A model of genetic variation for *Pinus ponderosa* in the Inland Northwest (U.S.A.): applications in gene resource management

G. E. REHFELDT

USDA Forest Service, Intermountain Research Station, Moscow, ID 83843, U.S.A.

Received August 1, 1990

Accepted April 19, 1991

REHFELDT, G. E. 1991. A model of genetic variation for *Pinus ponderosa* in the Inland Northwest (U.S.A.): applications in gene resource management. Can. J. For. Res. **21**: 1491–1500.

Models were developed to describe genetic variation among 201 seedling populations of *Pinus ponderosa* var. *ponderosa* in the Inland Northwest of the United States. Common-garden studies provided three variables that reflected growth and development in field environments and three principal components of six variables that reflected patterns of shoot elongation. Regression models were developed for describing genetic variation across the landscape. Using functions of latitude, longitude, and elevation as descriptors, these models produced values of R^2 that were as large as 0.66, while averaging 0.39. The models described genetic variation as occurring along relatively steep elevational clines and gentle geographic (i.e., latitudinal and longitudinal) clines. An exercise at validating the models with independent data supported their veracity. Predictions made by the models are applied to limiting seed transfer, designing breeding zones, planning gene conservation programs, interpreting phenotypic variation, and predicting the effects of environmental change on the adaptedness of populations.

REHFELDT, G. E. 1991. A model of genetic variation for *Pinus ponderosa* in the Inland Northwest (U.S.A.): applications in gene resource management. Can. J. For. Res. **21** : 1491–1500.

Des modèles ont été développés afin de décrire la variation génétique parmi 201 populations de semis de *Pinus ponderosa* var. *ponderosa* provenant de l'intérieur des terres dans la région du Nord-Ouest des États-Unis. Trois variables représentatives de la croissance et du développement au champ ainsi que trois composantes principales de six variables représentatives des patrons d'élongation de la pousse furent obtenues à partir des études en plantation comparative. Des modèles de régression furent mis au point afin de décrire la variation génétique d'origine géographique. La moyenne des coefficients R^2, obtenus à partir de modèles dont les descripteurs étaient la latitude, la longitude et l'altitude, était de 0,39, avec certaines valeurs aussi grandes que 0,66. La variation génétique décrite par les modèles était répartie selon des clines assez accentués de variation altitudinale, et des clines moins accentués de variation en latitude et longitude. La validation des modèles à l'aide de données indépendantes supportait leur véracité. Les prédictions faites par les modèles sont appliquées à la délimitation des zones de transfer de semences, à l'établissement des zones d'accouplement, aux programmes de conservation du patrimoine génétique, à l'interprétation de la variation phénotypique, et à la prédiction des effets de changements environnementaux sur l'adaptation des populations.

[Traduit par la rédaction]

Introduction

Throughout much of the 20th century, provenance research has been devoted to describing genetic variation and relating patterns of variation to environmental gradients. During the last 2 decades, however, a transition into a phase devoted to predicting genetic responses has occurred. Influencing this transition was the imaginative approach of Morgenstern and Roche (1969) for interpreting patterns of genetic variation according to concepts of quantitative genetics, an approach that Morgenstern (1978) later applied to provenance variation in *Picea mariana* (Mill.) B.S.P. Yet the impetus behind the transition has been R. K. Campbell's exemplary works with *Pinus sylvestris* L. (Campbell 1974*a*) and *Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii* (Campbell 1974*b*), which illustrated the development and application of models of genetic variation for guiding seed transfer in artificial reforestation. Today, functional models of genetic variation are available for introducing *Pinus contorta* Dougl. provenances to Sweden (Lindgren et al. 1976); matching *P. sylvestris* (Eriksson et al. 1980; Raymond and Lindgren 1989; Roberds and Namkoong 1989) and *Fraxinus americana* L. (Kung and Clausen 1984; Roberds et al. 1990) genotypes to planting sites; and controlling seed transfer in *P. contorta* (Rehfeldt 1988; Ying et al. 1989), *Picea sitchensis* (Bong.) Carr. (Campbell et al. 1989), *Pinus ponderosa* Laws. var. *scopulorum* (Rehfeldt 1990), *P. menziesii* var. *menziesii* (Campbell 1979, 1986), and *P. menziesii* var. *glauca* (Beissn.) Franco (Rehfeldt 1989).

Authenticity of a model depends on experimental procedures, statistical approaches, and validation. With regard to the experimental procedures, Campbell (1986) emphasizes that realistic models of genetic variation require, first, a sample of genotypes that is geographically and ecologically representative of a species in the region of study and, second, an evaluation of the genotypes under conditions that force expression of adaptive differences. As a result, the scope of a model for species inhabiting heterogeneous environments is often limited by the demands of sampling and testing. Statistical procedures and their safeguards are well documented, for example, by Draper and Smith (1981). These authors stress that the development of a model should be accompanied with some type of validation. Of the models of genetic variation available for forest trees, however, only that for *P. menziesii* var. *glauca* (Rehfeldt 1986*c*, 1989) has been verified with independent data.

In this paper, a model of genetic variation for ponderosa pine (*P. ponderosa* var. *ponderosa*) is developed for the Inland Northwest of the United States. Variation among populations in the Snake and Salmon river basins (Rehfeldt 1986*a*) is combined with that for the middle Columbia River system (Rehfeldt 1986*b*) to produce a model of genetic varia-

Printed in Canada / Imprimé au Canada

tion for the entire region. Validation is attempted with the independent data that are currently available, and predictions are applied to topics in gene resource management.

