

Predicting Swiss Needle Cast Disease Distribution and Severity in Young Douglas-Fir Plantations in Coastal Oregon

Pablo H. Rosso and Everett M. Hansen

First author: CSTARs Laboratory, Department of Land, Air, and Water Resources, University of California at Davis, One Shield Ave, Davis 96515; and second author: Department of Botany and Plant Pathology, Oregon State University, Cordley Hall 2082, Corvallis 97331.
Accepted for publication 24 January 2003.

ABSTRACT

Rosso, P. H., and Hansen, E. M. 2003. Predicting Swiss needle cast disease distribution and severity in young Douglas-fir plantations in coastal Oregon. *Phytopathology* 93:790-798.

Swiss needle cast (SNC), caused by the fungus *Phaeocryptopus gaeumannii*, is producing extensive defoliation and growth reduction in Douglas-fir forest plantations along the Pacific Northwest coast. An SNC disease prediction model for the coastal area of Oregon was built by establishing the relationship between the distribution of disease and the environment. A ground-based disease survey (220 plots) was used to study this relationship. Two types of regression approaches, multiple linear regression and regression tree, were used to study the relationship between disease severity and climate, topography, soil, and forest stand characteristics. Fog occurrence, precipitation, temperature, elevation, and

slope aspect were the variables that contributed to explain most of the variability in disease severity, as indicated by both the multiple regression ($r^2 = 0.57$) and regression tree (RMD = 0.27) analyses. The resulting regression model was used to construct a disease prediction map. Findings agree with and formalize our previous understanding of the ecology of SNC: warmer and wetter conditions, provided that summer temperatures are relatively low, appear to increase disease severity. Both regression approaches have characteristics that can be useful in helping to improve our understanding of the ecology of SNC. The prediction model is able to produce a continuous prediction surface, suitable for hypothesis testing and assisting in disease management and research.

Additional keywords: *Pseudotsuga menziesii*, Douglas-fir diseases, forest disease model, *Venturiaceae*.

Swiss needle cast (SNC), a foliar disease of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) caused by the ascomycetous fungus *Phaeocryptopus gaeumannii* (T. Rhode) Petr., was first discovered in Switzerland in the mid-1920s in young plantations established from seedlings brought from the United States. SNC also has been observed in other places where Douglas-fir was introduced (11), such as in New Zealand (12); therefore, the disease was considered a consequence of the host growing in nonoptimal conditions. In contrast, although the fungus is widespread and apparently native to Oregon, it has not caused extensive damage to forest trees until recently (1,10). In the late 1980s, however, chlorosis, premature defoliation, and growth loss associated with SNC were evident in many young, previously vigorous forest plantations on the north Oregon coast, and the disease has since intensified and spread (10).

The causes of this disease outbreak may be determined only if the conditions at which it occurs are well understood. Different environmental conditions, including climate, topography, soil characteristics, and forest stand characteristics (such as fire history, stand composition and structure, and so on), have been associated with the disease distribution in Oregon. For example, disease severity seems to be higher in low-elevation areas, southern slopes, sites previously occupied by spruce or alder, or sites with high rainfall or closer to the coast (10,13). Despite these observations, a systematic and detailed study of the variables that are associated with the disease distribution still needs to be done.

Disease surveys and observations indicate that SNC affects whole stands and large areas, suggesting that disease expression depends on phenomena operating at the landscape or regional

scale. SNC also has a strong seasonal pattern of development. In the Pacific Northwest, infection occurs in May or June, shortly after bud break. Germ tubes penetrate young needles through stomata, colonizing them intercellularly. They also colonize the needle surface (3). Pseudothecia form and mature in the stomatal chambers from November to February and, by May, ascospores are released and the infection cycle starts again. Needle colonization and stomatal occlusion by pseudothecia gradually impair the normal functioning of the needle (13,14), leading to premature needle shed when about 50% of stomata are occupied (10,13,14). Needle discoloration and loss may be evident in plantations as early as January, depending on the severity of the infection and climatic conditions.

To better understand disease outbreaks, it is important to determine under what conditions it is most prevalent. The natural spatial and temporal variations of the disease can be systematically studied to investigate the extent and nature of the environmental influence on disease severity.

The objective of this study was to determine what factors are associated with the geographical distribution and severity of SNC in Oregon and to build a statistical model to predict its occurrence. To achieve this goal, two main statistical tools, multiple linear regression and regression tree analysis, were used and their outcomes compared.

MATERIALS AND METHODS

To produce a disease prediction model, regression analyses were done using ground-based disease information as response (dependent) variables. This disease information was obtained from survey plots established in diseased and healthy Douglas-fir plantations in western Oregon. Explanatory (independent) variables were obtained by one of three methods: (i) direct measurement of survey plot characteristics (terrain slope, aspect, and so

Corresponding author: P. H. Rosso; E-mail address: phrosso@ucdavis.edu

Publication no. P-2003-0513-01R
© 2003 The American Phytopathological Society

on); (ii) overlay of digital disease survey plot maps on digital environmental information (temperature, precipitation, elevation, and so on), through the use of a geographic information system (GIS); and (iii) documented information (plantation seed source, stand site index, and so on).

Research area. The area of study, the northern coastal region of Oregon, extends from the city of Astoria (46°10' latitude N) south to Coos Bay (43°25' latitude N); and from the coast inland about 50 km to a north-south line at approximately 123°10' longitude, corresponding to the western edge of the Willamette Valley (Fig. 1) ($\approx 11,500$ km²). The Oregon coast topography is determined by the Coast Range, a moderately high range with most ridge tops within 450 to 750 m of altitude. Steep slopes and soils with relatively little development characterize the area. A wet and mild climate allows for the growth of dense and productive conifer forests where Douglas-fir has been extensively harvested and replanted by the timber industry.

Disease survey. A 1998 SNC survey conducted by the Oregon Department of Forestry provided the disease severity information and most other stand-based data. A total of 220 10- to 30-year-old plantations in the Oregon coastal area were selected according to accessibility and with no other evident bias (Fig. 1). Each stand was measured for needle retention, and a composite SNC severity rating was assigned.

To determine stand needle retention, a transect was laid through each stand starting from a haphazard point. Five points were established along the transect at 15-m intervals. The nearest two dominant or codominant trees were selected for sampling. The percentage of needles present on secondary lateral branches of the upper one-third of the crown of each tree was assessed. A value of 1 was assigned for each year's internode which had 100% of the needles present, as determined by visual assessment, and a fraction of 1, if less. For example, if a branch appeared to have the full complement of needles on the last 4 years' internodes and only half of the year before that, it would receive a rating of 4.5.

