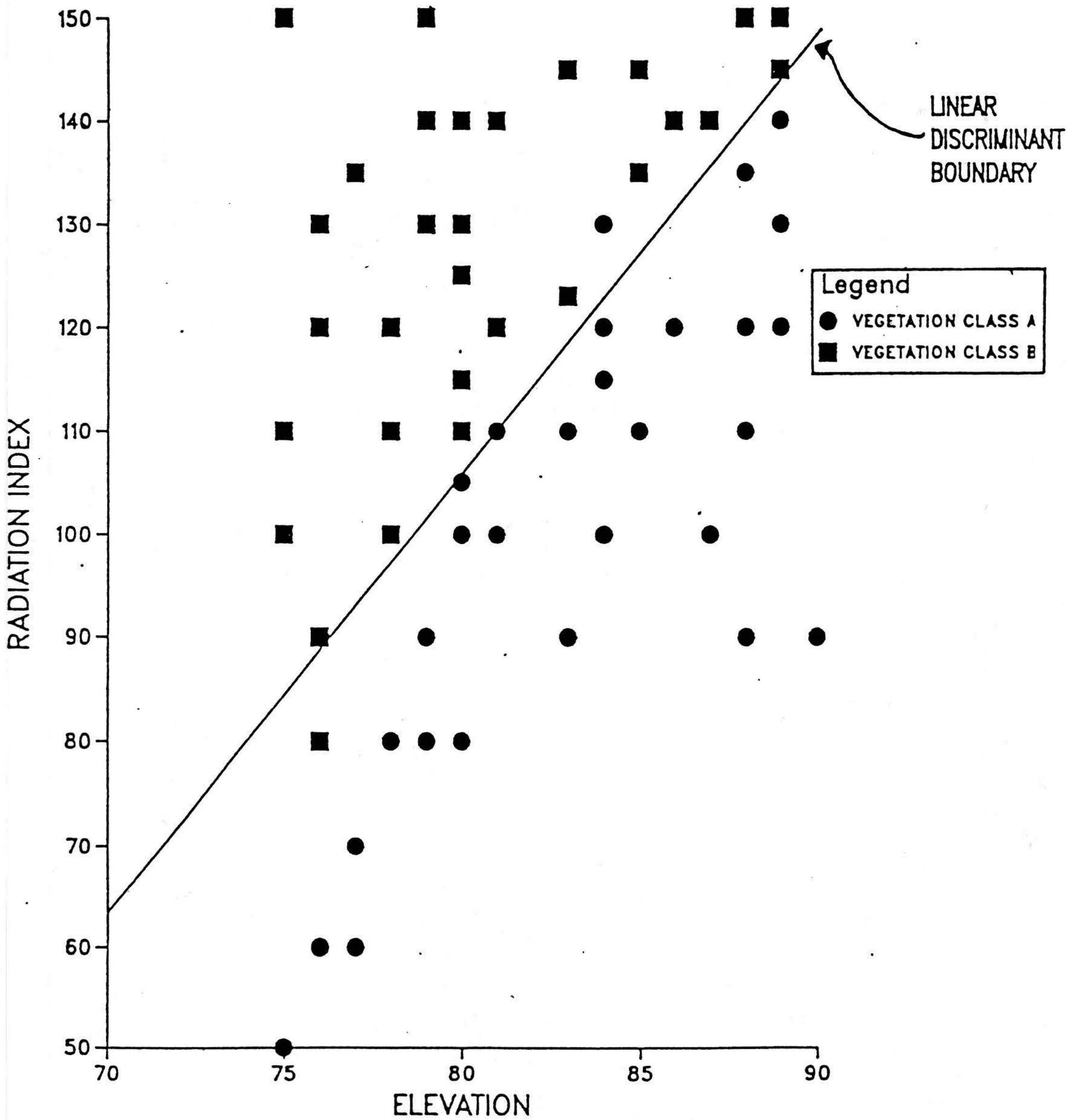
INTRODUCTION

Prediction of class membership is a common objective in site quality research. Linear discriminant analysis has been the predominant method used in past site class studies (Gasana and Lowenstein 1984, Harding et al. 1985, Turvey et al. 1986). A class discrimination method, called classification trees, has been recently developed (Breiman et al. 1984). The purpose of this chapter is to introduce this discrimination tool and to show that classification trees may perform better that linear discriminant analysis under certain conditions.

LINEAR DISCRIMINANT ANALYSIS

One objective of discriminant analysis is to weight and linearly combine selected predictor variables so that sample classes are separated by a linear boundary that maximizes the ratio of between-class to within-class variances. A hypothetical example is presented in Figure 3.1. All points to the right of the boundary are predicted to belong to vegetation class A, and all points to the left of the boundary are predicted to belong to vegetation class B. Two-class discriminant analysis can be viewed as analogous to linear regression with a (0,1) dummy dependent variable (Huberty 1972).

Figure 3.1. Linear discriminant function developed with 60 hypothetical sample cases.



Linear discriminant analysis has several limitations. Distorted effects can result if the assumption of equal covariance structure is not met (Huberty 1975, Williams 1983). Since the method maximizes the between-class to within-class variance ratio, outlier sample cases can drastically affect the results. These outliers are difficult to detect in the multivariate sample space (Harner and Whitmore 1980). Because of this sensitivity to outliers, discriminant function coefficients may be unstable when sample sizes are small (Morrison 1984).

Because of these limitations, statistically significant predictor variables in discriminant analysis may be meaningless (Williams 1983, Cavallaro et al 1980). To demonstrate this, random predictor variables were added to the sample cases displayed in Figure 3.1. Then linear discriminant analysis was run using the SPSSX statistical package (SPSS Inc. 1983). Despite a very strong bivariate relationship, random variables were included in the model by the discriminant analysis (Table 3.1). Because of this sensitivity to random noise, discriminant analysis results can be misleading if the sample size is small relative to the number of predictor variables tested.

Table 3.1. Discriminant analysis[a] based on cases presented in Figure 3.1 and additional random uniform predictor variables[b].

Variables selected	Number of candidate predictor variables
Elevation Radiation Random variable#5	5
Elevation Radiation Random variable#5	10
Elevation Radiation Random variable#5	25
Radiation Elevation Random variable#27 Random variable#32 Random variable#1 Random variable#12	50

[a]SPSSX discriminant analysis using the following options: Stepwise,Method=Wilk's lambda, Pin=0.05, Pout=0.05, Priors=size.

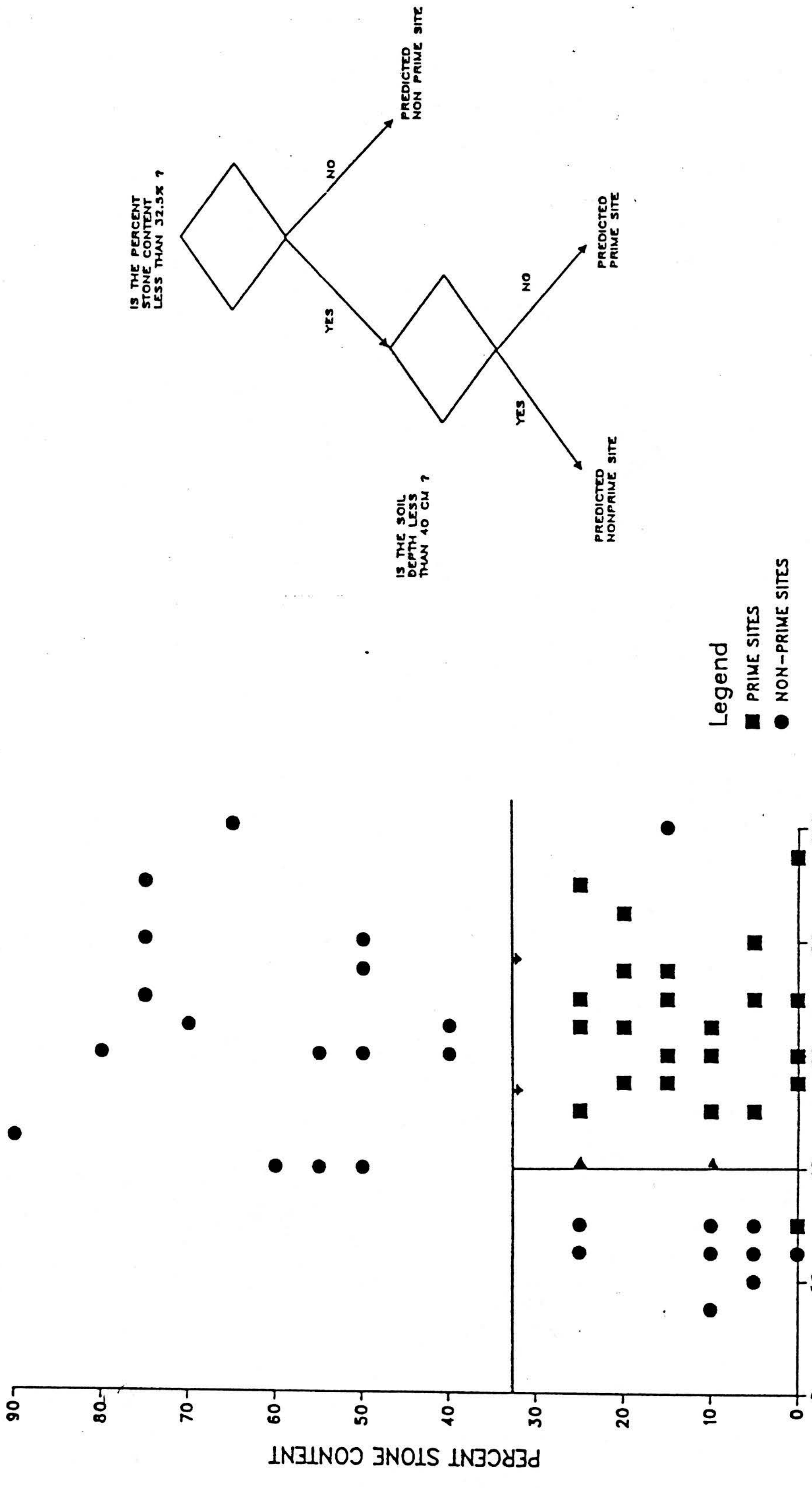
[b]Random variables generated with VAX-11 FORTRAN-77 V3.0 RAN function; values random uniform integers distributed from 0 to 99.