Procedures

Previous studies of population differentiation in ponderosa pine of the Inland Northwest have compared seedlings from 138 populations from the mid-Columbia River system (study 1, Rehfeldt 1986*b*) and 64 populations from the Snake and Salmon systems (study 2, Rehfeldt 1986*a*) (Fig. 1, Table 1). Both studies used randomized complete blocks in common gardens to assess genetic variation among populations in separate tests of (*i*) growth and development of 3-year-old trees in field environments and (*ii*) patterns of shoot elongation of 2-year-old trees in the greenhouse. The two studies, however, were completed in different years, involved a different set and number of populations, and were cultured under somewhat different environmental conditions. As a result, only nine of the numerous variables analyzed originally were measured under similar conditions in the two studies. Greenhouse tests provided six variables that described patterns of shoot elongation: the initiation, start, cessation, duration, rate, and amount of elongation. The initiation and start of elongation were defined as the day on which either 2 or 8 mm of elongation had occurred, respectively, and rate was defined as elongation per day during the period between which 20 and 80% of the total elongation had occurred. Field tests, conducted at low elevation (670 m) at the Priest River Experimental Forest (48.5°N, 116.7°W), provided the remaining three variables: 3-year height, leaf length, and deviation from regression of 3-year height on 2-year height. In both of the studies, field tests were managed under cultural regimes that were optimal for the growth of trees.

Of the 201 populations tested in the two studies, one, Three Mile, was common to both studies and thereby provides the link between them. Three Mile was from an elevation of 1340 m and was centrally located with regard to the geographic regions sampled by the two studies (Fig. 1). Even though seedlings from this population performed remarkably similarly in the two studies, overall means and variances differed (Table 2). This meant that data had to be transformed and scaled to obtain a single data set within which the population means obtained in one study were directly comparable with those obtained in the other. The procedures for doing this followed Rehfeldt (1989) and involved (*i*) transforming data within each block to standard normal deviates (Z) to produce data sets with common variance, a mean of zero, and an absence of block effects; (*ii*) deriving a scaling factor, which is the difference between the study 1 and study 2 mean performances of the Three Mile poplulation (Table 2); and (*iii*) scaling individual observations of study 2 to those of study 1 by adding the scaling factor to the mean standard deviate.

These procedures produced data sets of scaled standard deviates for three variables from field tests and six variables from greenhouse tests. Statistical analyses appropriate to these data, however, were conditioned by differing experimental procedures (Table 1). First, field tests included 201 populations, while tests conducted in the greenhouse included 161. Second, field tests involved different seedlings from those used in the greenhouse. Third, only plot means were available for variables measured in the field tests of study 2. And, finally, populations were represented by a mixture of wind-pollinated seeds from 10 trees in study 1 but from only 5 trees in study 2.

In an attempt to reduce the number of dimensions about which differentiation was being displayed, principal component analyses (SAS Institute Inc. 1982) were conducted on the correlation matrix of individual observations within the scaled data sets. Separate analyses were made for field and greenhouse data because covariances could not be calculated between variables measured in the two tests.

Although differences among populations for all of the original variables had been established in previous analyses, analyses of variance were performed primarily to provide an average error

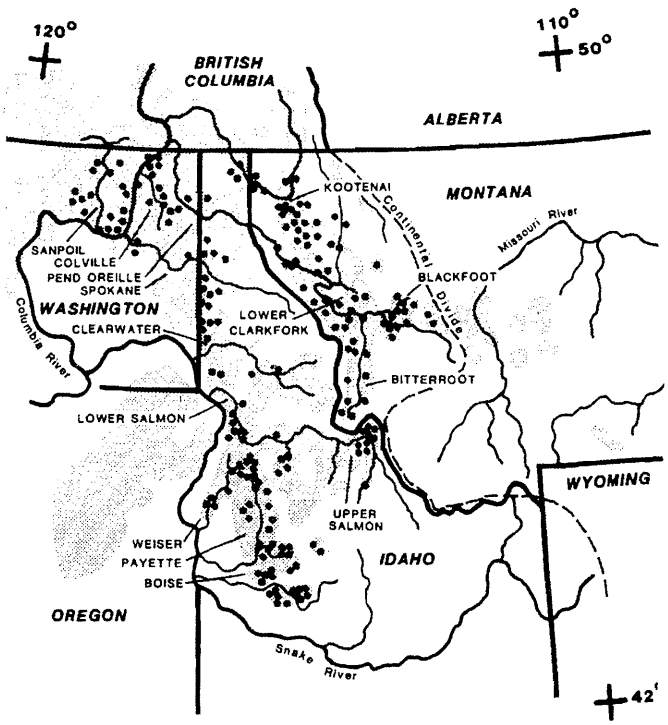

FIG. 1. Geographic distribution of ponderosa pine (stippling, from Little 1971) and populations sampled within 13 drainages in the Inland Northwest. The triangle in the Weiser drainage locates the Three Mile population.

TABLE 1. Composition and design of studies that assessed genetic variation in growth and development in a field environment and patterns of shoot elongation in a greenhouse

	Study 1		Study 2	
	Greenhouse	Field	Greenhouse	Field
Year completed	1984	1985	1983	1984
No. of populations tested	98	138	64	64
Blocks	3	3	3	5
Plot size[a]	9	9	9	9
Data base[b]	I	I	I	P

[a]Number of seedlings from a population planted in a row plot within each block.
[b]I, individual trees; P, plot means.

variance against which to judge differentiation. For these analyses, the effects of populations were tested against the pooled interaction of blocks and populations.