The SNC severity rating is a composite of the stand's average tree foliar retention, height growth, and foliage color. The lower value 1 represents a healthy stand, with 4 or more years of foliage, annual height growth of ≈ 1 m and normal green color. Values 2 to 5 were assigned to stand trees showing 4- to 1-year-old needle loss, respectively, and stand chlorosis ranging from slight to evident, respectively. In value 4 there is a slight height growth reduction and in value 5 there is at least 25% reduction in most trees. Value 6 is assigned to a severely diseased stand with substantial 1-year-old needle loss, 60% height growth reduction, and extremely chlorotic foliage. Both the stand SNC severity rating and crown needle retention were used alternatively as response variables in the models.

The location of the central geographic point of each stand was determined with a global position system (GPS) and digitized into a point coverage which was displayed on a Landsat image to assess the accuracy of digital stand location by comparison to detailed stand maps and aerial photos.

Independent variables. Temperature and precipitation data were obtained from raster (grid)-based outputs from the climate prediction model PRISM (4). These grids have an approximate cell resolution of 4 by 4 km (2.5 arc-minutes). Temperature and precipitation grids of the months considered most relevant from the epidemiological point of view, January, April to July, and November, were selected from the five most recent available years (1989 to 1993). These are the months that seem to correspond with critical steps in the life cycle of *P. gaeumannii* (3). Each of the 6 months was averaged over the 5 years. Ambient vapor pressure deficit (VPD), a measure of the water content of the atmosphere, was derived from temperature rasters following Monteith and Unsworth (15).

Degree-days (base 0°C) were calculated using an interpolation model that uses PRISM climate grids as inputs (Oregon degree-

day mapping calculator, version 1.2., developed by L. Coop, is available online on the Oregon State University Integrated Plant Protection Center website). Total radiation was calculated using SolarImg, a solar input model based on geographic location and digital topographic information. SolarImg is a component of the Solarrad model developed by M. Harmon and B. Marks, which is available online at the H. J. Andrews Experimental Forest, LTER, Forestry Science Lab, Oregon State University website.

Two types of fog data were used in the analysis: one obtained from satellite imagery and the other from a fog distribution model. Fog or low cloud occurrence was mapped using daily images obtained from the GOES satellite (National Oceanic and Atmospheric Administration, NOAA). This satellite produces an 8-km resolution raster image in which low clouds and high clouds are represented by different pixel values. Images from 1999 were downloaded from the Internet (website of the National Environmental Satellite Data and Information Service, NESDIS, NOAA) and processed to obtain a cumulative fog occurrence value for the spring-summer and fall-winter periods. The modeled fog map (C. Daly, unpublished data) is based on a coastal advection model which simulates the wet air penetration from the Pacific Ocean inland in the Oregon coast region. This simulation was adjusted to local variability with satellite and weather station information. The model produces a map with the predicted daily percentage of hours with fog over the entire month of July on the Oregon coast at a resolution of 1 km.

Two types of topographic variables were produced, one from measurements in situ and the other obtained from digital elevation models (DEMs). The in situ measurements of slope aspect and inclination were made on the ground using a handheld compass and clinometer. Measurements were approximated to represent the entire stand or plantation. Stand slope position also was assigned on the ground, describing the location of the stand with respect to the bottom of the slope, with four possible values: low, medium, high, or ridge top.

Stand characteristics, age, site index, tree density, and seed source were obtained from stand records. Elevation and shortest distance to the coast were extracted from maps.

A 60-m resolution DEM was resampled to 90- and 140-m resolution DEMs. Slope aspect and inclination were calculated in ERDAS-Imagine (version 8.3; Erdas Inc., Atlanta, GA) for each of the three DEMs. After exhaustive analysis of alternative slope representations, aspect was converted from azimuth angles to eight cardinal directions (N, NE, E, SE, S, SW, W, NW) and finally made into a binary variable assigning aspects: SE, S, SW, and W = 0, and NW, N, NE, and E = 1.

The 1936 stand vegetation variables were extracted from a polygon coverage produced by the Forest Service based on a map by Andrews and Cowlins (20). The coverage was transformed to raster, and a 1936 vegetation category was assigned to each stand. Stands were classified according to whether or not they had Douglas-fir or spruce hemlock as the dominant species in 1936.

Soil variables were obtained from the State Soil Geographic Data Base (STATSGO; U.S. Department of Agriculture). Stands were matched geographically with the corresponding map units, and the dominant soil types ("sequum" number) were selected from each map unit. The following soil characteristics were obtained: upper layer depth, calcium carbonate contents (percent by weight), cation exchange capacity, organic matter (percent by weight), particle size (sandy, loamy, silty, and so on), and soil order. Calcium carbonate and cation exchange capacity were later discarded due to lack of variability and missing information.

Analyses. Variables in raster format were overlaid on the SNC survey point coverage and a value of each variable was assigned to each survey plot (using the remote sensing software ERDAS-Imagine, version 8.3; Erdas Inc., Atlanta, GA).

All variables were plotted and correlated with the response variables to investigate the data structure. No single variable was

highly correlated with either disease rating or needle retention; therefore, a multivariate approach was considered.

The total number of variables evaluated, 68 (Table 1), was reduced to a manageable number (Table 1, underlined) using the

following procedure. Climate variables from different months within a group (i.e., maximum temperature in January, April, May, June, July, and November) were correlated by pairs and only one variable of each highly correlated subgroup was taken for further