CLASSIFICATION TREES

Classification trees discriminate by sequentially selecting the predictor variable that best partitions sample cases into the purest class memberships. The computer program CART (classification and regression trees; California Statistical Software 1985) first determines the best discriminant boundary value for each predictor variable. Consider the hypothetical example in Figure 3.2. In this case the values would be 32.5 for "percent stone content" and 40.0 for "soil depth" . Next, the program selects the predictor variable that best splits the sample cases into the purest class memberships. In this case a split at "percent stone content" of 32.5 would result in 49 cases correctly classified. A split at soil depth of 40 cm would result in 44 cases correctly classified. Therefore the variable "percent stone content" is chosen as the first predictor variable in the model. This process of selecting variables is continued until the number of cases remaining unclassified is less than five, or until all candidate predictor variables have been used.

The classification tree at this time is likely to contain many predictor variables. Therefore the tree is likely to classify sample cases well but would classify new cases with less than apparent accuracy. The CART program corrects for this overfitting of the model by pruning the classification tree. Ten-fold cross-validation is used to estimate the actual classification accuracy of the tree at each step as follows:

Figure 3.2. Classification tree developed on 60 hypothetical sample cases.



1) Ten subsamples are selected randomly without replacement from the original sample.
2) The *v*th subsample (*v* is initially 1) is excluded and the classification tree is developed on the remaining nine subsamples.
3) Classification actual accuracy is then estimated by classifying the excluded subsample cases. Since these cases were not used to develop the classifier, they will give a better estimate of the classifier's actual accuracy than would cases used in developing the model.
4) The excluded subsample is included for classification tree development and the next subsample is excluded for validation of the developed classifier.
5) Repeat steps 1) through 4) until all subsamples have been sequentially excluded for classification accuracy estimation. The cross-validated estimate of classification accuracy is then the mean of the accuracy estimates from step 3.

Cross-validation yields classification accuracy estimates for the tree at each pruning step. Typically, a large tree with too many predictor variables has a low estimated classification accuracy because it uses spurious relations that are artifacts of the sample. On the other hand, a small tree with few variables may have a low estimated classification accuracy if it does not use all important predictor variables. Therefore, the CART program chooses the tree size with the highest cross-validated accuracy estimate.

ADVANTAGES OF CLASSIFICATION TREES

Because CART uses cross-validation during the pruning process, insignificant predictor variables are less likely to be included in the final model. To illustrate this, I added random predictor variables to the sample cases displayed in Figure 3.2. Only the original predictor variables were chosen by the CART program, even when 100 random candidate predictor variables were added (Table 3.2).

TABLE 3.2. Classification tree analysis based on cases presented in Figure 3.2 and additional random uniform variables.[a]

Variables selected	Number of candidate predictor variables
Percent stone Soil depth	5
Percent stone Soil depth	10
Percent stone Soil depth	25
Percent stone Soil depth	50
Percent stone Soil depth	100

[b]Random variables generated with VAX-11 FORTRAN-77 V3.0 RAN function; values random uniform integers distributed from 0 to 99.

Outlier sample cases are a potential problem with most least squares procedures. Classification trees are robust with respect to outliers because each sample case carries the same weight in classifier development. Classification trees are nonparametric and can use nominal, ordinal, interval and ratio-scaled predictor variables. They can be understood by anyone familiar with dichotomous keys.

Missing predictor variables are commonly handled in discriminant analsysis by either deleting cases with missing values or by substituting mean values for missing values. CART handles missing values by keeping surrogate splits based on other predictor variables. If a predictor variable is missing for a sample case, CART uses the best available (nonmissing) surrogate split. Breiman _et al._ (1984) have found classification tree accuracy loss due to missing values to be slight if predictor variables are highly correlated, and therefore good surrogate splits are available.

Empirical medical studies have found classification trees to predict as well as discriminant analysis (Goldman _et al._ 1982, Dillman and Koziol 1983, Gilpin _et al._ 1983). Neither method is best under all conditions. Discriminant analysis performs well when linear combinantions of predictor variables are important (as in Figure 3.1). Classification trees perform well when threshold values are important in predicting class membership.

CHAPTER IV

A NEW APPROACH TO SITE QUALITY MODELING

INTRODUCTION

The typical soil-site quality study involves multiple regression with site properites measured from randomly located plots. Hundreds of studies have used this approach (see reviews by Carmean 1975, Hagglund 1981, Grey 1983). However, most statistically significant regression models developed using random selected plots reveal relationships that may be important only in the range of poor to good sites. Because of this, the conventional approach may be misleading in indentifying site properties associated with the best sites. The purpose of this chapter is to contrast the conventional regression approach with a new approach that discriminates prime vs. nonprime sites.

METHODS

Seventy-five 0.10 hectare plots were randomly established within the ponderosa pine (*Pinus ponderosa*) zone in the Dixie National Forest, Utah (see Appendix II maps). Site properties from these plots were used to develop a conventional site index regression model. In addition, forty-four 0.10 hectare plots were randomly established on the best ponderosa pine sites (according to Dixie National Forest silviculturists) so that models discriminating prime sites could be developed and compared with the regession model developed from randomly selected plots. Prime sites are defined as sites having site index base age 25 years (SI_{25}) of greater than 7.6 meters (25 feet).

Within each plot, the tree with best initial 25-year diameter growth (based on increment cores) was selected as a site tree. Site trees had no observable top-damage, had healthy appearing crowns, and a past history of regular radial growth to at least 25 years at breast height. SI_{25} was measured by taking increment cores about 25 whorls above breast height until the location corresponding to 25 years less than breast height age was found. Site index was then measured to the nearest .10 meter with a steel tape.

A soil pit was dug under the site tree crown and soil samples were taken from the 10 to 20, 20 to 30, 40 to 50, and 60 to 70 depth zones in the soil profile. Soil samples were placed in a cooler and frozen within 10 days of sampling. Frozen soils were later oven-dried at 60° C and sieved to 2 mm. Samples were then analyzed in the following manner: pH by glass electrode in a 1:1 paste, percent organic carbon by the Walkley-Black method (Nelson and Sommers 1982), Bray-Kurtz extractable phosphorus (Olsen and Sommers 1982), mineralizable nitrogen (Powers 1980) determined colorimetrically (Keeny and Nelson 1982), and percent sand, silt, clay by the hydrometer method (Bouyoucos 1962).

Elevation at each plot was estimated to the nearest 30 meter with a topographic map. Slope was measured with a relaskop and recorded to the nearest percent. Aspect was recorded and later transformed with a cosine function (Gaiser 1951). Potential July 1 solar radiation was estimated as a function of slope and aspect (Swift 1976). Habitat type was determined by using keys and descriptions presented in Youngblood and Mauk (1985). Habitat types were grouped into two classes: hot-dry low site potential (PIPO/ARPA, PIPO/ARNO, PIPO/PUTR) and cool-moist high site potential (PIPO/QUGA, PIPO/SYOR, PSME, and ABCO habitat types) (Table 4.1).

Table 4.1. Description of habitat types in the study area from Youngblood and Mauk (1985).

Habitat type	Dominant understory species	Typical site
	Low Site Potential Habitat Types	
PIPO/ARPA	*Arctostaphylos patula* *Juniperus scopulorum* *Purshia tridentata*	Shallow limestone soils; south and west facing slopes
PIPO/ARNO	*Artemesia nova* *Chrysothamnus viscidiflorus* *Juniperus scopulorum* *Tetradymia canescens*	Deep sandy plains at low elevations
PIPO/PUTR	*Purshia tridentata* *Artemesia tridentata* *Quercus gambelii*	Shallow basalt or sandstone slopes
	High Site Potential Habitat Types	
PIPO/QUGA	*Quercus gambelii* *Amelanchier alnifolia* *Symphoricarpos oreophilus* *Rosa woodsii* *Carex geyeri*	Wide variety of sites; most common on non-limestone soils
PIPO/SYOR	*Symphoricarpos oreophilus* *Berberis repens* *Populus tremuloides*	Moist benches; north and east facing slopes
PSME series	*Symphoricarpos oreophilus* *Berberis repens* *Juniperus communis*	Cool slopes above PIPO series
ABCO series	*Symphoricarpos oreophilus* *Berberis repens* *Populus tremuloides*	Cool slopes and benches above PIPO series

Foward stepwise regression (SPSS Inc. 1983; F-statistic probability to enter < 0.10) was used to develop a site index model with the measurements from randomly selected plots. Linear discriminant function analysis and classification tree analysis were used to discriminate prime sites from non-prime sites. From a previous study (see Chapter V) prime sites were expected to occur only in the high potential habitat group. Therefore plots with low site potential habitat types (PIPO/ARPA, PIPO/ARNO, and PIPO/PUTR) were excluded from the prime site discrimination. Potential predictor variables for regression and discrimination models are listed in Table 4.2.

Table 4.2. Candidate predictor variables from sample cases.