Developing a general model of genetic variation used procedures detailed earlier (Rehfeldt 1989): (*i*) deriving independent variables from latitude, longitude, and elevation that could serve as aliases for the complex environmental gradients that are operating in natural selection; (*ii*) subdividing the region of study into geographic subunits that could be included in regression models by means of dummy variables (Draper and Smith 1981) and within which independent variables could be nested; (*iii*) using multiple regression for determining which of the independent variables could be included in the same regression model without producing a singular variance–covariance matrix; (*iv*) screening the independent variables by stepwise regression, the best model of which was judged relative to statistical significance, the Mallows statistic, and patterns displayed by the residuals (Draper and Smith 1981); (*v*) refining a stepwise model with multiple regression to develop

TABLE 2. Means ($\bar{x}$) and standard deviations (SD) for all populations in each test, means for the Three Mile population, and scaling factors

Variable[a]	All populations				Three Mile population				
	Study 1		Study 2		Original units		Standard deviates		
	$\bar{x}$	SD	$\bar{x}$	SD	Study 1 ($\bar{x}$)	Study 2 ($\bar{x}$)	Study 1 ($\bar{x}$)	Study 2 ($\bar{x}$)	Scaling factor[b]
Field tests									
3-year ht. (cm)	48.4	5.1	41.3	4.7	45.6	46.2	−0.546	1.055	−1.601
Leaf length (cm)	13.2	0.9	15.2	0.8	14.1	14.8	0.949	−0.416	1.365
Deviation (cm)[c]	0.0	2.3	0.0	2.5	−2.9	−1.9	−0.921	−0.668	−0.253
Greenhouse tests[d]									
Initiation (days)	2.5	1.7	2.1	1.5	3.0	1.8	0.310	−0.183	0.493
Start (days)	7.7	1.6	6.7	1.6	8.2	7.1	0.299	0.266	0.033
End (days)	36.7	4.2	33.8	5.2	34.7	33.6	−0.481	−0.054	−0.427
Duration (days)	34.3	4.4	31.8	5.3	31.7	31.9	−0.568	−0.002	−0.566
Rate (mm/day)	5.1	1.2	4.4	1.0	4.8	4.3	−0.264	−0.111	−0.153
Amount (cm)	10.7	2.5	9.0	2.1	9.5	8.6	−0.488	−0.186	−0.301

[a]Units given in parentheses apply only to columns 2-7; they do not apply to columns 8-10, standard deviates and scaling factors.
[b]Study 1 mean minus the study 2 mean in standard deviates.
[c]Deviation from regression of 3-year height on 2-year height.
[d]Variables describe shoot elongation.

the most parsimonious model; (*vi*) finally, plotting elevational and geographic (latitudinal and longitudinal) patterns of variation to assure that the models were sensible biologically. A model, for instance, might be recognized as implausible and rejected if discontinuous patterns of genetic variation were described across the boundaries of geographic subunits and, particularly, if the size of the discontinuities approached the standard error of the mean ($s_{\bar{x}}$).

Rates of differentiation along geographic or elevational clines were interpreted relative to the least significant difference (LSD) (see, e.g., Steel and Torrie 1960) among populations at the 20% significance level (LSD 0.2). Values of LSD were used because stepwise models developed from numerous independent variables are subject to overfitting and overparameterizing (Draper and Smith 1981). Using LSD, therefore, guarded against accepting fallacious results. The 20% significance level was used to guard against accepting no differences among populations when differences actually exist (type II errors); such errors provide the greatest potential for disaster when models are applied. Values of LSD were calculated from the pooled interaction of blocks and populations in the analysis of variance.

Validity of the regression model was assessed by correlating predicted values with variables describing the field performance of 54 populations that were independent of the modeling. These populations were present in 12-year-old provenance tests on two field sites in northern Idaho.

TABLE 3. Eigenvector coefficients and eigenvalues for separate principal component analyses of variables measured in the field and greenhouse

Variable	Principal component		
	First	Second	Third
Field tests			
3-year ht.	0.67	0.14	0.73
Leaf length	−0.40	0.89	0.20
Deviation[a]	0.62	0.43	−0.65
Eigenvalue	1.55	0.91	0.53
Proportion of variance	0.49	0.34	0.17
Greenhouse tests[b]			
Initiation	−0.276	0.538	0.414
Start	−0.291	0.583	0.176
Cessation	0.408	0.507	−0.247
Duration	0.482	0.297	−0.378
Rate	0.389	−0.150	0.694
Amount	0.537	0.050	0.333
Eigenvalue	3.04	1.65	1.00
Proportion of variance	0.51	0.28	0.17

[a]Deviation from regression of 3-year height on 2-year height.
[b]Variables describe shoot elongation.

Results

The principal component analysis of field data provided three components that individually accounted for 49, 34, and 17% of the variance in plot means (Table 3). Analyses of variance indicated that populations differed significantly for all of the principal components as well as all of the original variables. Because the principal component analysis was not able to reduce the number of variates about which differentiation was being expressed, subsequent analyses of field data used the original variables.

Principal component analyses of greenhouse data attributed 51% of the variance to a first component, 28% to the second, 17% to a third, 4% to a fourth, and less than 1% to the remainder (Table 3). Analyses of variance indicated that populations differed significantly ($p < 0.05$) for only the first three of these components, hereafter designated as PCG1, PCG2, and PCG3. Even though resulting in a net loss of information (Johnson and Wichern 1982), principal components seem appropriate for subsequent analyses because they allow the number of variates to be halved while accounting for about 96% of the total variance. The eigenvector coefficients (Table 3) indicate that PCG1 primarily was determined by variables reflecting growth potential: the amount, rate, cessation, and duration of shoot elongation. With variables describing the beginning and cessation of shoot elongation as primary constituents, PCG2 seemed to reflect the portion of growth rhythm that was independent of growth potential. PCG3 was strongly controlled by the rate of shoot elongation.