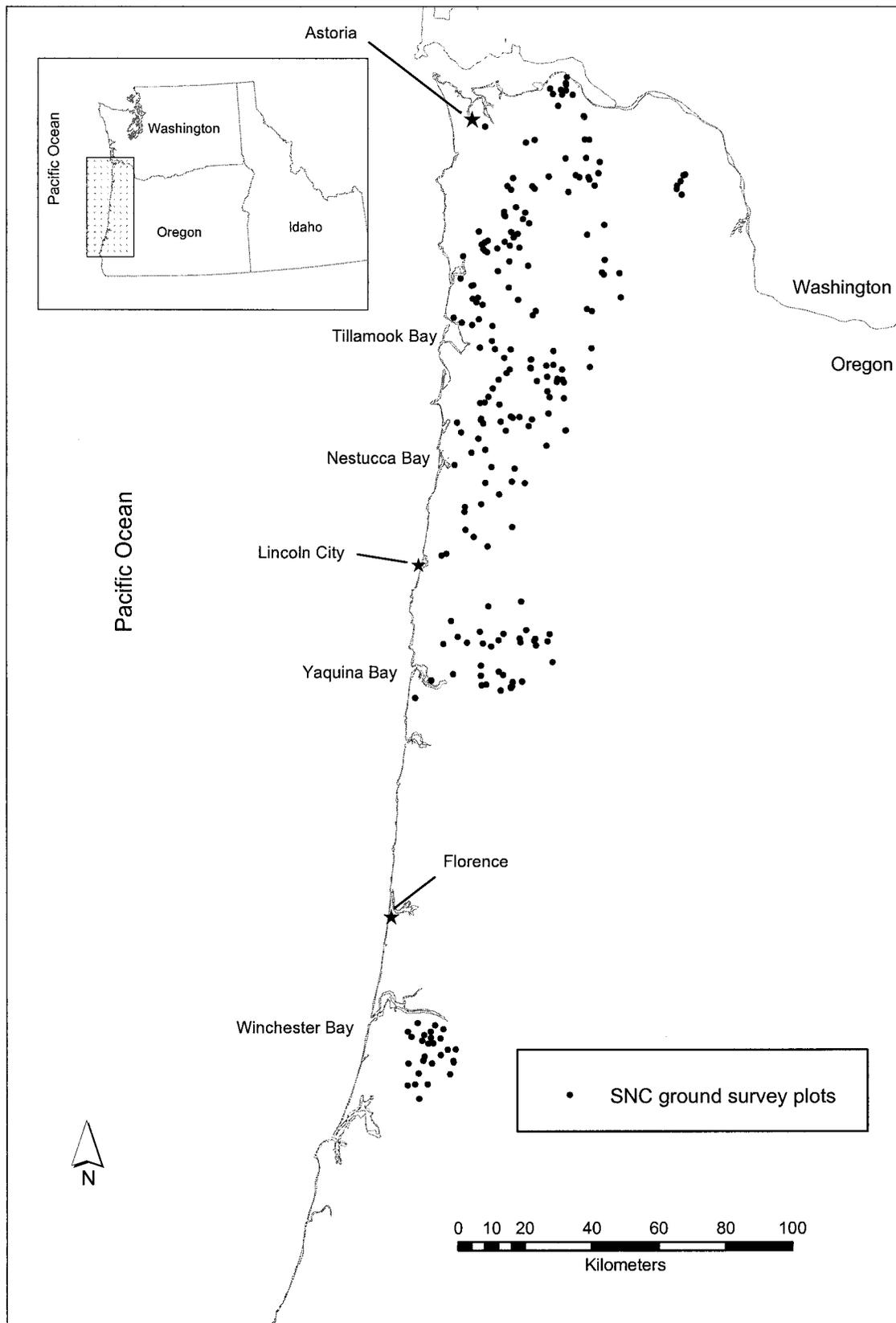


Fig. 1. Location of the Swiss needle cast (SNC) survey plots in northwest Oregon ($n = 220$). Main cities in the area (stars) have been included for geographic reference.

analysis. Other variables, such as site index, were excluded a priori because of excessive numbers of missing values. Categorical variables that could not be reduced to binary (e.g., soil order) also were excluded from this first part of the analysis because creating indicator variables for each category would increase the number of explanatory variables in an inefficient way. The addition of these variables would complicate the analysis and interpretation of results, besides utilizing degrees of freedom necessary to test interaction terms later in the analysis. In these cases, disease rating and needle retention were averaged by categories of each variable to verify the presence of trends in symptoms among categories. The explanatory value of these variables excluded a priori was assessed once a satisfactory model was built.

After this first selection, the number of explanatory variables was reduced to about a third of the initial number (Table 1, underlined variables). The resulting variables and the response variables were combined to build a model using two alternative analyses: multiple regression and a regression tree analysis (the latter often referred to as “CART analysis”) (2). Fifty stand plots (about one-quarter of the complete data set) were chosen randomly and excluded from the regression analysis to be used as a validation set for both the linear regression and regression tree models.

Multiple linear regression analysis. Variables were screened for fulfillment of the regression assumptions (18). Correlation coefficients were calculated to check for high correlation between any two variables. None of the selected variables had a correlation

coefficient higher than 0.8. A principal component analysis (PCA) was done to check for multivariate correlation (19) before the analysis was carried out, but none was observed. After the regression analysis, variance inflation was used to detect the presence of multicollinearity, but in no case did this exceed 3.5, indicating no substantial collinearity (17). In all cases, variance inflation Normality assumption of each variable also was tested using box plots and histograms. Normality of all the variables together was tested using a PCA analysis and analyzing histograms of the PCs (19). Variables (including the response variable) showing departures from normality were transformed as needed. July precipitation, 140-m resolution elevation, 60-m resolution elevation, elevation in situ, and organic matter were transformed to logarithmic scale. Exploratory regression analyses were carried out and scatterplots of the residuals were constructed to detect possible nonlinearity and nonconstant variance. All these assumptions were met by the variables selected.

For the selection of the best regression models, two variable selection techniques called “Cp statistic” and Schwartz’s Bayesian information criterion (BIC) were used (18). These techniques assign the lowest value to those regression models that best explain the response variable with the least number of independent variables. Models with lower Cp and BIC statistic values were selected for further analysis.

All the variables included in the selected models were pooled together and included in one regression model. All possible two-

TABLE 1. Variables considered in the analysis to build the Swiss needle cast (SNC) prediction model in coastal Oregon

Variables ^a	Layer type, resolution
Response	
Stand SNC severity rating (five categories, 1 = low to 5 = high severity)	Point coverage
Upper one-third crown needle retention (continuous, 0.8–3.5)	Point coverage
Independent	
<u>Universal Transverse Mercator X and Y coordinates</u>	
Climate directly from PRISM model	
Mean monthly maximum temperature (months: J,A,M,J,J, & N) (°C)	2.5 arc-minute (≈4 by 4 km) raster
Mean monthly minimum temperature (months: J,A,M,J,J, & N) (°C)	2.5 arc-minute (≈4 by 4 km) raster
Mean monthly precipitation (months: J,A,M,J,J, & N) (mm)	2.5 arc-minute (≈4 by 4 km) raster
Climate derived from PRISM outputs	
Mean vapor pressure deficit (VPD) (months: J,A,M,J,J, & N) (Pa)	2.5 arc-minute (≈4 by 4 km) raster
Degree days, base 0°C (months: J,A,M,J,J, & N) (°C)	2.5 arc-minute (≈4 by 4 km) raster
Total radiation (months: J,A,M,J,J, & N) (SolarImg model)	2.5 arc-minute (≈4 by 4 km) raster
<u>July fog occurrence (Fog model)</u>	1 by 1 km raster
Climate from GOES satellite	
<u>1999 spring–summer fog or low cloud occurrence</u>	8 by 8 km raster
1999 fall–winter fog or low cloud occurrence	8 by 8 km raster
Topography from digital elevation models	
<u>Elevation (m) (three different resolutions)</u>	60-, 90-, and 140-m-pixel size raster
<u>Slope inclination (%) (three different resolutions)</u>	60-, 90-, and 140-m-pixel size raster
<u>Slope aspect (SE–W versus NW–E) (three different resolutions)</u>	60-, 90-, and 140-m-pixel size raster
Heat load index (0–1) (two different resolutions)	60- and 140-m-pixel size raster
Topography measured in situ	
<u>Elevation (m)</u>	...
<u>Slope inclination (%)</u>	...
<u>Slope aspect (binary: SE–W versus NW–E)</u>	...
<u>Slope position (binary: low–mid versus high–ridge top)</u>	...
Stand characteristics	
Age (years)	...
Site index	...
Stem density (trees per hectare)	...
Seed source (binary: Local versus nonlocal)	...
Distance to the coast (miles)	...
Historical vegetation	
<u>Areas dominated by Douglas fir in 1936 (binary: yes or no)</u>	Polygon coverage
<u>Areas dominated by Spruce-Hemlock in 1936 (binary: yes or no)</u>	Polygon coverage
Soil characteristics from STATSGO database	
<u>Organic matter (percent weight)</u>	Polygon coverage
Upper layer depth (cm)	Polygon coverage
Soil order (4 categories)	Polygon coverage
Particle size (6 categories)	Polygon coverage
Geology or lithology (15 categories)	Polygon coverage