Potential predictor variables	Prime site plots Mean	S.D.	N	Non-prime plots Mean	S.D.	N
Elevation (m)	2489	67	45	2445	86	75
Potential solar radiation ($cal/cm^2/day$)	972	22	45	965	51	75
Slope (%)	9.6	6.8	45	9.4	11.1	75
Organic matter (%)						
at 15 cm	2.5	0.9	45	3.5	1.3	75
at 25 cm	1.7	0.8	45	2.9	1.4	70
at 45 cm	1.1	0.7	45	2.2	1.3	66
at 65 cm	1.0	0.6	39	1.8	1.1	50
Extractable P (ppm)						
at 15 cm	105	52	45	31	45	75
at 25 cm	93	56	45	25	42	70
at 45 cm	91	61	45	17	36	66
at 65 cm	87	70	39	12	30	49
Mineralizable N (ppm)						
at 15 cm	8.8	11.6	45	11.5	5.6	75
at 25 cm	5.5	6.0	45	8.2	4.0	70
at 45 cm	4.2	4.0	45	5.6	2.7	66
at 65 cm	3.8	4.2	39	4.5	1.7	49
Soil pH						
at 15 cm	6.2	0.4	45	7.3	0.7	75
at 25 cm	6.3	0.5	45	7.4	0.7	70
at 45 cm	6.4	0.5	45	7.5	1.1	66
at 65 cm	6.5	0.6	39	7.7	0.7	49
Sand (%)						
at 15 cm	49	10	45	31	15	75
at 25 cm	48	11	45	32	13	70
at 45 cm	47	9	45	33	13	66
at 65 cm	47	10	39	30	12	49

Table 4.2 (continued).

Potential predictor variables	Prime site plots			Non-prime plots		
	Mean	S.D.	N	Mean	S.D.	N
Silt (%)						
at 15 cm	25	6	45	33	8	75
at 25 cm	24	6	45	30	7	70
at 45 cm	20	5	45	28	8	66
at 65 cm	20	5	39	29	7	49
Clay (%)						
at 15 cm	25	7	45	36	10	75
at 25 cm	28	8	45	37	11	70
at 45 cm	33	9	45	39	11	66
at 65 cm	34	9	39	40	9	49
Habitat type (0,1 dummy)	1.0	0	45	0.5	0.5	75

RESULTS AND DISCUSSION

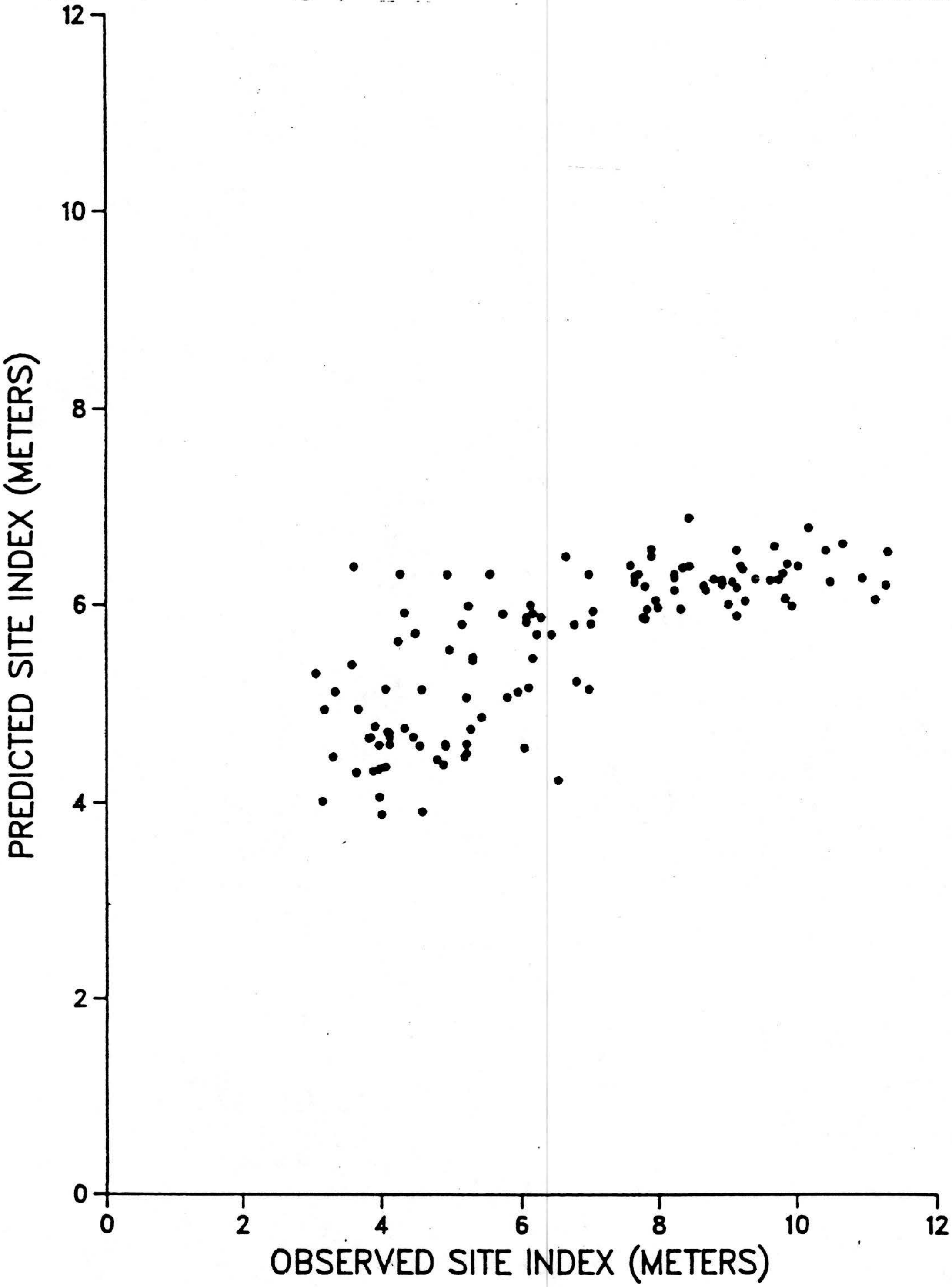
Three variables were selected by the stepwise regression procedure (Table 4.3). The weak linear relationship revealed by the regression occurred only in the site range the model was developed fromf (poor to average sites). When the regression model was applied to nonprime and prime site sample cases it consistently underestimated site index of prime site cases (Figure 4.1). Therefore extrapolation beyond the range of sites sampled was misleading. Significant regression predictor variables from randomly selected samples may not be important if one is interested in identifying site factors associated with relatively rare prime sites.

Table 4.3. Multiple regression model developed with random plots.

Variable	Coefficient	Significance
Habitat type	1.4	< 0.01
Silt (15cm)	-0.04	0.02
Slope percent	-0.02	0.07
Regression constant	6.07	< 0.00

Adjusted R^2=0.29 Standard error of estimate=3.5

Figure 4.1. Predicted site index from regression developed with nonprime plots vs. observed site index from all plots.



Both linear discriminant and classification tree analyses discriminated prime sites as a function of sand percent and pH (Figures 4.2, 4.3). Both classifiers had high apparent classification accuracy: the classification tree correctly classified 71 of 77 sample cases (92 percent) while the linear discriminant function correctly classified 66 of 77 sample cases (85 percent). The actual acuracy of the classification tree was estimated with 10-fold cross-validation (Breiman _et_ _al._ 1984) to be 88 percent. The actual accuracy of the linear discrimant function was estimated to be 80 percent by using the random bootstrap (Efron 1982).

Model reliability was also be assessed by examining misclassified sample cases. The cases misclassified by the classification tree tended to occur near the prime site - nonprime site boundary of 7.6 meters (25 feet) site index (Figure 4.4).

Figure 4.2. Linear discriminant function developed with plots having high site potential habitat types.

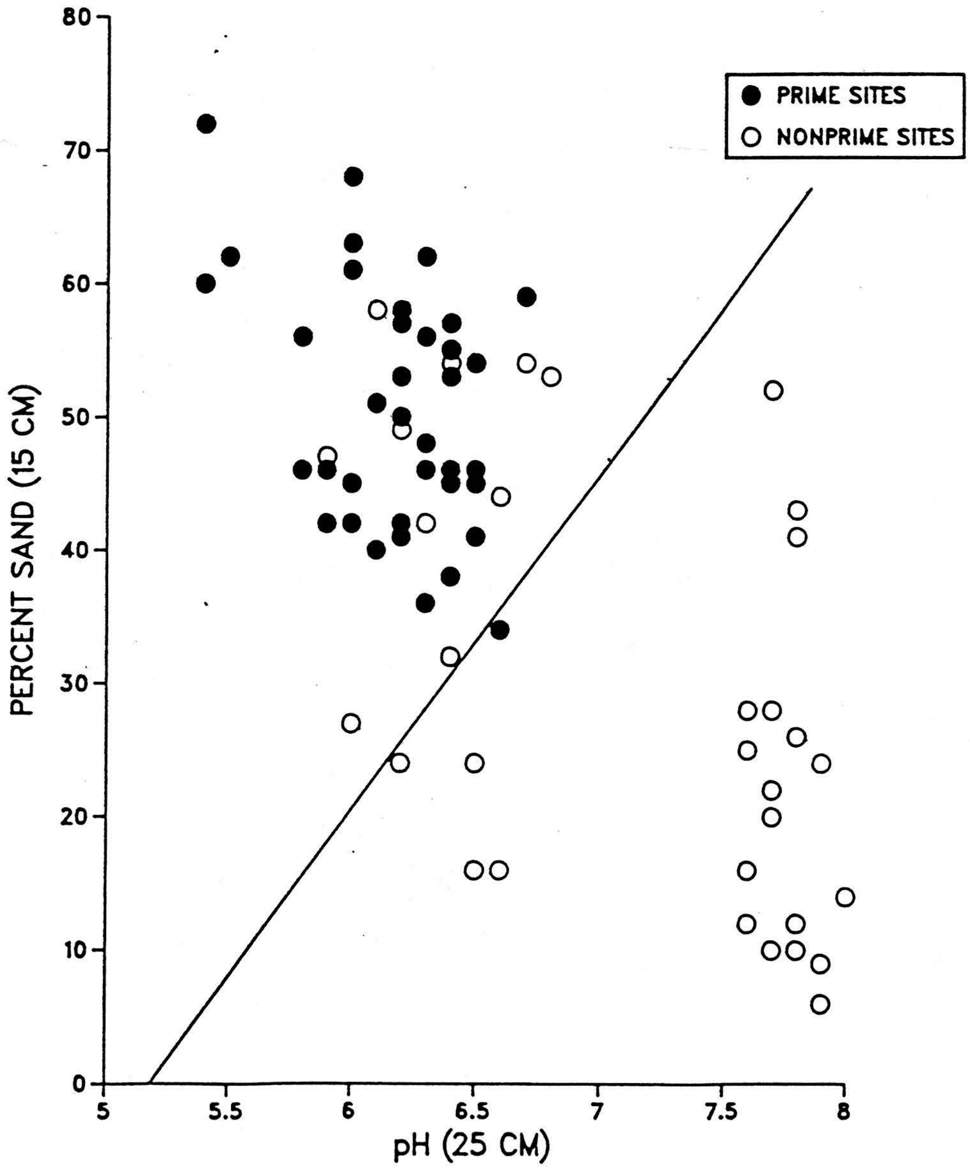


Figure 4.3. Classification tree developed with plots having high potential habitat types.