TABLE 4. Comparison of two regression models used to describe patterns of genetic variation among populations

Variable	Model 4[a] R^2	Model 4[a] Adjusted[c] R^2	Model 4[a] Residual mean square	Model 4[a] Independent variables	Model 15[b] R^2	Model 15[b] Adjusted[c] R^2	Model 15[b] Residual mean square	Model 15[b] Independent variables
Field tests								
3-year ht.	0.68	0.66	0.4307	7	0.70	0.68	0.4058	8
Leaf length	0.54	0.53	0.3921	6	0.53	0.51	0.4060	5
Deviation	0.10	0.08	0.4794	3	0.13	0.11	0.4221	6
Greenhouse tests								
PCG1	0.55	0.53	0.3708	6	0.55	0.53	0.3672	6
PCG2	0.29	0.23	0.1584	9	0.29	0.25	0.1586	9
PCG3	0.18	0.17	0.0974	3	0.20	0.16	0.0972	7

[a]No geographic subunits.
[b]Two geographic subunits: (*i*) Montana plus Idaho and Washington north of 48°N latitude and (*ii*) Idaho plus Washington south of 48°N latitude.
[c]R^2 adjusted for the degrees of freedom.

TABLE 5. Probability (p) of statistical significance for regression coefficients included in model 4

Variables	Independent variables[a] Elev.	Elev.^{2}	Lat.	Long.	NW	SW	Lat.^{2}	Long.^{2}	NW^2	SW^2	Lat.^{3}	Long.^{3}	NW^3	SW^3
Field tests														
3-year ht.	<0.01		<0.01	<0.01			<0.01	<0.01		<0.01			<0.01	
Leaf length	0.04		0.01			0.01				<0.01		0.01		<0.01
Deviation	<0.01								0.05				0.05	
Greenhouse tests														
PCG1	<0.01			<0.01		0.04		<0.01		0.04				0.04
PCG2	<0.01		<0.01			<0.01		<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	
PCG3						0.05				0.04				0.05

NOTE: Independent variables included in model 4 are indicated by the probability that the coefficient was statistically significant.
[a]NW, latitude × longitude; SW, longitude ÷ latitude.

Multiple regression analyses of the six dependent variables considered 16 different groups of independent variables, 15 of which included various geographic subunits (Idaho or Montana, for instance) as dummy variables. All but one of these 15 models predicted discontinuous patterns of genetic variation across subunit boundaries, with the size of the discontinuities approaching $s_{\bar{x}}$. Results of this type are attributable to the modeling procedures, likely reflect overfitting, and therefore, can be rejected because they are biologically implausible. The exception, group 15 (Table 4), produced sensible results but provided models that were only slightly better statistically than those of group 4 (Table 4), which used no geographic subunits. Because the models produced by both groups described essentially the same patterns of variation, only the models produced by group 4 will be considered.

All of the regression models accounted for a significant ($p < 0.01$) proportion of the variance in the dependent variables (Table 3). Values of R^2 were as high as 0.68, while averaging 0.39 (Table 4). The best fitting models involved variables reflecting growth potential (PCG1 and 3-year height) and leaf length; the worst model accounted for only 10% of the variance (deviation from regression of 3-year height on 2-year height). The level of significance of the individual coefficients was generally high, with nearly all coefficients significant at a probability far less than 0.01 (Table 5). The regression equations are available from the author.

Like the results of the two studies on which the current analyses were based, the models described genetic variation as occurring along both elevational and geographic clines. In fact, an elevational cline was a component of all regression models, except that for PCG3 (Table 5). The steepest of these clines was for 3-year height, and this cline is presented in Fig. 2 for 13 geographic intercepts. Rates of differentiation along the elevational cline in relation to LSD 0.2 suggest that populations within the same drainage that are separated by 416 m in elevation tend to be genetically different (80% confidence level) for 3-year height. Although not illustrated, the elevational interval associated with differences equalling LSD 0.2 were 542 m for PCG1, 1168 m for the deviation from regression, 1191 m for PCG2, and 2123 m for leaf length. The last three of these intervals exceed the elevational distribution of ponderosa pine within any of the drainages studied.

A geographic component to genetic variation also is illustrated in Fig. 2 by the regression lines of different intercept. This component is detailed for all six variables in Fig. 4, where genetic variation among populations growing at an elevation of 1250 m is represented by isopleths. The interval between isopleths equals 0.5 LSD 0.2, and therefore, populations separated by a geographic distance equalling two intervals are expected to differ at the 80% confidence level. According to this figure, when populations from the same elevation are compared, those from north central Idaho tend to have the highest growth potential (3-year height and PCG1); those from southern Idaho, the longest leaves and highest values of PCG3; and those from north central Washington and central Montana, the lowest values of PCG2.

TABLE 6. Simple correlations of predicted values of six variables with independent data from long-term field tests of 54 provenances

	Predicted variables					
Independent data	3-year ht.	Leaf length	Deviation	PCG1	PCG2	PCG3
3-year ht.[a]	0.40**	0.12	0.28*	0.40**	0.30*	0.21
6-year ht.[a]	0.24	−0.20	0.04	0.13	0.03	−0.08
12-year ht.[a]	0.21	−0.20	−0.02	0.09	−0.00	−0.11
% needle cast[b]	−0.45**	0.35**	−0.22	−0.39**	−0.22	−0.28*
% pitch resin midge[c]	−0.30*	0.51**	−0.01	−0.24	−0.14	0.42**

NOTE: *, statistical significance at the 5% level; **, statistical significance at the 1% level.
[a]Data are from two plantations provided by the Inland Empire Tree Improvement Cooperative.
[b]From Hoff (1986).
[c]From Hoff (1989).