^a The independent variables selected for the final regression analyses are underlined. Months J, A, M, J, J, & N = January, April, May, June, July, and November, respectively.

way interaction and quadratic terms were added to this model. Then, the variable with the highest *P* value was removed. This step was repeated until all variables in the model showed coefficients significantly different from zero (0.05 significance level).

Visual inspection of predicted versus residual graphs, the “Dffits” function, and studentized residuals (18) were used to identify outliers and influential observations. No observation was found to be influential enough to be removed from the data set.

A final model refinement was attempted by plotting the residuals against those variables excluded a priori. All linear regression analyses were performed in SAS (version 8.0; SAS Institute, Cary, NC).

Regression tree analysis. Regression tree analysis (RT) classifies sampling units based on a sequential binary partitioning of the data into homogeneous subsets, considering one explanatory variable at a time. This method has some advantages over traditional regression in that it is not based on linear and additive relationships between factors, and variables do not need to fulfill distributional assumptions (5,16,21). For this reason, explanatory

variables were used in their original format. RT also is robust with respect to outliers and missing values, and any type of variable, numerical or categorical, can be used without transformation. To make both regression results comparable, variables transformed into binary were kept as such also for RT. The statistical package S-Plus (version 4; Insightful Corporation, Seattle, WA) was used to perform the RTs.

Validation. A statistical test by Freese (6) was used on the 50 plots previously set aside to judge the predictive capability of the model and to compare models. Freese’s χ^2 statistic helps to determine whether a model achieves a certain degree of accuracy when a validating data set is used. Accuracy is chosen a priori and depends on the maximum tolerated departure from the observed values, and the desired probability level. In this case, a maximum departure of 0.5 SNC rating units and a 0.05 probability level were chosen.

Spatial disease distribution analysis. The assumption of independence of observations in linear regression (18) might be violated when samples come from a geographical setting, because land features tend to be autocorrelated in space (7,16). To test for the presence of autocorrelation, two indices, Moran’s *I* and Geary’s *C*, were calculated. Both tests calculate the difference in values of the variable of interest (in our case, disease prediction residuals) between any two points, weighted by the distance between the points to emphasize the differences between nearby points. Moran’s *I* ranges from -1 to 1 , indicating no autocorrelation when it equals 0. In Geary’s index, a value of 1 means no autocorrelation, <1 means positive autocorrelation, and >1 , a negative one (7,8). Calculations were done using S-Plus (Insightful Corporation).

Prediction map. The best regression model was applied to geographically referenced information in the form of grids, and a map of the predictions was produced to better understand the spatial distributional characteristics of the predictions, and to allow for future accuracy assessment of spatial prediction.

RESULTS

Disease survey. The resulting SNC severity rating ranged from 1 to 6, with an average of 2.59. Only two stands were classified as 6; therefore, their rating was changed to 5 and category 6 was eliminated from the analysis. Needle retention ranged from 0.80 to 3.50 with an average of 1.87.

Analyses. Disease severity distribution was best explained by climate and topographic variables that typically have an important influence on a stand’s temperature and moisture status. Climate variables from summer months appeared more important than from any other seasons.

Multiple linear regression analysis. The best regression model found was: Mean {SNC rating} = $3.30 - 0.13$ mean July maximum temperature + 0.58 mean July precipitation (log scale) + 0.018 July fog occurrence + 0.23 SE–W slope aspect at 60-m pixel size + 0.22 SE–W slope aspect in situ – 0.24 elevation in situ (log scale).

Standard errors of the coefficients are (starting with the intercept): 1.38, 0.05, 0.17, 0.004, 0.1, 0.1, and 0.09. These variables explain $\approx 60\%$ ($r^2 = 0.57$, $n = 170$) of the observed variation in SNC rating. All coefficients of the included variables were significantly different from zero at a 0.05 level. The relationship between predicted and observed disease levels of the stand plots as applied to data used to develop the regression model is shown in Figure 2A.

Residuals of this model were plotted against the quantitative variables excluded a priori: stand age, average stem density, and site index. None of these variables showed an obvious trend with respect to the residuals. Residuals also were averaged by categories of those soil categorical variables excluded a priori: upper soil depth, soil order, particle size, and geology or lithology, but no

