Classification and Regression Tree Analysis for Assessing Hazard of Pine Mortality Caused by *Heterobasidion annosum*

F. A. BAKER, Department of Forest Resources, Utah State University, Logan 84322-5215; DAVID L. VERBYLA, Department of Forest Resources, University of Idaho, Moscow 83843; C. S. HODGES, JR., Department of Plant Pathology, North Carolina State University, Raleigh 27607; and E. W. ROSS, USDA Forest Service, Washington, DC 20250

ABSTRACT

Baker, F. A., Verbyla, D. L., Hodges, C. S., Jr., and Ross, E. W. 1993. Classification and regression tree analysis for assessing hazard of pine mortality caused by *Heterobasidion annosum*. Plant Dis. 77:136-139.

A classification tree based on soil variables was developed to predict the extent of mortality of slash (*Pinus elliottii*) and loblolly (*P. taeda*) pines caused by *Heterobasidion annosum*. Data were obtained from unthinned plantations inoculated with *H. annosum*. The numbers of dead trees plus living trees with lethal infections (bearing basidiocarps of *H. annosum*) on each of 152 plots on 16 sites were counted after 5 yr. Hazard of lethal root disease was arbitrarily scored as high if seven or more trees on a plot were affected and low if six or fewer trees were affected. Classification and regression tree analysis (CART) revealed that percent silt in the A horizon and pH of the A horizon were important predictor variables for hazard category. The apparent accuracy of model classification was 85%, and a 10-fold cross-validation estimate of model classification accuracy was 80%. CART provides a useful method for improving the accuracy of disease hazard rating.

Annosus root disease, caused by *Heterobasidion annosum* (Fr.:Fr.) Bref., is one of the most destructive root decays of North American conifers. The disease is especially damaging in thinned plantations (11). Although *H. annosum* may not kill southern pines replanted on infested sites (8), treatment to control the disease is recommended if stands on high-hazard sites are thinned. Site hazard ratings have been based on certain chemical and physical soil characteristics that were recognized as determinants of mortality caused by *H. annosum* (2,5,10). Site hazard ratings have been used to determine the need for applying preventive treatments (1,6,9).

Journal Paper No. 3942 of the Utah Agricultural Experiment Station, Utah State University, Logan 84322.

Accepted for publication 13 June 1992.

In 1964, two of us (CSH and EWR) began a study to refine the relationship between soil characteristics and the mortality resulting from annosus root disease and thus increase the accuracy of hazard ratings. This study demonstrated wide variation in the extent of mortality on sites with different soil characteristics, but no statistically significant correlations between soil characteristics and the extent of mortality were detected by the methods available then, such as analysis of variance, stepwise regressions, and discriminant analysis (Hodges and Ross, *unpublished*).

Breiman et al (3) recently developed a statistical method called classification and regression tree analysis (CART). CART attempts to construct a binary decision tree by selecting the most useful variables from a set of candidate predictor variables. CART has been used in Montana to predict the proportion of stands with mortality caused by root diseases (4). We applied CART to the data of Hodges and Ross and found significant correlations between certain soil characteristics and mortality caused by annosus root disease. The results of this analysis are reported here.

MATERIALS AND METHODS

Plot establishment. During the winter of 1964–1965, 16 study sites were selected in South Carolina, North Carolina, Georgia, and Virginia; nine were in the coastal plain and seven were in the Piedmont (Table 1). Fifteen of the sites were in 13- to 20-yr-old unthinned plantations and one was in a natural stand of loblolly (*Pinus taeda* L.) or slash (*P. elliottii* Engelm.) pine with no mortality in dominant or codominant trees and no evidence of annosus root disease. Sites ranged from high to low in annosus root disease hazard ratings. Ten 0.04-ha plots were established at each of the 15 plantations and five plots were established at the natural stand, for a total of 155 plots initially. Subsequently, three plots at the Bertie County, North Carolina, site were logged, which left a total of 152 plots for analysis.