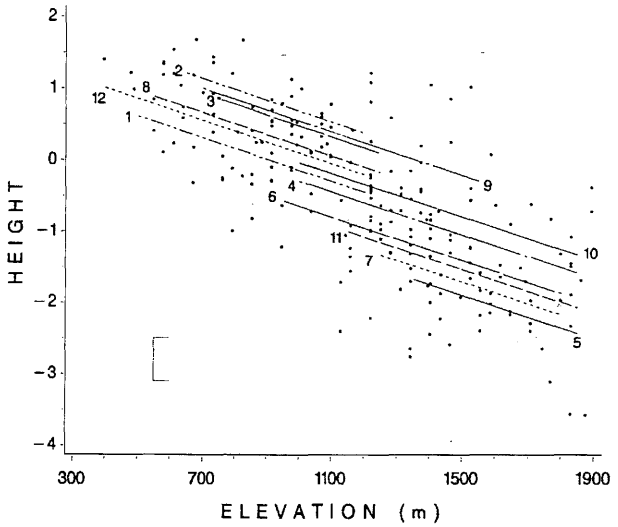

FIG. 2. Population means (scaled units of Z) for 3-year height plotted by seed source elevation. The vertical bracket near the origin quantifies LSD 0.2. Regression lines denote individual drainages that are located in Fig. 1: 1, Sanpoil; 2, Spokane; 3, Clearwater; 4, lower Salmon; 5, upper Salmon; 6, Payette and Weiser; 7, Boise; 8, Kootenai; 9, lower Clarkfork; 10, Bitterroot; 11, Blackfoot; and 12, Colville.

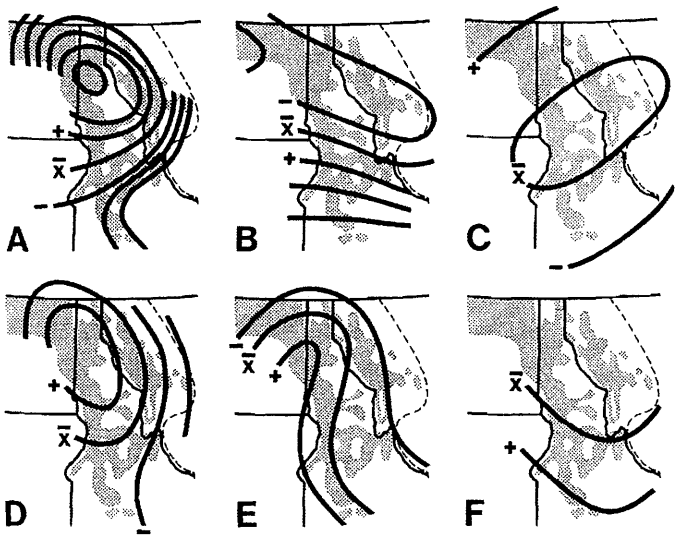

FIG. 3. Geographic patterns of genetic variation predicted by regression models for a constant elevation (1250 m) and presented as isopleths of equal performance. The interval between isopleths equals 0.5 LSD 0.2. Shading represents the distribution of ponderosa pine. (A) Three-year height. (B) Leaf length. (C) Deviation from regression of 3-year height on 2-year height. (D) First principal component of greenhouse data (PCG1). (E) PCG2. (F) PCG3.

The exercise in validating the model involved correlating predicted values with observed variables for populations other than those on which the models were based (Table 6). The independent data involved the growth, insect resistance, and disease resistance of 54 populations growing in 12-year-old provenance trials on two sites in northern Idaho. Most notably, the mean height of these provenances was significantly ($p < 0.05$) correlated with predicted values of 3-year height and PCG1 at age 3 but not at subsequent ages (Table 6). In addition, the incidence of needle cast (*Lophodermium baculiferum* Mayr) was negatively correlated with growth potential (3-year height and PCG1) and positively associated with leaf length. And finally, a significant correlation was obtained between predicted leaf lengths and the incidence of the pitch resin midge (*Cecidomyia piniinopis* Osten Sacken), which, perhaps incidentally, pupates in the leaves. The longer the leaves, the greater the frequency of the insect. Quite inexplicably, the incidence of both pests also correlated significantly with PCG3.

Discussion

The results describe genetic variation among populations along elevational and geographic clines. Interpretation of genetic differentiation along the clines, however, is essentially the same as described in the two studies (Rehfeldt 1986*a*, 1986*b*) on which the current analyses are based. As the elevation of the seed source increases, growth potential and its component variables sharply decrease. This response presumably results from selection associated with the length of the frost-free season, a variable that declines by about 90 days across an elevational interval of 1000 m (Baker 1944). Because populations separated by about 400 m tend to differ in growth potential, sites that differ by about 35 days in average frost-free period tend to support genetically different populations. For variables other than those determining growth potential, however, elevational clines are so flat that practical interpretations are absurd.