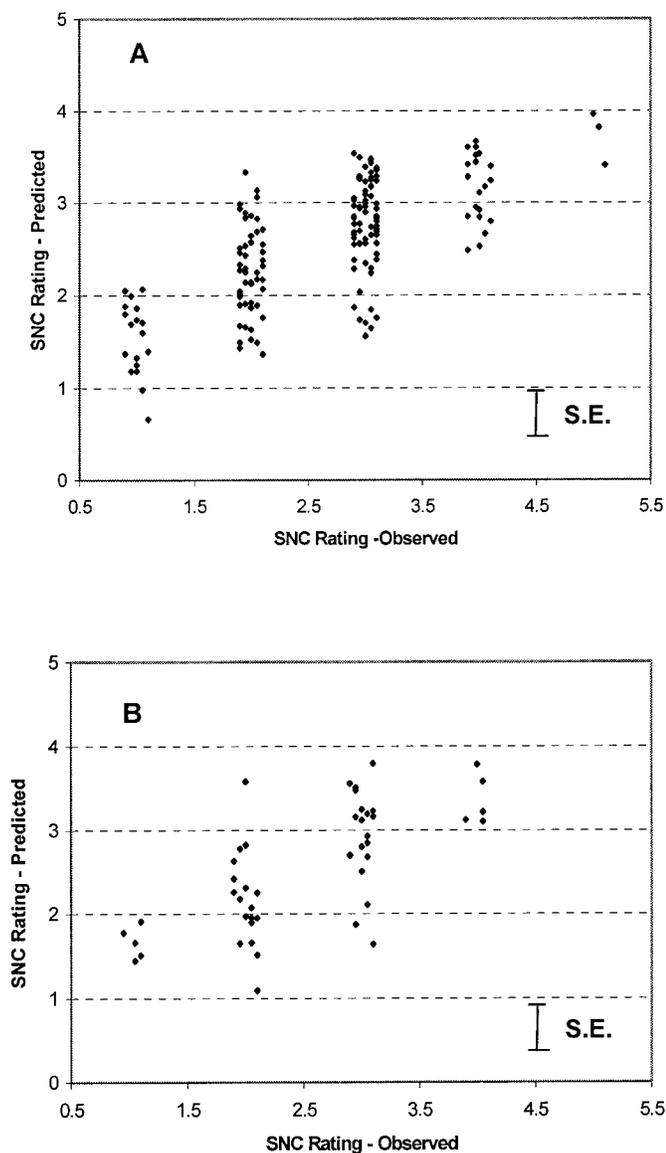


Fig. 2. Comparison of predicted (*Y* axis) and observed (*X* axis) Swiss needle cast (SNC) ratings for the **A**, model building plots ($n = 170$), and **B**, validation plots ($n = 50$), based on the multiple regression model described in the text. Observed values have been horizontally displaced to facilitate visualization. S.E. = standard error bar.

obvious trends were observed. Seed source was finally discarded because, when the information was reliable, most of it corresponded to local sources.

When the regression analysis was run with stand needle retention as the response variable, the maximum variability explained was too low ($r^2 = 0.29$) to justify further consideration.

RT. Initially, a regression tree of 25 nodes (that is, points at which splitting occurs) was obtained. Accordingly, it had 25 terminal nodes (leaves), in which to place all the survey plots to obtain a disease prediction for each plot. Residual mean deviance (RMD), a measure of the average heterogeneity within prediction groups, equaled 0.19. The original tree was pruned to avoid overfitting. The S-plus software uses a cost-complexity function to build a curve of deviance respect to the size of the final tree. This curve can be used to assess the minimum tree size (simplest model) at which deviance (and thus, misclassification) does not substantially increase. A limit of 40% increase in RMD was set with this approach. The final size of the pruned tree was 10 nodes

(Fig. 3) with an RMD of 0.27. Climate variables found to be relevant for the model were, in decreasing order of importance, July fog occurrence, November average minimum temperature, July VPD, and April degree days. Topography variables included: elevation at 90-m pixel size, slope aspect at 60 m, slope inclination at 140 m, and elevation in situ. Geographic location along the Y axis (north-south) also was among the most important variables. The relationship between predicted and observed SNC values for the plots used in developing the model is shown in Figure 4A.

Validation. When the linear regression model equation was applied to the validation set of stand plots, the mean of the residuals of the validation set was -0.01 , indicating a bias satisfactorily close to zero. The variance of the residuals was lower for the validation set (0.30 versus 0.37) indicating a comparable precision of the validation set compared with the model set (Fig. 2A and B).

For the Freese's χ^2 test, a maximum tolerated departure of 0.5 units of SNC disease rating was initially chosen and a prob-