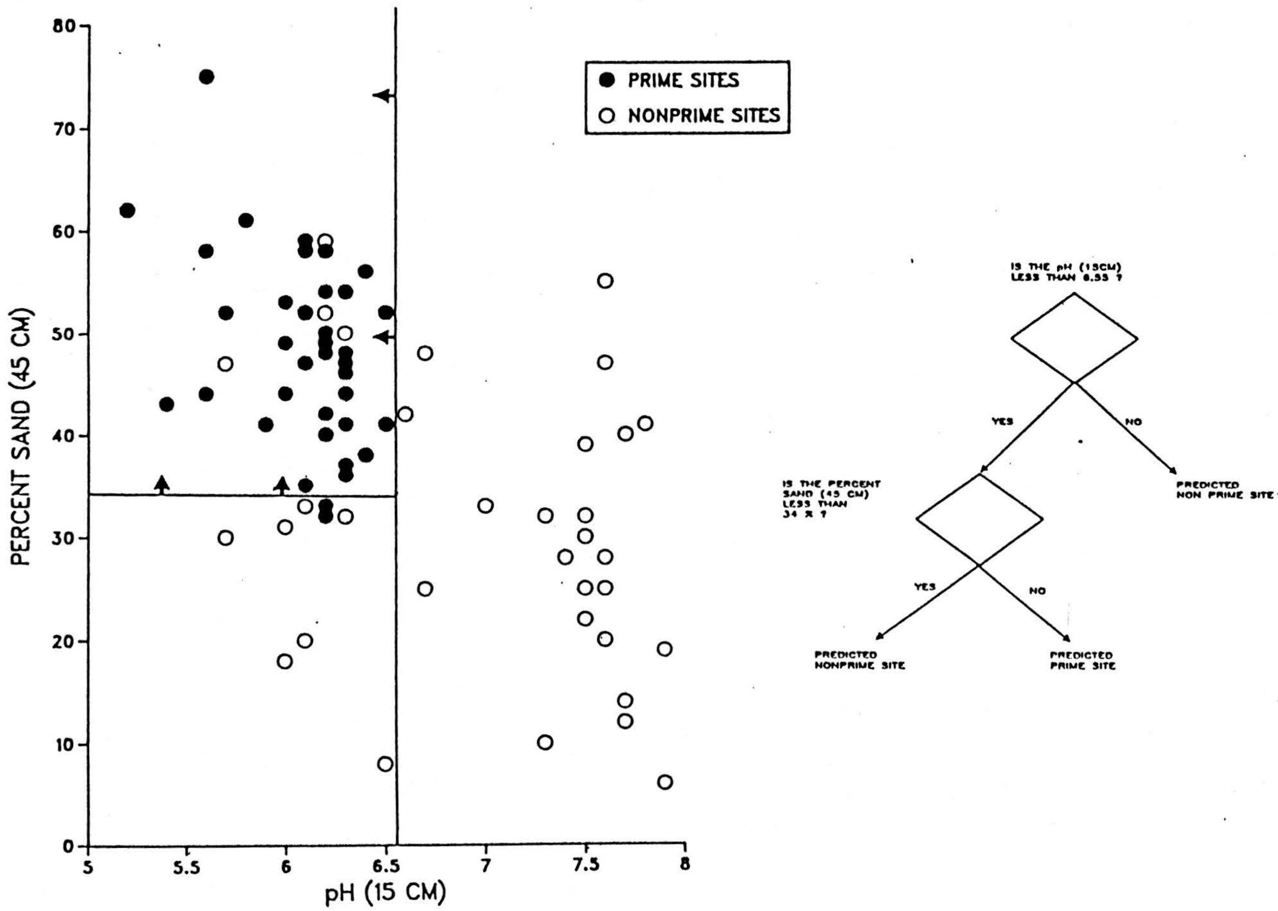
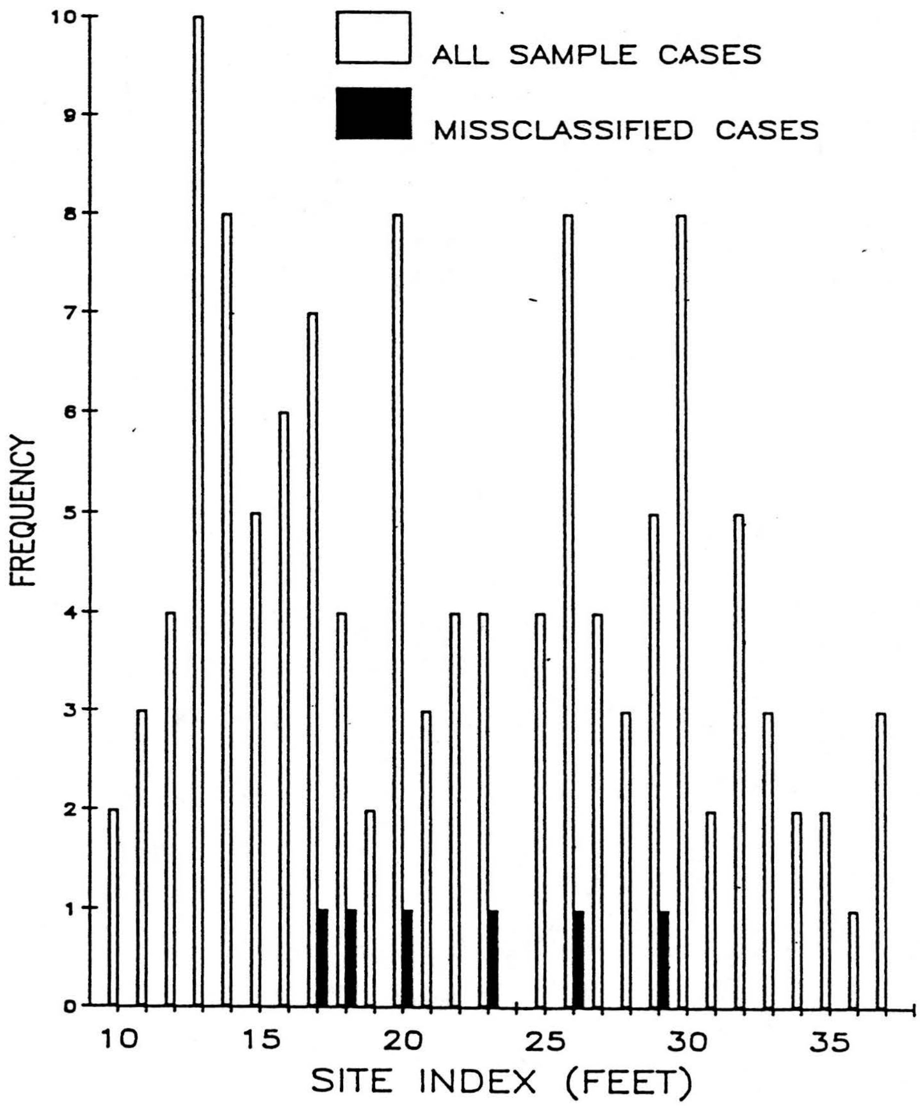


Figure 4.4. Distribution of cases misclassified by classification tree.



Prime sites were associated with high potential habitat types that had a high percent sand content. Frequent short duration thunderstorms in the study area are common throughout the summer. Therefore, deep sandy sites may have more available water for root extraction due to high infiltration and low runoff during these storms. Also these sites may have more rapid root extraction of water due to high hydraulic conductivity. Williams et al. (1963) found the best ponderosa pine sites to have the highest hydraulic conductivity in the Zuni Mountains, New Mexico.

A low pH associated with prime sites seems to be consistent with past studies. Zinke (1958) found the ponderosa pine site index to be the greatest at pH's of 6.0 to 6.5 in northwestern California. Howell (1932) found water culture ponderosa pine seedling height and root growth to be best at pH 4.0 .

Soil pH could affect many biological and chemical properties of soil that influence tree growth. For example, in this study extractable phosphorus and pH at 15 cm were negatively correlated (Pearson's $r=-0.68$, $P<0.001$).

Many of the potential predictor variables were correlated and therefore were not included in the model even though they may be useful predictors. For example, percent sand and percent clay at 15 cm were strongly correlated ($r=0.89$, $P<0.001$). Therefore, once percent sand at 15 cm is included in the discriminant model, percent clay explains little of the remaining variation and is excluded.

CONCLUSIONS

1) A regression model developed over a range of poor to good sites was biased when applied to prime sites. Models may be misleading if extrapolated beyond the site range within which they were developed.

2) Classification tree and linear discriminant analysis accurately discriminated prime sites. Prime sites were associated with high site potential habitat types that had high percent sand content and low soil pH.

CHAPTER V

PONDEROSA PINE HABITAT TYPES AS AN INDICATOR OF SITE QUALITY IN THE DIXIE NATIONAL FOREST, UTAH

INTRODUCTION

The USDA Forest Service habitat type land classification system (Pfister and Arno 1980) has been shown to be useful in predicting tree growth potential (Roe 1967, Monserud 1984, Mathiasen et al. 1986). These studies examined differences in mean tree growth across habitat type series (a series is a collection of habitat types having the same dominant by the dominant tree species at climax). To my knowledge, no studies have examined within-series habitat types as indicators of site production potential. The objective of this study is to examine site index among five habitat types within the *Pinus ponderosa* series.