The individual plots were selected to contain a minimum of missing trees in an attempt to equalize the chance of root-to-root spread from an inoculated plot center. Disease centers were initiated by cutting four trees, two in each of two rows, in the center of each plot and then immediately inoculating the stumps. Each stump surface was sprayed to run-off with a conidial suspension of *H. annosum* isolates from several locations. The surface was then covered with a 2-cm layer of sawdust previously inoculated with *H. annosum*, and a 10-cm trunk section from the same tree was placed on the sawdust to reduce contamination by potentially antagonistic microorganisms.

Each year for 5 yr after inoculation, the residual trees in each plot were checked for mortality caused by *H. annosum*. Dead trees with basidiocarps of *H. annosum* were considered to have been killed by the fungus. Roots of dead trees without basidiocarps were examined for typical decay or were cultured to detect the fungus. If one or more basidiocarps were found at the base of a living tree, this tree was considered to have a lethal infection. Maximum extent of mortality was determined 5 yr after inoculation by measuring from the plot center (the middle of the four inoculated stumps) to the base of the most distant dead tree or the most distant living tree with basidiocarps at its base.

Soil sampling and analysis. Soil samples for chemical and mechanical analysis were collected from each plot at each site. The A horizon was sampled 0–15 cm deep with a soil sampling tube by making a composite of about 15 cores collected at random locations on each plot. The B horizon was sampled about 76 cm deep with a soil auger. A composite sample was made from five borings on each plot.

Soil chemicals were analyzed with the methods described by Jackson (7). Nitrogen was determined by macro-Kjeldahl. Exchangeable potassium, calcium, and magnesium were extracted with 1.0 N NH_4OAc and analyzed with a Perkin-Elmer 303 atomic absorption spectrophotometer. Phosphorus was extracted with 0.1 N HCl and 0.03 N NH_4F (Bray No. 2 solution) and determined by the molybdenum blue procedure. Soil pH was determined in a soil paste with a glass electrode. Organic matter was determined by the chromic acid (Walkley-Black) method. Mechanical analysis was done by the hydrometer method.

Statistical analysis. The hazard of pine mortality caused by *H. annosum* was arbitrarily scored as low or high before statistical analysis. The hazard class was defined as low if six or fewer trees within a plot were dead or had lethal infections 5 yr after inoculation and high if seven or more trees were dead or infected. Our rating was more conservative than that used by Froehlich et al (5), who considered high hazard sites as those with five or more diseased trees. By our

Table 1. Location, soil texture, drainage, and initial annosus root disease hazard rating of study sites in South Carolina, North Carolina, Georgia, and Virginia

Area	County, state	Soil texture	Drainage	Hazard rating[a]
Coastal plain	Colleton, SC[b]	Sandy to sandy loam	Moderate to good	High
	Decatur, GA	Sandy to sandy loam	Moderate to good	High
	Camden, GA	Sandy to sandy loam	Moderate to good	High
	Aiken, SC	Sandy to sandy loam	Moderate to good	High
	Georgetown, SC			
	Site 1	Loamy sand to sandy loam	Poor	Medium or low
	Site 2	Loamy sand to sandy loam	Poor	Medium or low
	Isle of Wight, VA			
	Site 1	Loamy sand to sandy loam	Poor	Medium or low
	Site 2	Sandy to sandy loam	Moderate to good	High
	Bertie, NC	Loamy sand to sandy loam	Poor	Medium or low
Piedmont	Warren, NC	Sandy loam to sandy clay loam	Moderate to good	Medium
	Union, SC			
	Site 1	Sandy loam to sandy clay loam	Moderate to good	Medium
	Site 2	Sandy clay loam	Good	Low
	Green, GA			
	Site 1	Sandy loam to sandy clay loam	Moderate to good	Medium
	Site 2	Sandy loam to sandy clay loam	Moderate to good	Medium
	Site 3	Sandy loam to sandy clay loam	Moderate to good	Medium
	Site 4	Sandy loam to sandy clay loam	Moderate to good	Medium

[a]According to Morris and Frasier (10).
[b]Silt content in A horizon averaged 25%, higher than in any other coastal plain plot.