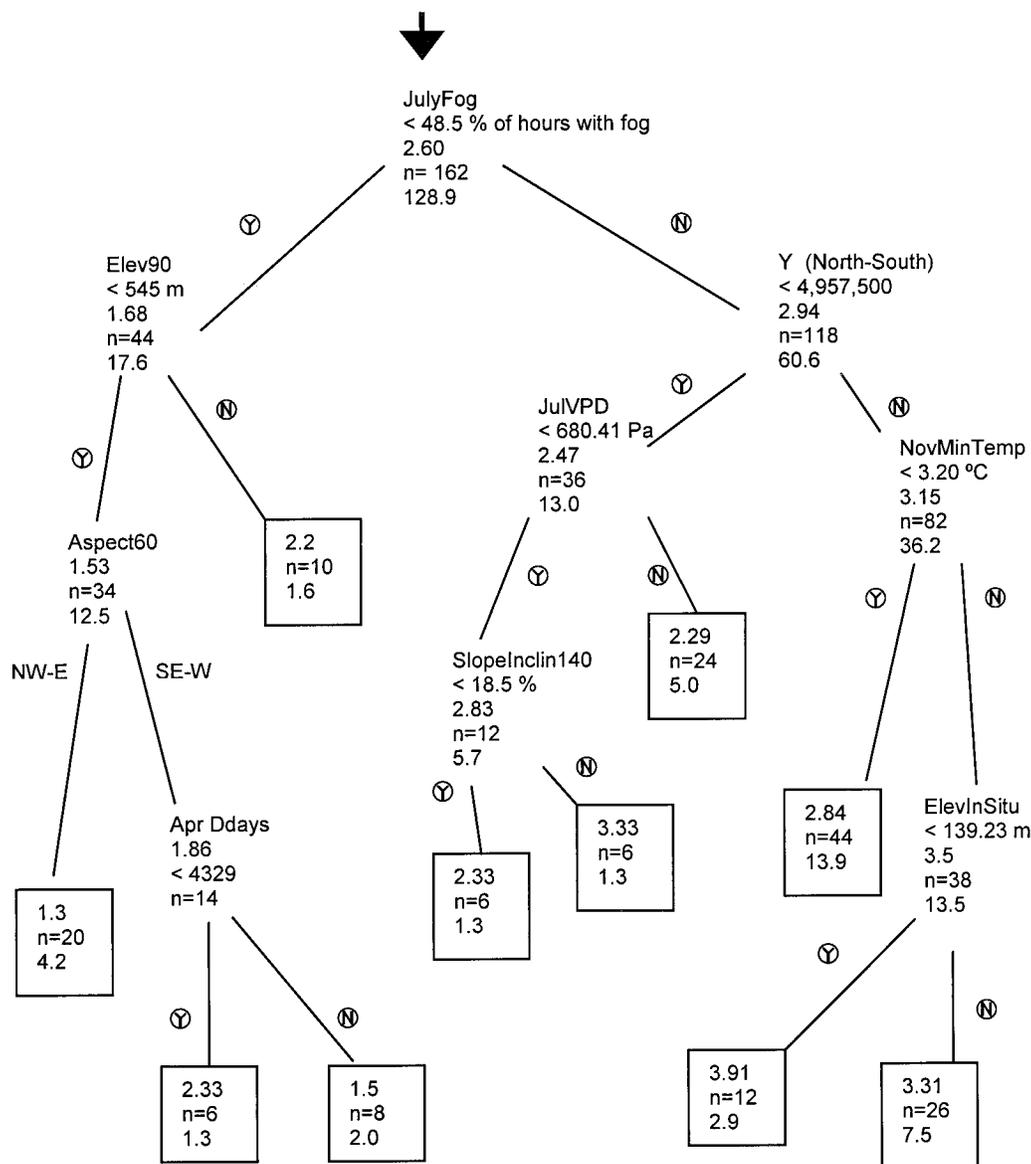


Fig. 3. Classification of Swiss needle cast (SNC) plots using the regression tree analysis. Binary splits are indicated by open text and terminal points (leaves) by boxes. First line of each binary split is the name of the variable used for the split; second line, the value at which split was done; third line, the SNC average at that node; fourth line, the sample size; and fifth line, the deviance at that point. Y and N indicate whether or not the value of a given stand is less than the split value of the node. First line of each leaf represents the predicted value, which is the average of the SNC rating values of the plots included in that leaf; second line represents the sample size; and third line represents the deviance. The total sample size ($n = 162$) resulted from the subtraction of the validation set and the exclusion of some plots with missing values in the explanatory variables.

ability level of 0.05. The calculated χ^2 statistic was higher (221.13) than the critical χ^2 statistic from tables (65.17), indicating an unacceptable level of accuracy. However, when the tolerated departure was raised to 1 unit, the calculated χ^2 statistic was lower (55.28), providing statistical evidence that 1 unit is about the highest accuracy the model could attain with that specific validation set.

When the validation set was applied to the pruned regression tree model, it showed a mean of the residuals close to zero (0.03) indicating virtual absence of bias. Its precision was comparable to that of the original model (variance of residuals 0.32 and 0.37, respectively) (Fig. 4A and B).

As with the linear regression model, a departure of 1 and a probability level of 0.05 were chosen to apply the Freese's χ^2 test to the regression tree model. The calculated χ^2 statistic was somewhat lower (62.29) than the critical χ^2 statistic from tables (65.17), indicating that, similarly to the linear regression model, 1 unit is about the highest accuracy the model could achieve.

Spatial disease distribution analysis. All the survey plots south of Winchester Bay ($n = 27$) (Fig. 1) were excluded from

the spatial analysis because the extensive gap between them and the rest of the plots would have produced error in the calculations.

Moran's I test equaled 0.04 (two-sided P value = 0.01, null hypothesis of spatial independence) and Geary's C equaled 0.93 (two-sided P value <0.001, null hypothesis of spatial independence), providing evidence that there is some degree of autocorrelation among residuals. The same results were obtained using tests based on normality and on randomization assumptions (7,9). The magnitude of this autocorrelation, however, is very low, indicating very little spatial dependence between points. Following these results, it is concluded that, even though there might be some bias in the regression model due to spatial dependence of the response variable, this bias is not important enough to invalidate the estimates of the model's precision.

Prediction map. Although both the multiple linear regression and the regression tree produced similar results in terms of model accuracy, the former was chosen to construct the prediction map because it was able to produce a continuous SNC prediction variable. In contrast, the regression tree is able to generate only the prediction values obtained at each leaf (Fig. 3). In situ values of elevation and slope aspect were not available for all the cells of the grid; therefore, in situ values were compared with elevation and aspect from DEMs at different scales. Both elevation and aspect at 90-m pixel size DEMs showed the best correlation with in situ measurements and were used for the prediction map.

A regression analysis was done with the substitute elevation and aspect variables. Results indicate that the substitute variables are associated with SNC rating (P value ≈ 0 , F statistic = 29.6, 6 and 155 degrees of freedom), and the fraction of total variance explained when using these variables ($r^2 = 0.53$) was somewhat lower than with the original variables ($r^2 = 0.57$).

The SNC prediction layer (Fig. 5) shows a strong disease gradient decreasing from the coast-inland, with some discrete areas of very high severity in the Tillamook Bay area, the Nehalem Bay area (the bay just north of Tillamook in Fig. 1), and west of the city of Astoria. Transition zones between areas of contiguous severity levels have a complex, finger-like trace which seems to be determined mostly by topographic features.

DISCUSSION

The association between SNC disease severity and some environmental variables was found to be strong enough ($\approx 60\%$ of the total disease variability observed) to uncover some ecological trends and to encourage future research in this direction to better understand the epidemiology of SNC of Douglas-fir.

The relationship between predicted and observed values was consistent in the regression models and their validations (Figs. 2 and 4). The unexplained model variability takes mainly two forms: (i) the prediction-observed line has a slope lower than the ideal one-to-one line and an intercept greater than zero, producing an underestimate of the SNC rating of the stands with higher disease and an overestimate of stands with lower SNC disease levels, respectively; and (ii) a substantial prediction error at each disease level, as illustrated by the vertical dispersion of points.

Both the under- and over-predictions has to be taken into account when interpreting the geographic distribution of the disease prediction (Fig. 5). However, beyond absolute disease severity values, the general distributional pattern of the prediction matches the common understanding that disease tends to be more severe near the coast. This pattern resembles variables selected as relevant by the linear regression model such as fog occurrence, temperature, and terrain elevation. All these variables have a marked west-east gradient from the coast inland in lines roughly parallel to the coast line. Consistently, fog occurrence, elevation, and temperature were among the main variables the RT model used to separate stand disease severity.