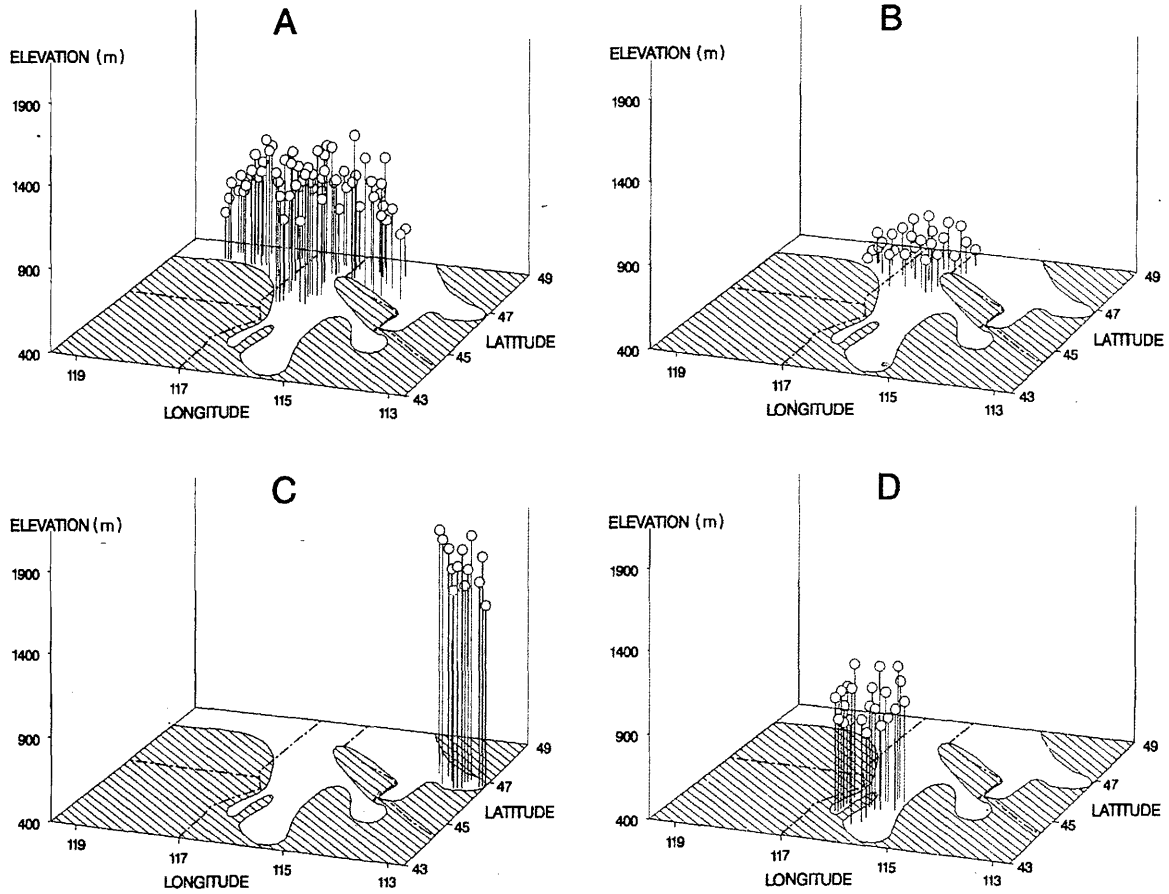

FIG. 4. Recurrence of selected genotypes as predicted by the model of genetic variation. (A) Populations with growth potentials slightly above average in otherwise average individuals. (B) Populations with the highest growth potential. (C) Populations with low growth potential and short leaves. (D) Populations with long leaves on otherwise average individuals. State boundaries are denoted on the floor of the diagram. Hatching marks the area not inhabited by ponderosa pine.

The geographic clines (Fig. 3) describe patterns of genetic variation that correspond to gradients in a rather large number of environmental variables (United States Department of Commerce 1968). In this region, elevations at the valley floor increase from northwest to southeast. The climate in the north central portion of the region is typified by winter precipitation and summer drought. To the east and southeast, the climate grades into an arid continental pattern, with the scant precipitation generally occurring in brief summer storms. In the extreme northwestern portion of this region, the climate is mild but extremely arid. Geographic patterns seem to reflect these general environmental gradients.

The procedures employed produced a model of genetic variation that was sensible biologically and corroborative of previous results. The procedures, however, tacitly assume that no biases were introduced by scaling of data sets. Indeed, scaling based on the performance of a single population provided a tenuous link between studies. Experimental errors biasing the performance of this population in either study would bias the distribution of population means in the entire scaled data set. A lack of bias would also demand no interactions between populations and the environmental conditions in the two studies.

Previous analyses that used similar scaling techniques (Rehfeldt 1989) allowed the effectiveness of scaling to be vindicated in a manner that was not possible for the current data. To be sure, the similarities in means and variances between studies for both the common population and all populations (Table 2) provide some assurance that experimental errors and genotype by environment interactions were not responsible for overt scaling biases. Also pertinent is an experimental approach (see Rehfeldt 1984) that emphasizes control of extraneous environmental effects while assessing genetic variation in individual traits rather than in general performance. Such programs tend to minimize experimental errors while relegating interactions between genotype and environment to scale effects rather than to wholesale changes in rank. For these reasons, biases that may have been introduced by the procedures are believed to have little effect on practical interpretations.

The exercise of verifying the models with independent data tended to be supportive. One must be aware, however, that long-term provenance tests by nature are composed of material most of which is not adapted to the planting site. As a result, environmental effects eventually mask the growth potential of all but the adapted provenances. In the present case, predicted values of variables related to growth potential were significantly correlated with the observed 3-year height. But, as the trees aged, environmental effects increased, and the strength of the correlations decreased. Nevertheless, predicted leaf lengths were strongly correlated with, first, the incidence of a needle disease and, second,