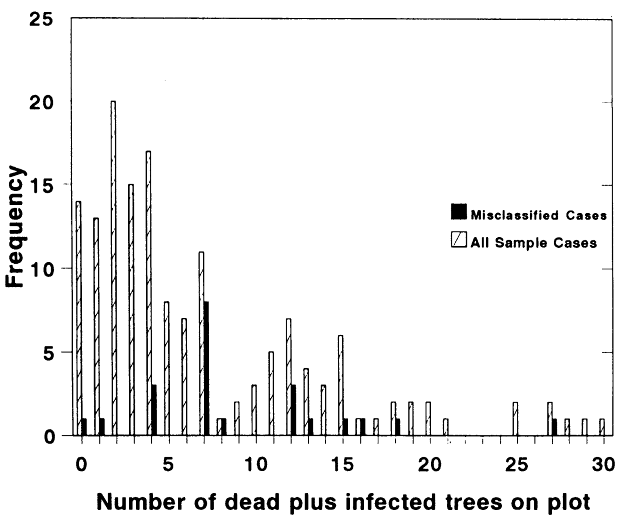

Fig. 1. Frequency of dead and infected trees in plots correctly and incorrectly classified for hazard of lethal infection of pines by *Heterobasidion annosum* according to the classification tree. Low hazard class = six or fewer infected trees per plot; high hazard class = seven or more infected trees per plot.

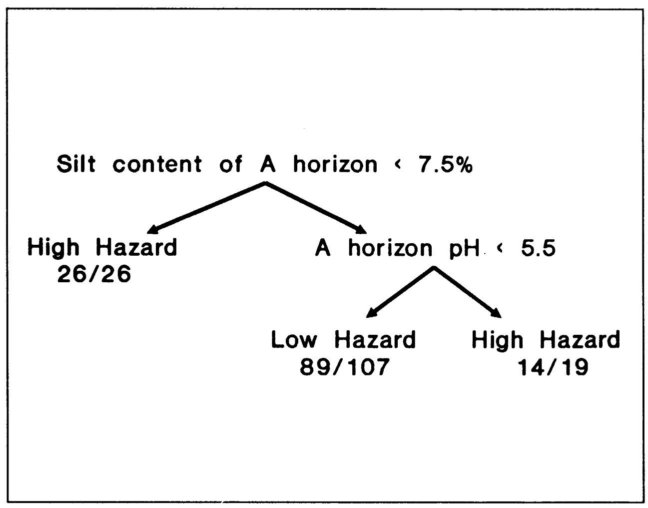

Fig. 2. Classification tree for predicting pine mortality caused by *Heterobasidion annosum* on 152 plots in the southeastern United States. Fractions are proportions of arbitrarily classified plots that were correctly identified by classification and regression tree analysis (CART).

criteria, 94 sample plots were on low-hazard sites and 58 were on high-hazard sites (Fig. 1).

Statistical procedures to predict site hazard class were then conducted. Twenty-two soil measurements were available as candidate predictor variables (Table 2). Stepwise linear discriminant analysis was initially used to predict class membership, but no strong linear relationships were found. Classification tree analysis was then used. This technique uses predictor variables to sequentially split the sample into groups with purer class membership (13).

Tenfold cross-validation was used to validate the classification tree. This process involves randomly excluding sample cases, developing the model, and testing the model with the excluded cases. Because the model is always tested with sample cases that were not used for model development, cross-validation provides a nearly unbiased estimate of how well the model would predict the result from a new sample of cases from the same population (12,14).

RESULTS AND DISCUSSION

Six months after inoculation, all inoculated stumps were colonized by *H. annosum* and most had basidiocarps. Thus, each plot center had adequate inoculum to initiate spread of the fungus. After 5 yr, *H. annosum* had killed trees at every location (Table 3). All dead and diseased trees were grouped around the inoculated stumps and apparently represented spread of *H. annosum* from these stumps.