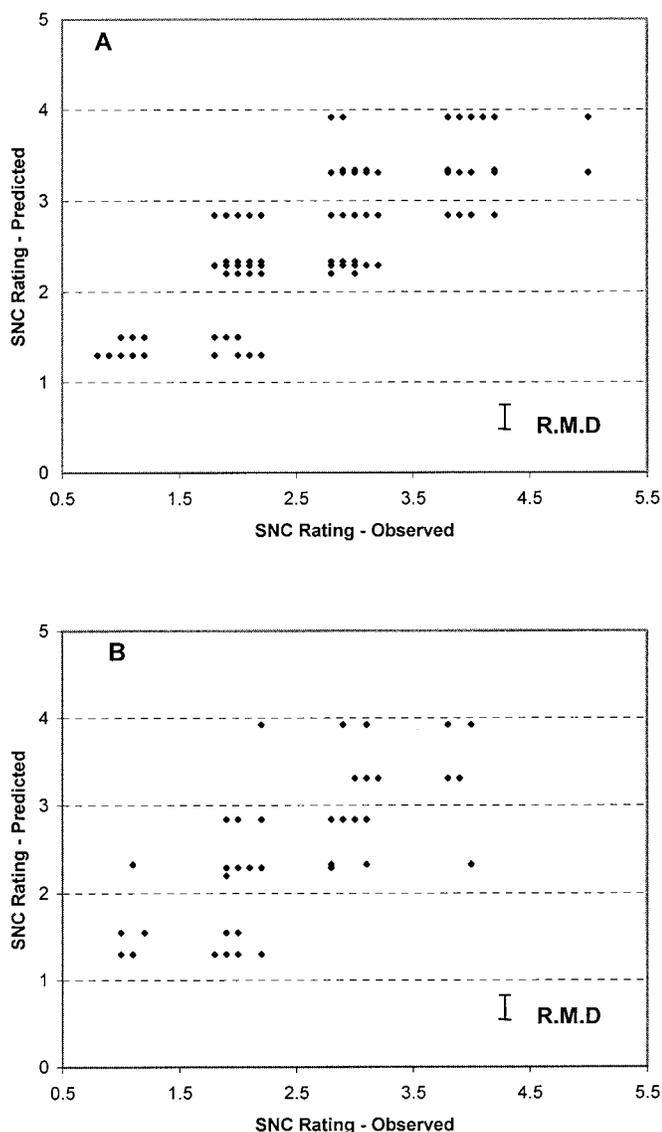


Fig. 4. Comparison of predicted (Y axis) and observed (X axis) Swiss needle cast (SNC) ratings for the **A**, model building plots ($n = 162$) and **B**, validation plots ($n = 50$), based on the regression tree model described in the text. Observed values have been horizontally displaced to facilitate visualization. R.M.D = residual mean deviance (equivalent to standard error in linear regression).

What both the multiple regression and RT models seem to suggest is that moisture (as determined by fog occurrence, precipitation, vapor pressure deficit, altitude, and aspect), warm conditions (temperature in winter, vapor pressure deficit, altitude, and aspect), but not excessive heat (temperature in summer) seem to correlate with higher SNC severity. The fact that July maximum temperature tends to correspond to lower disease values is probably because high temperatures in summer combined with low moisture may inhibit the development of the fungus. In areas with higher fog prevalence (fog in the coast is very prevalent in summer in the

Pacific Northwest), the negative effect of high temperature and dryer conditions on the fungus during the summer may be attenuated. In contrast, SE-W slope aspects, which receive more solar energy all year around, seem to favor disease development or expression. As Manter (13) indicates, infection levels alone do not fully explain differences in disease severity between north- and south-facing slopes, suggesting that microclimate also has an impact on the physiology of the infected host itself.

The fact that summer conditions seem very important when modeling SNC is not surprising, given that (i) critical develop-

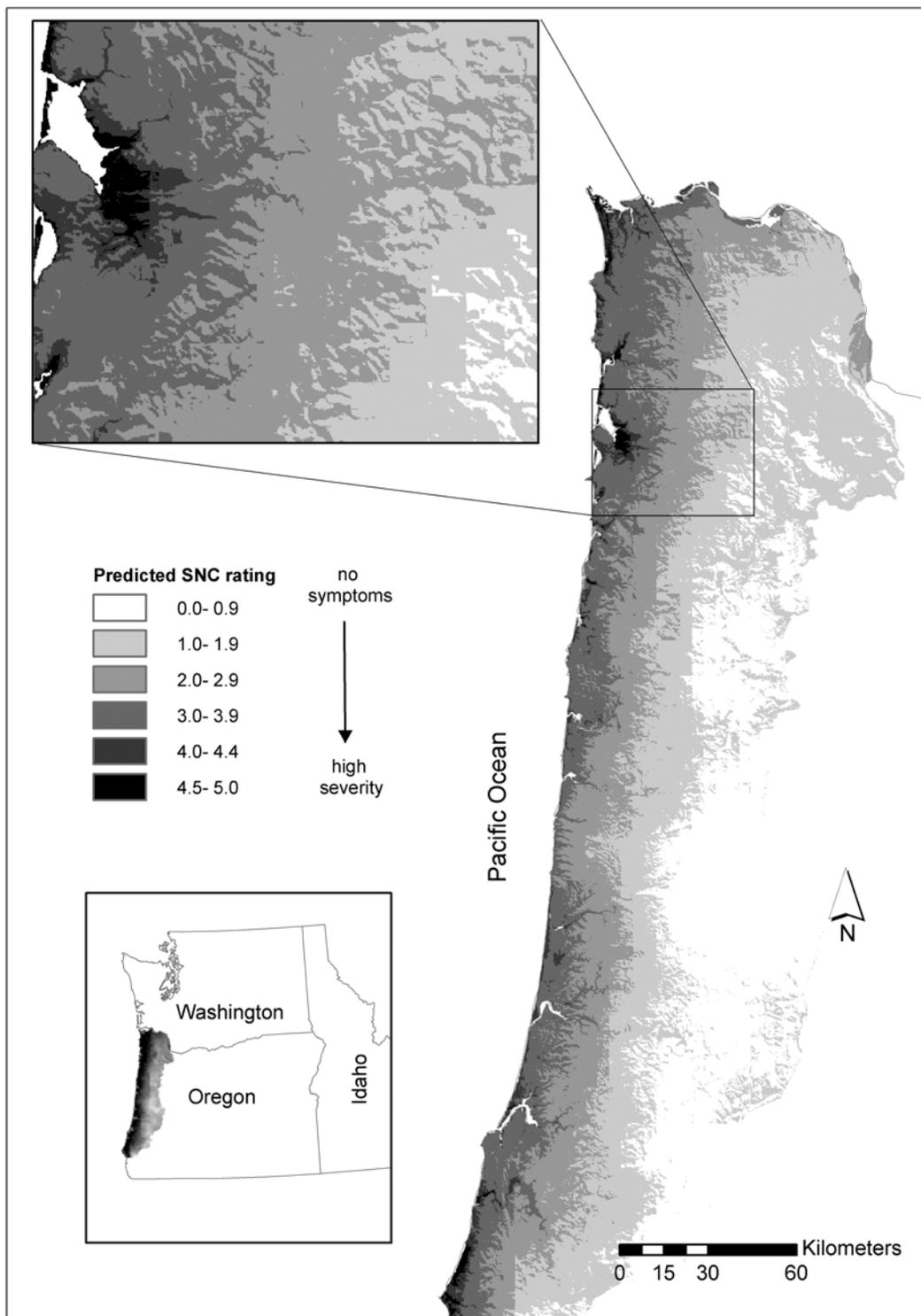


Fig. 5. Map of Swiss needle cast (SNC) disease severity ratings as predicted by the multiple linear regression model. Higher predicted severity areas are depicted with darker tones. To the left, a section of the Tillamook Bay area is displayed at higher resolution to show the typical SNC prediction pattern.

ment stages of the pathogen, such as germination, occur in late spring to summer (3,12), and (ii) the Pacific Northwest has a strong seasonality, especially in precipitation, which influences the functioning and development of both pathogen and host.