METHODS

Site index base age 25 years (SI_{25}) was measured for 172 dominant ponderosa pine (*Pinus ponderosa*) trees from randomly established 0.10 hectare plots within the ponderosa pine zone of the Dixie National Forest in southern Utah (Appendix II). SI_{25} was chosen rather than site index base age 100 years because many stands in the study area were 25 to 75 years old. Height growth differences due to site differences usually begin early. For example, Oliver (1972) found that a six-year height intercept in ponderosa pine seedlings accounted for 81 percent of the variation in SI_{100}. The tree with best initial 25-year diameter growth (based on increment cores) within each plot was selected as a site tree. Site trees had no observable top damage, had healthy appearing crowns and a past history of regular radial growth to at least 25 years at breast height. SI_{25} was measured by taking increment cores in the area 25 whorls above breast height until the location corresponding to 25 years less than breast height age was found. Site index was then measured to the nearest .10 meter with a steel tape.

Habitat type was determined by using keys and descriptions published by Youngblood and Mauk (1985). Five habitat types within the ponderosa pine series were sampled (Table 5.1). Unclassified plots or those not in the ponderosa pine series were excluded from the analysis. To avoid bias in habitat type identification, habitat type was determined before site index was measured.

Table 5.1. Ponderosa pine habitat types sampled.

Habitat type	Dominant understory species	Typical site
PIPO/ARPA	*Arctostaphylos patula* *Juniperus scopulorum* *Purshia tridentata*	Shallow limestone soils; south and west facing slopes
PIPO/ARNO	*Artemesia nova* *Chrysothamnus viscidiflorus* *Juniperus scopulorum* *Tetradymia canescens*	Deep sandy plains at low elevations
PIPO/PUTR	*Purshia tridentata* *Artemesia tridentata* *Quercus gambelii*	Shallow basalt or sandstone slopes
PIPO/QUGA	*Quercus gambelii* *Amelanchier alnifolia* *Symphoricarpos oreophilus* *Rosa woodsii* *Carex geyeri*	Wide variety of sites; most common on non-limestone soils
PIPO/SYOR	*Symphoricarpos oreophilus* *Berberis repens* *Populus tremuloides*	Moist benches; north and east facing slopes

RESULTS

Site index variances differed significantly among habitat types (Bartlett F-max test, $P < 0.01$). Therefore a rank-order nonparametric test was used instead of analysis of variance in testing the null hypothesis that mean site index was the same among habitat types. Mean site index differed significantly among the five habitat types (Kruskal-Wallis test, $P < 0.0001$). However, there was considerable site index variation among the five habitat types (Fig. 5.1, Table 5.2). Because of this variation, no habitat type contained consistently good sites or consistently poor sites.

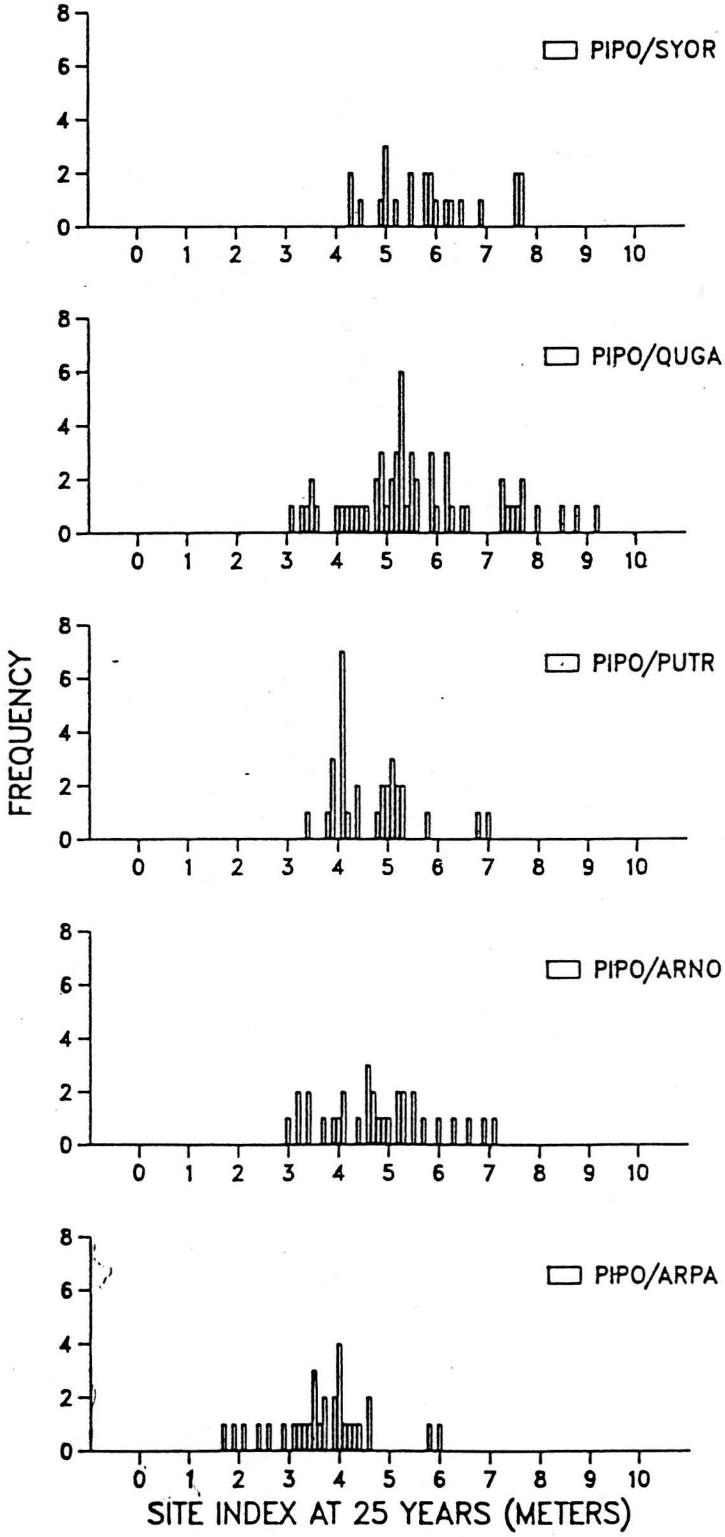
There are several possible reasons for the wide site index variation within a habitat type. A given habitat type may reflect a wide range of site conditions. For example, PIPO/SYOR is considered a moist, cool habitat type within the ponderosa pine series. However, this habitat type can range from steep slopes with at least 35 percent exposed rock to level bottoms with deep loam soils (Youngblood and Mauk 1985).

Vegetation composition on some sites may be affected by factors such as seed source prior to disturbance, time elapsed since last disturbance, and type of disturbance. For example, shrub species characteristic of the PIPO/ARPA habitat type (*Ceanothus* and *Arctostaphylos*) tend to germinate rapidly and re-sprout following surface fires (Daubenmire 1959). *Purshia tridentata* usually does not resprout and may be re-established by rodents caching *Purshia* seeds after a fire disturbance (West 1968, Sherman and Chilcote 1972). Therefore the habitat type of a given site may not be entirely dependent on environmental conditions; it may also depend on site history and chance.

Table 5.2. Random plot SI_{25} (meters) descriptive statistics by habitat type.

Habitat type	Mean	Standard deviation	Sample size
PIPO/ARPA	3.6	0.96	31
PIPO/ARNO	4.8	1.09	31
PIPO/PUTR	4.7	0.84	30
PIPO/QUGA	5.6	1.41	57
PIPO/SYOR	5.9	1.07	23

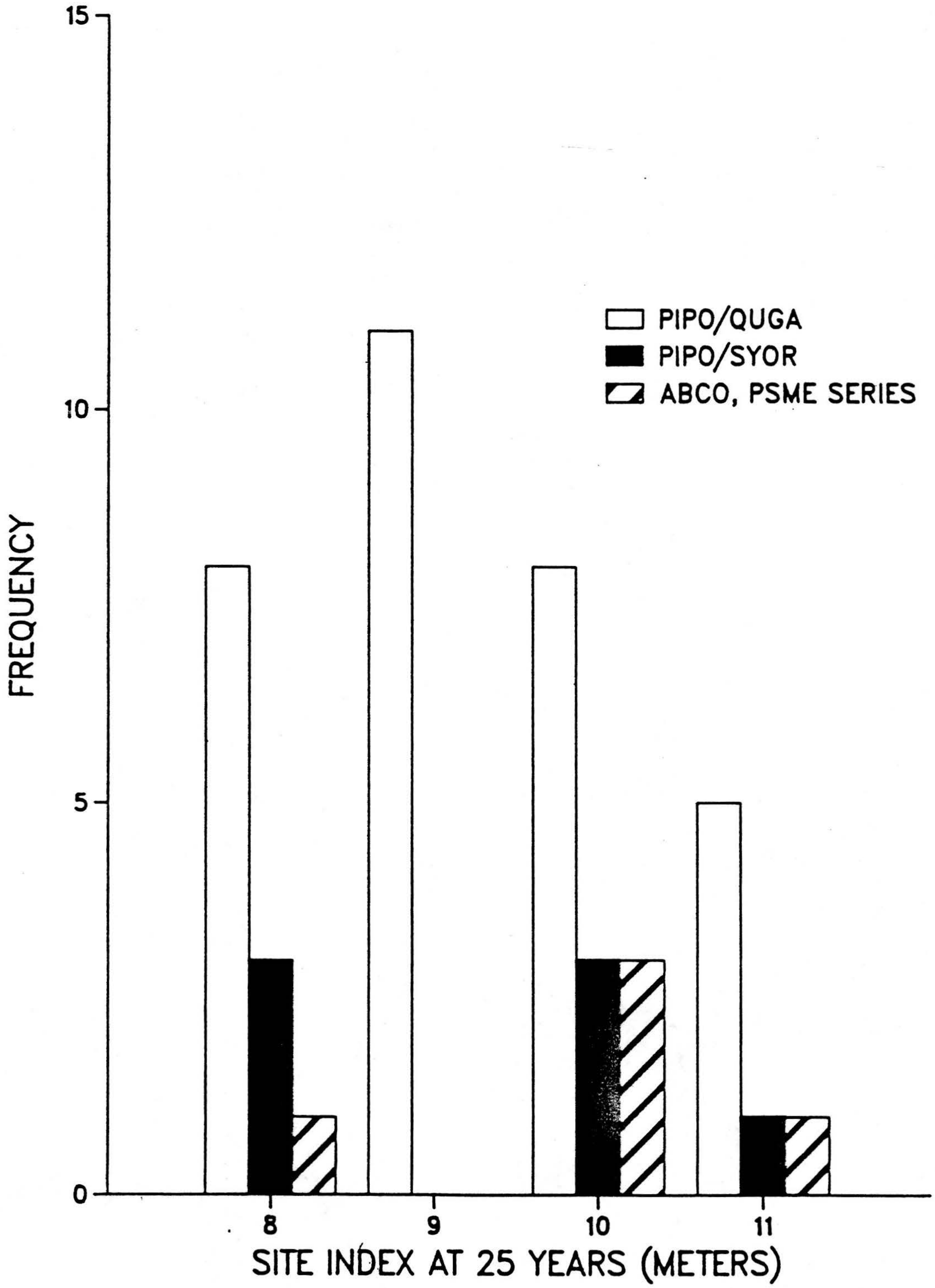
Figure 5.1. Distribution of random ponderosa pine SI_{25} among PIPO habitat types.