A two-variable classification tree correctly predicted hazard class for 85% of the sample cases (Fig. 1). The model indicates that soils with A horizons low in silt content or high in pH are usually high-hazard sites (Fig. 2). This relationship is consistent with results of previous studies (5,9,10), but it does not hold true on sites with high water tables. Hazard on such sites is usually low regardless of the silt content. Most of the misclassified sites, with the exceptions described below, had seven or eight dead and diseased trees.

Eight misclassified plots each had more than 10 infected trees, and all of these plots were on one site in Colleton County, South Carolina. The soil on this site is a loamy fine sand over a fine sandy loam to a depth of 64 cm, has a silt content averaging 25%, and is moderately well drained. This site would be recognized as a high-hazard site by Morris and Frasier (10) by virtue of its soil texture and depth. We did not analyze soil depth in our study. Although the classification tree correctly classified 85% of the samples, the inclusion of other variables such as soil depth could possibly improve site classification.

With the use of CART, a soil sample can be analyzed and the hazard of pine mortality caused by *H. annosum* can be assigned. CART provided an easily interpreted hazard rating system that correctly identified the hazard status of 129 of 152 plots. We have not evaluated this classification tree for use in areas other

Table 2. Chemical and physical characteristics of soils as related to hazard of pine mortality caused by *Heterobasidion annosum*[a]

Measurement	Low-hazard plots Mean	Low-hazard plots SD	High-hazard plots Mean	High-hazard plots SD
A horizon				
pH	5.0	0.36	5.2	0.41
Sand (%)	69	11	77	12
Silt (%)	16	5	13	8
Clay (%)	15	9	10	5
Carbon (%)	0.67	0.28	0.64	0.28
Organic matter (%)	1.19	0.62	1.13	0.57
N (μg/g)	354	164	290	113
P (μg/g)	32	27	34	25
K (μg/g)	36	18	22	15
Ca (μg/g)	154	95	109	82
Mg (μg/g)	32	27	18	9
B horizon				
pH	5.0	0.35	5.0	0.33
Sand (%)	47	12	63	20
Silt (%)	15	5	10	7
Clay (%)	38	11	27	15
Carbon (%)	0.29	0.13	0.22	0.10
Organic matter (%)	0.50	0.22	0.39	0.18
N (μg/g)	260	77	188	54
P (μg/g)	8	7	14	12
K (μg/g)	50	28	32	23
Ca (μg/g)	254	98	194	145
Mg (μg/g)	77	53	50	40

[a] Hazard was arbitrarily scored as high if seven or more trees were dead or had lethal infections and low if six or fewer trees were affected.

Table 3. Pine mortality and extent of spread of *Heterobasidion annosum* 5 yr after inoculation of stumps[a]

Location	Number of trees Dead	Number of trees Infected	Maximum spread[b] (m)	Average spread across row (m)	Average spread along row (m)
Camden County, GA	10.7	13.9	9.1	7.1	8.2
Colleton County, SC	9.1	9.2	7.4	5.6	6.5
Green County, GA, site 4	9.8	6.3	6.6	5.4	5.6
Aiken County, SC	7.1	6.0	6.6	5.4	5.6
Decatur County, GA	2.8	7.2	7.4	4.9	6.5
Union County, SC, site 1	4.4	3.5	5.1	4.1	4.4
Isle of Wight County, VA, site 2	6.6	1.1	5.3	...[c]	...
Georgetown County, SC, site 1	2.3	2.7	5.7	2.8	4.1
Green County, GA, site 3	2.1	2.8	4.1	2.9	3.1
Green County, GA, site 2	2.7	1.8	4.4	3.1	3.4
Isle of Wight County, VA, site 1	3.4	1.0	4.1	2.8	3.5
Georgetown County, SC, site 2	2.2	2.2	5.2	3.9	4.1
Warren County, NC	3.2	1.1	4.8	3.4	3.5
Greene County, GA, site 1	2.1	1.2	3.5	2.5	3.1
Bertie County, NC	1.9	1.4	4.7	3.1	3.9
Union County, SC, site 2	0.6	0.9	2.7	2.0	1.6