Is it possible to develop a model to more accurately predict disease occurrence? To answer this question it is necessary to consider possible sources of disease variability other than soil, climate, and topography. Some potential sources of disease variability, such as host or pathogen genetic variations, and inoculum load, have not been fully considered in this study due to lack of adequate information.

Another important source of variability can be attributed to the way the SNC rating is assigned to stands during the ground survey. The SNC rating system was designed to integrate different disease manifestations, mainly discoloration, early defoliation, and growth reduction. Different values of these three attributes have to be combined to produce a rating scale with few values or categories (1 to 6). This is a difficult task because the combination of the three attributes has to produce a detectable and meaningful categorization. In an effort to make the categorization meaningful, the present SNC rating system is based on some combinations of attribute values, but not on all possible ones. For example, SNC rate 2 consists of foliage color lighter than the normal dark green, 4 or more years of needle retention in most trees, and no reduction in height growth increment. Rate 3 consists of obvious foliage yellowing, most trees retaining 3 years of needles, and no reduction in height growth increment. But what happens when the color is light green, as in rate 2 but the retention is less than 4 years, as in rate 3? In practice, the surveyor has to make arbitrary decisions, which usually are based on previous knowledge and hence highly influenced by the context; that is, immediate past experience, general disease level of the surroundings, and so on. At this point there is enough accumulated experience to devise a rating system that incorporates all the possible combinations of symptoms, which at the same time could be equivalent to the present rating to ensure consistency and continuity.

Our analyses and final model point at the influence of variables consistent with forest manager's experience and scientific findings about SNC distributional patterns. However, our study indicates a more complex scenario than simply the effect of one or two variables. Our observations about possible influences of climate on the disease are consistent with the suggestions of Hansen et al. (10) that climate could have played a role in the increase of SNC symptoms. Hypotheses supporting the idea that SNC is more severe in places where Douglas-fir plantations were established in replacement of other species were not highly supported by our results because none of the 1936 vegetation variables consistently came out as main factors in explaining disease distribution. However, it is important to note that using the vegetation distribution at only one point in time may not be sufficient to understand the complexity of the process of stand replacement. None of the soil variables came up as important, which would have shown some consistency with the soil nutrient imbalance hypotheses (22). However, it is not clear whether any of the soil variables used in this study has any direct relationship with soil nitrogen or any other nutrient abundance.

The results of the model can be used as a hazard prediction to orient integrated forest management. The model also is useful in generating hypotheses and emphasizing the need for information and understanding. Future research should focus on better understanding processes at finer spatial and temporal scales of resolution, such as weather and canopy hydrology, guided by the main findings of this study at the regional level. For example, it seems

reasonable to suspect that fog and ambient temperature, both found to be important variables in this study, must have a strong effect on air moisture and leaf wetness, which in turn might be closely related to fungal spore germination and hyphal development (13). In other words, findings at the scale of the present study need to be corroborated and integrated with studies at lower scales. This includes scales at which the pathogen interacts with the host, mainly histologically, physiologically, and genetically. In this way, relationships and patterns can be integrated with plausible mechanisms.

ACKNOWLEDGMENTS

We thank the Oregon Department of Forestry for its invaluable contribution in providing financial support, survey results, and overall cooperation; and C. Daly and L. Coop for providing climate information.

LITERATURE CITED

1. Boyce, J. S. 1940. A needle cast of Douglas-fir associated with *Adelopus gaeumannii*. *Phytopathology* 30:649-659.
2. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. 1984. *Classification and Regression Trees*. Wadsworth Inc., Belmont, CA.
3. Capitano, B. 1999. The infection and colonization of Douglas-fir by *P. gaeumannii*. M.S. thesis. Oregon State University, Corvallis.
4. Daly, C., Neilson, R. P., and Phillips, D. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteorol.* 33:140-158.
5. Dobbertin, M., and Biging, G. S. 1998. Using the non-parametric classifier CART to model forest tree mortality. *For. Sci.* 44:507-516.
6. Freese, F. 1960. Testing accuracy. *For. Sci.* 6:139-145.
7. Goodchild, M. F. 1986. *Spatial Autocorrelation*. CATMOG 47, Geo Books, Norwich.
8. Griffith, D. A. 1987. *Spatial Autocorrelation: A Primer*. Resource Publications in Geography, Association of American Geographers, Washington, DC.
9. Haining, R. 1993. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, New York.
10. Hansen, E. M., Stone, J. K., Capitano, B. R., Rosso, P., Sutton, W., Winton, L., Kanaskie, A., and McWilliams, M. G. 2000. Incidence and impact of Swiss needle cast in forest plantations of Douglas-fir in coastal Oregon. *Plant Dis.* 84:773-778.
11. Hood, I. A. 1997. Swiss needle cast. Pages 55-56 in: *Compendium of Conifer Diseases*. E. Hansen and K. Lewis, eds. The American Phytopathological Society, St. Paul, MN.
12. Hood, I. A., and Kershaw, D. J. 1975. Distribution and infection period of *Phaeocryptopus gaeumannii* in New Zealand. *N. Z. J. For. Sci.* 5:201-208.
13. Manter, D. K. 2000. Physiological impacts of Swiss needle cast on Douglas-fir. Ph.D. diss. Oregon State University, Corvallis OR.
14. Manter, D. K., Bond, B. J., Kavanagh, K. L., Rosso, P. H., and Filip, G. M. 2000. Pseudothecia of the Swiss needle cast fungus, *Phaeocryptopus gaeumannii*, physically block stomata of Douglas-fir, reducing CO₂ assimilation. *New Phytol.* 148:481-491.
15. Monteith, J. L., and Unsworth, M. H. 1990. *Principles of Environmental Physics*. 2nd ed. Edward Arnold, London.
16. Morissette, J. T., Khorram, S., and Mace, T. 1999. Land-cover change detection enhanced with generalized linear models. *Int. J. Remote Sens.* 20:2703-2721.
17. Neter, J., Wasserman, W., and Kunter, M. H. 1990. *Applied Linear Statistical Models*. 3rd ed., Irwin, Homewood, IL.
18. Ramsey, F. L., and Schafer, D. W. 1997. *The Statistical Sleuth*. Duxbury Thomson Learning, Pacific Grove, CA.
19. Tabachnik, B. G., and Fidell, L. S. 1989. *Using Multivariate Statistics*. 2nd ed. Harper Collins, New York.
20. U.S. Forest Service. 1936. Forest type map, State of Oregon. NW quarter. U.S. Pacific NW Forest Experimental Station.
21. Verbyla, D. L. 1987. Classification trees: A new discrimination tool. *Can. J. For. Res.* 17:1150-1152.
22. Waring, R. H., Boyle, J., Cromack, K., Maguire, D., and Kanaskie, A. 2000. Researchers offer new insights into Swiss needle cast. *West. For. Nov/Dec 2000:10-11.*