Although there was considerable site index variation within habitat types, the best sites occurred only in the PIPO/QUGA and PIPO/SYOR habitat types. Therefore habitat types are useful when used in conjunction with additional site factors in identifying prime sites (see chapter IV).

In a related study (chapter IV), I sampled the best ponderosa pine sites (according to Dixie National Forest silviculturalists) within the Dixie National Forest. Forty-four 0.10 hectare plots were randomly established within these prime site areas and SI_{25} at each plot was measured from the dominant ponderosa pine with the best initial 25-year diameter increment. Random plots within these areas were always identified as PIPO/QUGA, PIPO/SYOR or habitat types within the Douglas-fir (*Psuedotsuga menziesii*) or white fir (*Abies concolor*) series (Figure 5.2).

Figure 5.2. Distribution of prime ponderosa pine SI_{25} by habitat type.



CONCLUSIONS

1. The best ponderosa pine sites sampled consistently occurred in PIPO/QUGA, PIPO/SYOR, PSME or ABCO habitat types.
2. The range of SI_{25} within any one habitat type was broad. Therefore habitat type should not be used alone in predicting the best ponderosa pine sites. However, habitat type when used with other site variables is useful in predicting prime sites.

CHAPTER VI

EFFECT OF ASPECT ON PONDEROSA PINE HEIGHT AND DIAMETER GROWTH

INTRODUCTION

Site quality studies have generally transformed aspect azimuth using a cosine function (see reviews by Carmean 1975, Hagglund 1981). This transformation produces a maximum at the northeast and a minimum at the southwest aspect. Empirical studies in the eastern United States have found south and west facing slopes to be least productive, and north and east facing slopes to be most productive (Trimble and Weitzman 1956, Lee and Sypolt 1974, Auchmoody and Smith 1979, Tajchman and Wiant 1983, Hicks and Frank 1984). It is believed that northerly aspects in the northern hemisphere have greatest forest production because of greater available soil moisture (Werling and Tajchman 1984), more rapid nutrient cycling (Hicks and Frank 1984), and more favorable ambient and soil temperatures (Lee and Sypolt 1974).

Optimal aspect for forest production may be the aspect receiving maximum solar radiation subject to available water and optimal temperature constraints. Therefore optimal aspect may not always be northerly. For example, at extreme northerly latitudes in the northern hemisphere, southerly aspects appear to have the best black spruce (*Picea mariana*) sites (Lowry 1975).

Roise and Betters (1981) theorized that the optimal aspect for tree growth varies with elevation. They showed that aspen (Populus tremuloides) occurs more frequently on north-facing slopes at low elevations and on south-facing slopes at high elevations. However, to my knowledge this relationship has not been empirically examined for individual tree growth. Running (1984) used computer simulation to show in theory that south slopes may produce higher seasonal photosynthesis than north slopes when energy, rather than water, becomes the factor limiting physiological activity. The first objective of this chapter is to determine whether ponderosa pine (Pinus ponderosa) at its upper elevational limit exhibits greatest growth on southerly slopes.

Optimal aspect may also vary depending upon the measurement of tree growth used. Stage (1976) found the optimal aspect for western white pine (Pinus monticola) site index to be northeast and the optimal aspect for individual basal area growth to be southwest. Greater diameter growth on southwest slopes may have been due to a lower tree density. However, diameter growth may be greatest on southerly slopes early in the life of a stand when density has minimal effect on individual diameter growth. The second objective of this chapter is to examine mean initial 10-year diameter increment and mean site index on various aspects to determine whether the greatest mean site index and diameter increment occur on the same aspect.

METHODS

Two cinder cones, Bowers Knoll and Henrie Knoll, on the Cedar City District, Dixie National Forest were chosen as study sites (see Appendix II maps). Both sites are free of topographic shading, have uniform parent material, and are in the upper elevational limit of ponderosa pine. Bowers Knoll ranges in elevation from 2530 to 2604 meters; Henrie Knoll ranges from 2710 to 2821 meters above sea level.

Dominant and codominant ponderosa pine (*Pinus ponderosa*) were sampled at mid-slope as site trees. Site trees had no observable top damage, had healthy appearing crowns, and a past history of unsuppressed radial growth. Each site tree was increment cored at breast height (1.4 m above ground level) to determine total breast height age and initial 10-year diameter growth increment. Tree height was measured indirectly with a relaskop and computed after measuring slope distance to the site tree (Long and Mohai 1986). Site index at base age 100 years was computed using procedures described by Fisher (1980). Slope aspect was determined with a hand compass as either north, northeast, east, southeast, south, southwest, west, or northwest. Aspects were grouped for analysis into northerly aspects (north, northeast, northwest) and southerly aspects (south,southeast, southwest).

RESULTS AND DISCUSSION

Mean initial ten-year diameter increment was significantly greater on southerly slopes (Table 6.1). Mean site index was significantly greater on southerly slopes at Henrie Knoll; no significant difference was observed from Bowers Knoll samples.

Greater individual initial diameter growth on southerly slopes might be due to a lower tree density. Diameter growth was measured for the first ten years at breast height to minimize this density effect. Greater mean diameter growth on southerly slopes might be due to an extended diameter growing season. Ponderosa pine diameter growth usually coincides with the rainy late summer period on the Colorado plateau and stops once a minimum ambient temperature occurs (Mace and Wagle 1964). At the upper elevational limit, ponderosa pine diameter growth on northerly slopes may be slower due to lower ambient temperatures as the season progresses when compared with southerly slopes (Figure 6.1).

Figure 6.1. Potential solar radiation on 30 percent slopes at 38 degrees north latitude.

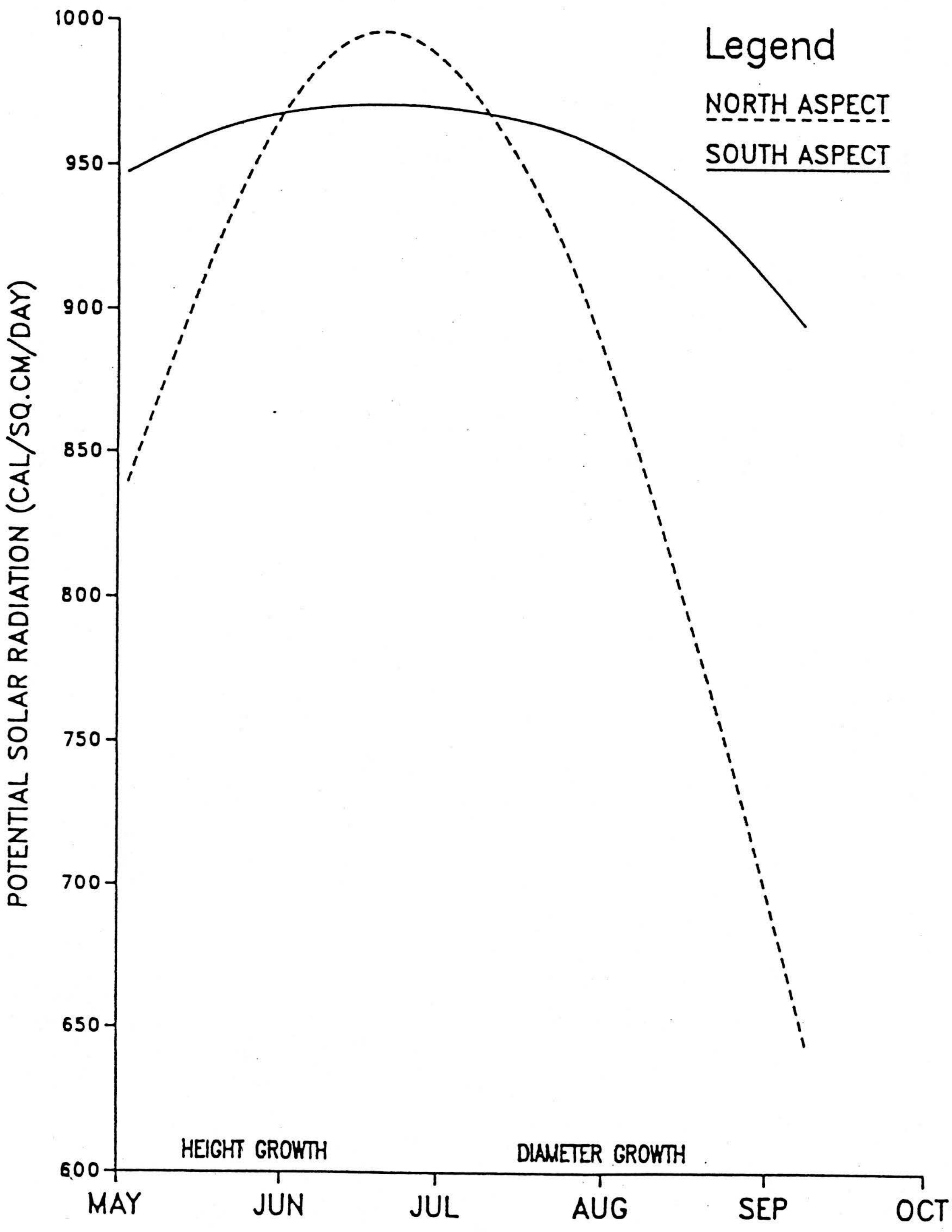


Table 6.1. Student T-test of mean site index$_{100}$ and mean 10 year diameter increment by aspect group.