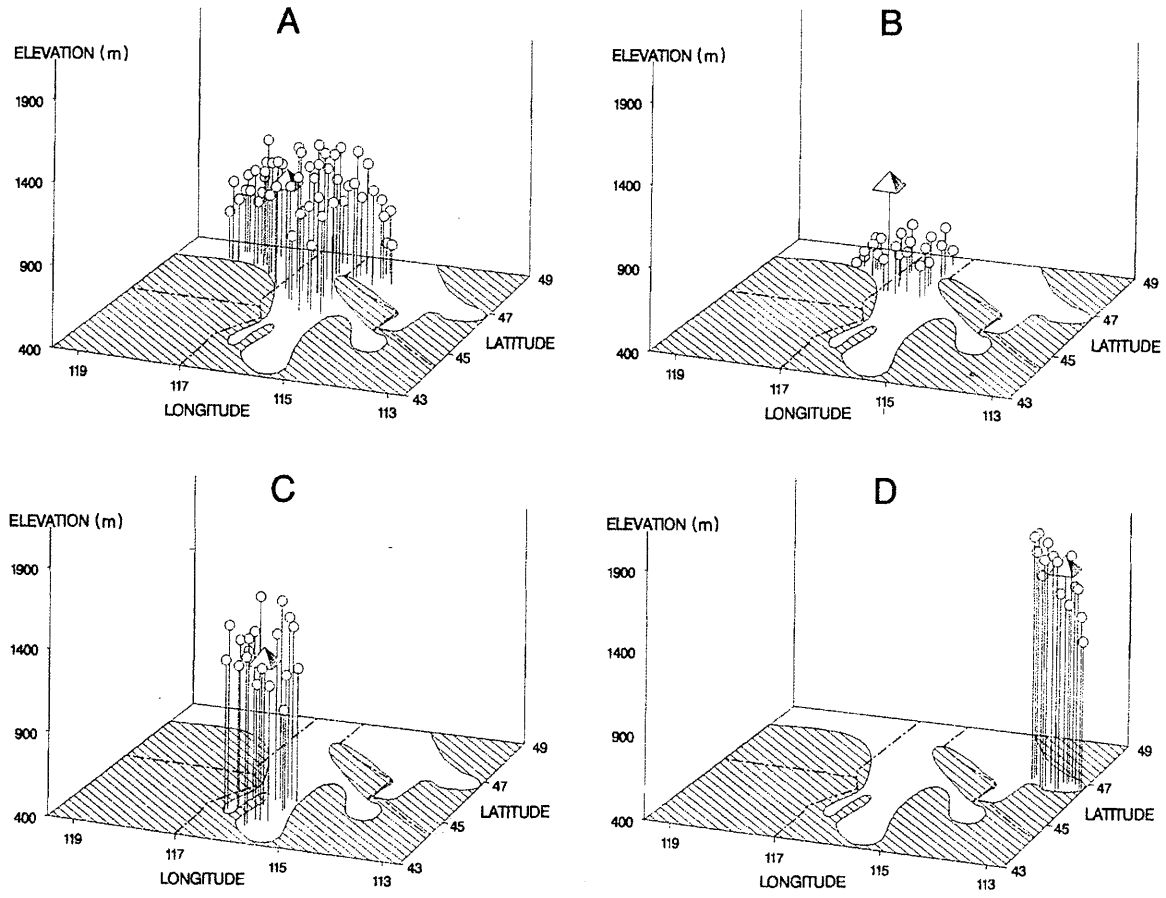

FIG. 5. Using models to guide seed transfer. (A) The pyramid locates a site at 1100 m in the Spokane drainage, and balloons locate compatible sites. (B) Balloons locate populations with a growth potential 10% greater than the population native to the Spokane site. (C) The pyramid locates a site at 1300 m in the Payette drainage, and balloons locate compatible sites. (D) The pyramid locates a site at 1800 m in the Blackfoot drainage, and balloons locate compatible sites.

the incidence of an insect that pupates in the leaves. In total, therefore, the exercise supports the veracity of the model, even though complete vindication is still impossible.

Concomitant variation along elevational and geographic clines describes complex patterns across the landscape. Figure 2 suggests, for instance, that populations inhabiting the same elevation in different drainages tend to be different genetically. Or, conversely, similar genotypes are expected to recur at different elevations in different drainages. Thus, populations with a moderately high growth potential (3-year height = 1.0 in scaled units of Z) are expected to occur at about 400 m in the Colville drainage, 700 m in the Clarkfork drainage, and 800 m in the Spokane drainage (Fig. 2). But, examining the frequency with which similar genotypes recur with respect to all variables is facilitated by using the regression equations to generate a data base containing predicted values for the entire geographic and elevational distribution of the ponderosa pine within the region of study. By assuming that each observation in the data base is surrounded by a confidence interval of ± 0.5 LSD 0.2, one can readily group similar genotypes.

The recurrence of similar genotypes across the landscape is illustrated in Fig. 4. A broad distribution is expected for individuals that are slightly above average in growth potential but have average values of the other characters (Fig. 4A). Although limited to an elevational distribution of 400 m within a single drainage, this genotype is expected to occur in 6 of the 13 drainages and to have an elevational distribution that encompasses 800 m. By contrast, genotypes of highest growth potential are expected at the lowest elevations (400 to 600 m) in only three of the drainages (Fig. 4B). Figure 4C indicates that populations with short leaves and low growth potential are expected in the region that Conkle and Critchfield (1988) designated as transitional between the long-leaved *P. ponderosa* var. *ponderosa* and the short-leaved var. *scopulorum*. And finally, Fig. 4D shows that long leaves on otherwise average individuals are expected in genotypes distributed across only 300 m of elevation within the Weiser, Payette, and Boise drainages. In this latter area, no potentially interbreeding populations occur to the south and east, while those to the west are distant. As a result, the recurrence of similar genotypes appears to be limited.

While contributing to an understanding of microevolution, the models also have direct practical application in gene resource management. All uses, however, require that one accepts the untested assumption that natural populations are well adapted to the sites on which they grow. As a guide for limiting the transfer of seeds in artificial reforestation, the models use to advantage the principle that similar genotypes recur across the landscape. For instance, one may seek either the localities for which a given seed source is