[a] Data are averages for 10 plots at 14 locations, 7 plots in Bertie County, and 5 plots in site 2 (a natural stand) of Isle of Wight County. Sites are listed according to number of dead plus infected trees. Trees with basidiocarps of *H. annosum* at the base were considered infected.
[b] Distance from inoculated stumps to farthest dead tree or farthest living tree with basidiocarps of *H. annosum*.
[c] Not applicable, since site was in a natural stand.

than where our data were collected. On the basis of the 10-fold cross-validation, however, the expected accuracy of the classification tree applied to similar populations is 80%. Because the sites examined in this study are typical of a wide range of annosus root disease sites in the southeastern United States, the classification tree should be quite stable. Pathologists and forest managers should find CART useful for improving the accuracy of disease hazard classification.

ACKNOWLEDGMENTS

We thank D. W. Roberts for his suggestions and assistance with the analysis and we also thank the reviewers and editor for their improvements to this manuscript.

LITERATURE CITED

1. Alexander, S. A. 1989. Annosus root disease hazard rating, detection and management strategies in the southeastern United States. Pages 111-116 in: Proc. Symp. Res. Manage. Annosus Root Dis. (*Heterobasidion annosum*) west. North Am. U.S. For. Serv. Gen. Tech. Rep. PSW-116.
2. Alexander, S. A., Skelly, J. M., and Morris, C. L. 1975. Edaphic factors associated with the incidence and severity of disease caused by *Fomes annosus* in loblolly pine plantations in Virginia. Phytopathology 65:585-591.
3. Breiman, L., Freidman, J. H., Olshen, R. A., and Stone, C. J. 1984. Classification and Regression Trees. Wadworth, Inc., Belmont, CA.
4. Byler, J. W., Marsden, M. A., and Hagle, S. K. 1990. The probability of root disease on the Lolo National Forest, Montana. Can. J. For. Res. 20:987-994.
5. Froehlich, R. C., Dell, T. R., and Walkinshaw, C. H. 1966. Soil factors associated with *Fomes annosus* in the Gulf States. For. Sci. 12:356-361.
6. Froelich, R. C., Kuhlman, E. G., Hodges, C. S., Wiess, M. J., and Nichols, J. D. 1977. *Fomes annosus* root rot in the South: Guidelines for prevention. U.S. For. Serv. South. Area State Private For.
7. Jackson, M. L. 1958. Soil Chemical Analysis. Prentice-Hall, Englewood Cliffs, NJ.
8. Kuhlman, E. G. 1986. Impact of annosus root rot minimal 22 years after planting pines on root rot infested sites. South. J. Appl. For. 10:96-98.
9. Kuhlman, E. G., Hodges, C. S., Jr., and Froelich, R. C. 1976. Minimizing losses to *Fomes annosus* in the southern United States. U.S. For. Serv. Res. Pap. SE-151.
10. Morris, C. L., and Frasier, D. H. 1966. Development of a hazard rating for *Fomes annosus* in Virginia. Plant Dis. Rep. 50:510-511.
11. Robbins, K. 1984. Annosus root rot in eastern conifers. U.S. For. Serv. For. Insect Dis. Leafl. 76.
12. Verbyla, D. L. 1986. Potential prediction bias in regression and discriminant analysis. Can. J. For. Res. 16:1255-1257.
13. Verbyla, D. L. 1987. Classification trees: A new discrimination tool. Can. J. For. Res. 17:1150-1152.
14. Verbyla, D. L., and Litvaitis, J. A. 1989. Resampling methods for evaluating class accuracy of wildlife habitat models. Environ. Manage. 13:783-787.