	Northerly Aspects			Southerly Aspects			
Site Index (m)	N	Mean	S.D.	N	Mean	S.D.	P-value
Bowers Knoll	26	18.9	2.5	22	19.0	1.7	0.95
Henrie Knoll	20	17.1	2.0	18	18.7	1.2	0.008
10 year diameter increment(cm)	N	Mean	S.D.	N	Mean	S.D.	P-value
Bowers Knoll	26	2.7	0.6	22	3.4	0.6	<0.001
Henrie Knoll	20	2.1	0.5	18	3.0	0.6	<0.001

Soil moisture has a strong influence on height growth in pines (Perala 1985). This may be especially true on the Colorado Plateau because May and June are usually droughty. Pearson (1918) found a strong correlation between precipitation in April/May and ponderosa pine height growth in northern Arizona. Northerly slopes during the height growth period may be the more favorable sites due to a slower spring snowmelt and therefore greater available soil water during height growth. This may be why maximum site index is often reported for northerly aspects, except at extreme upper elevational limits (such as Bowers Knoll) where low temperatures rather than low soil moisture may be limiting.

Stage (1976) found southerly aspects to be superior for individual basal area growth. This may be due to less individual tree competition on the lower density southerly slopes. It may also be due to differences in radiation loading as the diameter season progresses allowing for greater diameter growth and a longer diameter growing season on southerly slopes. Further research with controls on density is needed to determine whether optimal aspects for diameter and height growth differ.

Aspect is often used in digital terrain models to predict site index (Ferguson 1981, Fox et al. 1985). These models may be misleading if applied to areas with different elevations or latitudes. For example, a digital terrain model developed to predict aspen production in Colorado as a function of slope,aspect, elevation and parent material may predict poorly when applied to aspen stands in Alaska.

CONCLUSIONS

1. The optimal aspect for ponderosa pine growth was not constant but changes with elevation. At low elevations north-facing slopes have relatively better sites. The highest elevation site studied had greatest mean site index and mean diameter increment on southerly slopes.
2. Height growth and diameter growth occur at different times of the year. Maximum diameter growth at a species upper elevational limit may occur on warmer aspects due to higher ambient temperatures during diameter growth and an extended growing season. Diameter growth was significantly greater on southerly slopes witin the two upper elevation zones studied.
3. Maximum height growth below a species upper elevational limit may occur on northerly aspects that have the highest soil moisture content during the height growth period in the spring. Height growth was significantly greater on southerly slopes only on the highest elevation zone studied.

CHAPTER VII

GENERAL SUMMARY AND CONCLUSIONS

The conventional site quality approach is to use multiple regression procedures with data from randomly selected plots to develop a model that predicts site index as a function of soil,topographic and vegetation properties. This approach is poor if many candidate predictor variables are available relative to sample size because the resulting model is likely to contain biologically insignificant predictor variables. Such a model is misleading and may perform poorly if validated with new independent data.

Another problem with the conventional approach is that relationships revealed by models developed with randomly selected plots may not hold for relatively rare prime sites. Since intensive silviculture often is only economical on prime sites, models that discriminate prime vs. nonprime sites may be more useful.

A new approach is presented using classification tree analysis and linear discriminant function analysis to discriminate prime vs. nonprime ponderosa pine sites. Classification tree analysis uses cross-validation to develop a model with the best estimated actual accuracy. Both the classification tree and linear discriminant function predicted prime and nonprime sites as a function of habitat type, percent sand, and soil pH. Cross-validation was used to estimate the actual classification tree accuracy as 88 percent. The random bootstrap was used to estimate the linear discriminant function classification accuracy as 80 percent. A multiple regression model was developed with the nonprime site plots and was biased when applied to prime site plot data.

Forest habitat types within the ponderosa pine series were examined as site quality indicators. The range within any one habitat was broad. However, the best ponderosa pine sites consistently occurred in only *Pinus ponderosa*/*Quercus gambelii* and *Pinus ponderosa*/*Symphoricarpos oreophilus* habitat types, or within the *Pseudotsuga menziesii* and *Abies concolor* series. Therefore forest habitat type when used with other site variables may be useful in predicting prime sites.

The effect of aspect at the upper elevational limit of ponderosa pine was examined by comparing mean site index and mean initial 10-year diameter increment on southerly and northerly slopes from two cinder cones. Southerly aspects had on both cinder cones had greater mean diameter increment. Southerly aspects on the highest elevation cinder cone had the greatest mean site index. There was no significant difference in mean site index on the lower elevation cinder cone. The effect of aspect is often assumed to be constant in most site quality models. Optimal growth probably varies with elevation, latitude, and species. The optimal aspect may be the aspect receiving maximum solar radiation subject to available water and optimal temperature constraints.

RECOMMENDATIONS FOR FURTHER RESEARCH

Most forest site quality models use relatively easily measured variables that correlate with measured tree growth. For example, sand percent is correlated with tree growth presumably because it affects available soil water. Soil pH is correlated with tree growth presumably because it affects nutrient availability. Such models are fine for planning purposes. However, they are weak in explaining processes that affect site quality.

Austin et al. (1984) have argued that more biologically relevant variables need to be developed or measured. For example, slope and transformed aspect are often included in plant distribution models. These variables are associated with the amount of solar radiation a site receives. However, site potential solar radiation can be calculated as a function of slope, aspect and latitude (Swift 1976). A potential radiation index may be more biologically relevant than slope and transformed aspect since it more closely reflects a factor (solar radiation) affecting plant growth. Austin (1984) has found the potential radiation index to be a better predictor of plant species distributions than simple measures of slope and aspect. The potential solar radiation index would probably also be a better predictor variable than slope and aspect in forest site quality models.

Many processes affecting site quality are dynamic. Yet most variables used in site quality models are static measurements taken at one time during the growing season. Also many soil measurements are taken in the laboratory on disturbed soil samples rather than in the field on relatively undisturbed samples. New measurement techniques need to be developed that can be used on-site, with minimum soil disturbance, and can be monitored throughout the growing season. For example, the resin bag method (Binkley and Matson 1983) can be used with intact soil cores for assessing soil nitrogen availability, nitrogen mineralization, and nitrification (DiStefano and Gholz 1986). Such a field technique is valuable because patterns of nitrogen availability and loss rather than a static meausure of available nitrogen may be important. Similar field techniques need to be developed to measure available soil water and other nutrients.

Practical site quality models that use easily measured variables will continue to be valuable. In addition, models using measured dynamic site factors such as nutrient and water availability need to be developed to improve our understanding of processes affecting site quality.

LITERATURE CITED

Austin, M. P., Cunningham, R. B., and P. M. Fleming. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. Vegatatio 55:11-27.

Auchmoody, L. R. and H. C. Smith. 1979. Oak soil-site relationships in northwestern West Virginia. USDA For. Serv. Res. Pap. NE-434.

Berk, K. N. 1978. Comparing subset regression procedures. Technometrics 20:1-6.

Binkley, D. and P. Matson. 1983. Ion exchange resin bag method for assessing forest soil nitrogen availabilty. Soil Sci. Soc. Am. J. 47:1050-1052.

Breiman, L., Friedman, J. H., Olshen, R. A., and C. J. Stone. 1984. Classification and Regression Trees. Wadsworth, Inc. Belmont, Ca.

Bouyoucos, G. J. 1962. Hydrometer method improved for making particle size analysis of soils. Agron. J. 54:464-465.

Buckland, S. T. 1984. Monte carlo confidence intervals. Biometrics 40:811-817.

Califorinia Statistical Software, Inc. 1985. Classification and regression trees (CART). Lafayette, Ca.

Carmean, W. H. 1975. Forest site quality evaluation in the United States. Adv. Agron. 27:209-269.

Cavallaro, J. L., Menke, J. W. and W. A. Williams. 1980. Use of discriminant analysis and other statistical methods in analyzing microhabitat utilization of dusky-footed woodrats. In The use of multivariate statistics in studies of wildlife habitat, D. E. Capen, ed. USDA For. Serv. Gen. Tech. Rep. RM-87. pp. 222-231.

Corns, I. G. W. and D. J. Pluth. 1984. Vegetational indicators as independent variables in forest growth prediction in west-central Alberta, Canada. For. Ecol. Manage. 9:13-25.

Daubenmire, R. F. 1959. Plants and environment. A textbook of plant autecolgy. John Wiley and Sons, New York. 422 pp.

Diehr, G. and D. R. Hoflin. 1974. Approximating the distribution of the sample R^2 in best subset regressions. Technometrics 16:317-320.

Dillman, R. O. and J. A. Koziol. 1983. Statistical approach to immunosuppresssion classification using lymphocyte surface markers and functional assays. Cancer Res. 43:417-421.

DiSefano, J. F. and H. L. Gholz. 1986. A proposed use of ion exchange resins to measure nitrogen mineralization and nitrification in intact soil cores. Comm. Soil Sci. Plant Anal. 17:989-998.

Draper, N. R., Guttman, I. and H. Kanemasu. 1971. The distribution of certain regression statistics. Biometrika 58:295-298.

Efron, B. 1982. The jacknife, the bootstrap and other resampling plans. SIAM Publications, Philadelphia, PA.

Efron, B. 1983. Estimating the error rate of a prediction rule: improvements on cross-validation. J. Am. Stat. Assoc. 78:316-331.

Ferguson, B. K. 1981. The use of overlays in site quality mapping. Can. J. For. Res. 11:361-369.

Fisher, R. F. 1980. Site index and height growth curves for ponderosa pine in southern Utah. Unpublished.

Flack, V. F. and P. C. Chang. 1987. Frequency of selecting noise variables in subset regression analysis: a simulation study. Am. Stat. 41:84-86.

Fox, L., Brockhaus, J. A. and N. D. Tosta. 1985. Classification of timberland productivity in northwestern California using LANDSAT, topographic, and ecological data. Photogram. Eng. and Remote Sensing 51:1745-1752.

Frank, P. S. Jr., Hicks, R. R. Jr. and E. J. Harner Jr. 1984. Biomass predicted by soil-site factors: a case study in north central West Virginia. Can. J. For. Res. 14:137-140.

Freedman, D. A. 1983. A note on screening regression equations. Am. Stat. 37:152-155.

Gaiser, R. N. 1951. Relation between topography, soil characteristics, and the site index of white oak in south-eastern Ohio. USDA For. Serv. Cent. States Exp. Stn. Tech. Paper 121.

Gasana, J. K. and H. Loewenstein. 1984. Site classification for Maiden's Gum, Eucalyptus globus subsp. Maidenii, in Rwanda. For. Ecol. Manage. 8:107-116.

Gilpin, E., Olshen, R., Henning, H., and J. Ross Jr. 1983. Risk prediction after myocardial infarction: comparison of three multivariate methodologies. Cardiology 70:73-84.

Goldman, L., Weinberg, M., Weisberg, M., Olshen, R., Cook, F., Sargent, R. K., Lamas, G. A., Dennis, C., Deckelbaum, L., Fineberg, H., Stiratelli, R., and the Medical Housestaffs at Yale-New Haven Hospital and Brigham and Women's Hospital. 1982. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. New England J. Med. 307:588-596.

Gong, G. 1986. Cross-validation, the jacknife, and the bootstrap: excess error estimation in forward logistic regression. J. Am. Stat. Assoc. 81:108-113.

Gordon, L. and R. A. Olshen. 1985. Tree-structured survival analysis. Cancer Treat. Rep. 69:1065-1068.

Grey, D. C. 1983. The evaluation of site factor studies. S. Afr. For. J. 127:19-22.

Hagglund, B. 1981. Evaluation of forest site productivity. For. Abst. 42:515-527.

Harding, R. B., Grigal, D. F. and E. H. White. 1985. Site quality evaluation for white spruce plantations using discriminant analysis. Soil Sci. Soc. Am. J. 49:229-232.

Harner, E. J. and R. C. Whitmore. 1980. Robust principal component and discriminant analysis of two grassland bird species habitat. In The use of multivariate statistics in studies of wildlife habitat, D. E. Capen, ed. USDA For. Serv. Gen. Tech. Rep. RM-87. pp. 209-221.

Hicks, R. R. Jr. and P. S. Frank Jr. 1984. Relationship of aspect to soil nutrients, species importance and biomass in a forested watershed in West Virginia. For. Ecol. Manage. 8:281-291.

Howell, J. Jr. 1932. Relation of western yellow pine seedlings to the reaction of the culture solution. Plant. Physiol. 7:657-671.

Huberty, C. J. 1972. Regression analysis and 2-group discriminant analysis. J. Exp. Ed. 41:39-41.

_____________. 1975. Discriminant analysis. Rev. Ed. Res. 45:543-598.

Jones, J. R. 1969. Review and comparison of site evaluation methods. USDA For. Serv. Res. Paper RM-51.

Keeny, D. R. and D. W. Nelson. 1982. Nitrogen---inorganic forms. <u>In</u>: A. L. Page (ed). Methods of soil analysis. Part 2. Agron. 9:672-676. Am. Soc. Agr., Madison, Wis.

Lee, R. and C. R. Sypolt. 1974. Toward a biophysical evaluation of forest site potential. For. Sci. 20:145-154.

Long, J. N. and P. Mohai. 1986. A simplified procedure for measuring tree heights. West. J. Appl. For. 1:21-22.

Lovell, M. C. 1983. Data mining. Rev. Econ. Stat. 65:1-12.

Lowry, G. L. 1975. Black spruce site quality as related to soil and other site conditions. Soil Sci. Soc. Am. Proc. 39: 125-131.

Mace, A. C. Jr. and R. F. Wagle. 1964. A measure of the effect of soil and atmospheric moisture on the growth of ponderosa pine. For. Sci. 10:454-460.

Mathiasen, R. L., Blake, E. A., and C. B. Edminster. 1986. Estimates of site potential for Douglas-fir based on site index for several southwestern habitat types. Great Basin Nat. 46:277-280.

Monserud, R. A. 1984. Height growth and site index curves for inland Douglas-fir based on stem analsis data and forest habitat. For. Sci. 30:943-965.

Morrison, M. L. 1984. Influence of sample size on discriminant function analysis of habitat use by birds. J. Field Ornithol. 55:330-335.

Nelson, D. W. and L. E. Sommers. 1982. Total carbon, organic carbon, and organic matter. In: A. L. Page (ed). Methods of soil analysis. Part 2. Agronomy 9:570-571. Am. Soc. Agr., Madison, Wis.

Neter J. and W. Wasserman. 1974. Applied linear statistical models. Richard D.Irwin, Inc. Homewood, Illinois.

Oliver, W. W. 1972. Height intercept for estimating site index in young ponderosa pine plantations and natural stands. USDA For. Serv. Res. Note PSW-276.

Olsen, S. R. and L. E. Sommers. 1982. Phosphorus. In: A. L. Page (ed). Methods of soil analysis. Part 2. Agronomy 9:416-418. Am. Soc. Agr., Madison, Wis.

Perala, D. A. 1985. Predicting red pine shoot growth using growing degree days. For. Sci. 31:913-925.

Pfister, R. D. and S. F. Arno. 1980. Classifying forest habitat types based on potential climax vegetation. For. Sci. 26:52-70.

Powers, R. F. 1980. Mineralizable soil nitrogen as an index of nitrogen availability to forest trees. Soil Sci. Soc. Am. J. 44:1314-1320.

Rencher A. C. and F. C. Pun. 1978. Inflation of R^2 in best subset regression. Technometrics 22:49-53.

Roe, A. L. 1967. Productivity indicators in western larch forests. USDA For. Serv. Res. Note INT-59. 4 pp.

Roise, J. P. and D. R. Betters. 1981. An aspect transformation with regard to elevation for site productivity models. For. Sci. 27:483-486.

Page, G. 1976. Quantitative evaluation of site potential for spruce and fir in Newfoundland. For. Sci. 22:131-143.

Pearson, G. A. 1918. The relation between spring precipitation and height growth of western yellow pine saplings in Arizona. J. For. 16:677-689.

Pope, P. T. and J. T. Webster. 1972. The use of an F-statistic in stepwise regression procedures. Technometrics 14:327-340.

Sherman, R. J. and W. W. Chilcote. 1972. Spatial and chronological patterns of _Purshia tridentata_ as influenced by _Pinus ponderosa_. Ecology 53:294-298.

SPSS Inc. 1983. SPSSX user's guide. McGraw-Hill. Chicago, Illinois.

Stage, A. R. 1976. An expression for the effect of aspect, slope, and habitat on tree growth. For. Sci. 22:457-460.

Sutherland, D. H., Olshen, R., Cooper, L. and S. L. Y. Woo. 1980. The development of mature gait. J. Bone Joint Surg. [Am.] 62:336-353.

Swift, L. W. 1976. Algorithm for solar radiation on mountain slopes. Water Resour. Res. 12:108-112.

Tom, C. H. and L. D. Miller. 1979. Forest site productivity mapping in the coniferous forests of Colorado with LANDSAT imagery and landscape variables. In: Proceedings of the thirteenth international symposium on remote sensing of the environment, The Univ. of Michigan, Ann Arbor, April 23-27, pp. 675-692.

__________________________. 1980. Forest site index mapping and modeling. Photogram. Eng. and Remote Sensing. 46:1585-1596.

Trimble, G. R. Jr. and S. Weitzman. 1956. Site index studies of upland oaks in the northern Appalachians. For. Sci. 1:162-173.

Turvey, N. D., Rudra, A. B. and J. Turner. 1986. Characteristics of soil and productivity of *Pinus radiata* (D. Don) in New South Wales. I. Relative importance of soil physical and chemical parameters. Aust. J. Soil Res. 24:95-102.

Werling, J. A. Jr. and S. J. Tajchman. 1984. Soil thermal and moisture regimes on forested slopes of an appalachian watershed. For. Ecol. Manage. 7:297-310.

West, N. E. 1968. Rodent-influenced establishment of ponderosa pine and bitterbrush seedlings in central Oregon. Ecology 49:1009-1011.

Williams, B. K. 1983. Some observations on the use of discriminant analysis in ecology. Ecology 64:1283-1291.

Williams, J. A., Leven, A. A. and H. E. Dregne. 1963. Relation of soil properties to ponderosa pine production in the Zuni Mountains, New Mexico. Proc. 2nd N. Am. For. Soils Conf. Corvallis, Oregon. pp381-398.

Youngblood, A. P. and R. L. Mauk. 1985. Coniferous habitat types of central and southern Utah. USDA For. Serv. Gen. Tech. Rpt. INT-87. 89pp.

Zar, J. H. 1974. Biostatistical Analysis. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Zinke, P. J. 1958. Site quality for Douglas-fir and ponderosa pine in northwestern California as related to climate, topography, and soils. Proc. Soc. Am. For. 167-171.