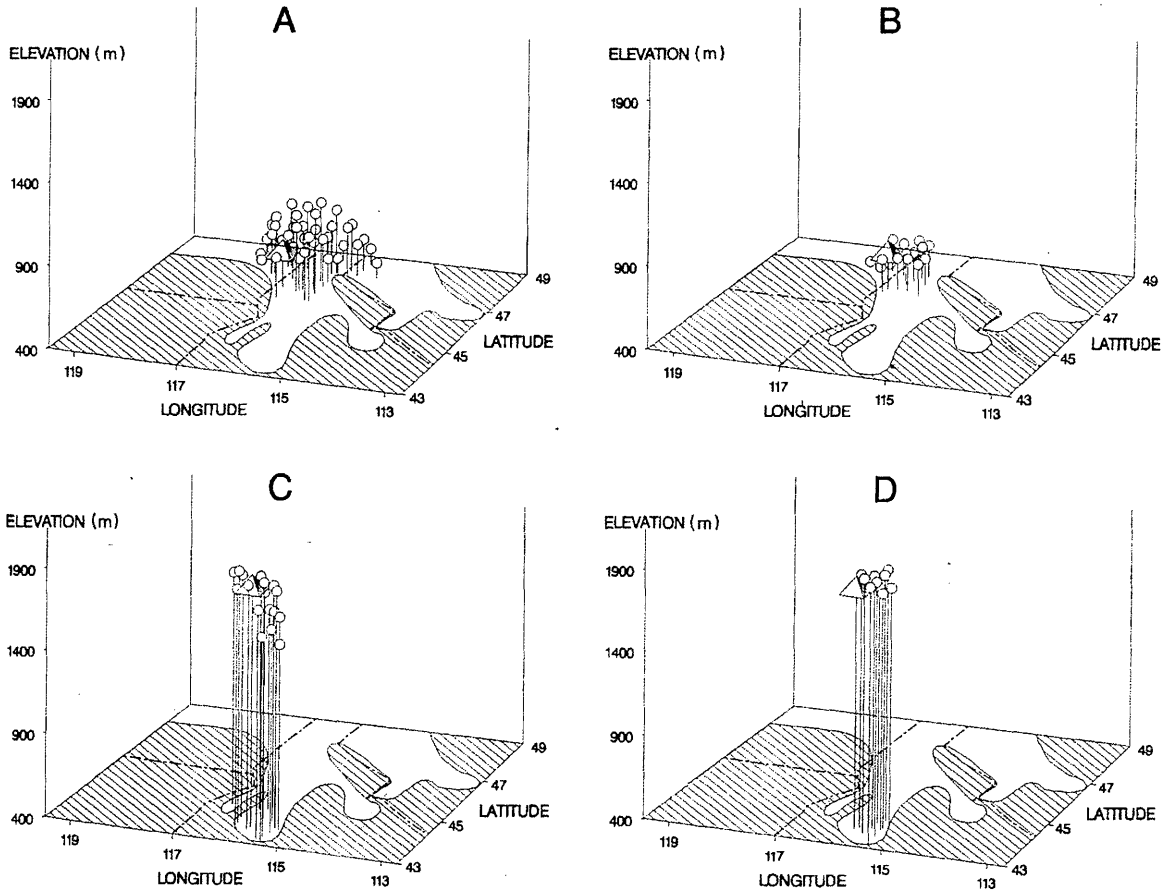

FIG. 6. Using models to predict the effects of environmental change on the adaptedness of populations. (A) The pyramid locates a site at 650 m in the Spokane drainage, and balloons locate populations expected to be compatible in the contemporary climate. (B) Balloons locate populations expected to be compatible with the Spokane site if the frost-free period increases by 20 days. (C) The pyramid locates a site at 1750 m in the Boise drainage, and balloons locate populations expected to be compatible in contemporary climates. (D) Balloons locate populations expected to be compatible with the Boise site if the frost-free period decreases by 20 days.

adapted or the seed sources that are adaptationally suited for a particular site. As shown in Fig. 5A, to maintain the adaptedness typical of natural populations on a planting site at 1100 m in the Spokane drainage, seeds could be imported from elevations between 700 and 900 m in the Sanpoil drainage, 800 and 1200 m in the Colville drainage, 800 and 1300 m in the Clearwater drainage, 600 to 1100 m in various portions of the Kootenai drainage, and 800 to 1200 m in both the Clarkfork drainage and the Spokane drainage itself. Alternatively, seeds from a seed production area located at the same site could be planted at the sites indicated without a loss in adaptedness. To be sure, this Spokane site is centrally located within the region, and therefore, recurrence is high. Similar examples from the periphery of the region produce much different results (Figs. 5C and 5D). Without doubt, the degree of recurrence associated with peripheral populations is expected to be much less than that for the Spokane.

A forest manager might also choose to disregard the risk of poor adaptation and plant the Spokane site with trees exhibiting a 3-year height that is 10% greater (original units of measure) than that of the local source (Fig. 5B). In this case, the appropriate seeds would come from low elevations (400 to 700 m) where the climate is more mild than at the planting site itself.

Another use of the model might involve designing the composition of seed orchards. Seed orchards ordinarily serve a single seed zone or breeding zone within which all genotypes are assumed to be adaptively similar. Thus, discrete seed zones might be constructed from patterns of variation such as those detailed in Figs. 2 and 3. If one accepts the confidence intervals used in this paper, the geographic extent of a discrete seed zone should not exceed two intervals between isopleths for any of the variables (Fig. 3). This means that about 12 geographic zones would be required for the region of study. Because each zone should encompass only 400 m of elevation (Fig. 2), a total of 30 seed zones would be required to represent the species' ecological distribution. As summarized by J. Manzak,[1] however, less than 4 million ponderosa pine seedlings are planted each year in the entire region. This means that for many of the seed zones, the number of seedlings required each year would be so small that seed procurement and nursery practices would be both inefficient and uneconomical. Discrete seed zones, however, fail to consider that

[1]J. Manzak, Champion International Corporation, Milltown, MT. Unpublished report presented at the Annual Meeting of the Inland Empire Tree Improvement Cooperative, 18 Feb. 1989, Post Falls, ID.

similar genotypes recur at different elevations across the landscape. By using the models of genetic variation to construct seed zones on the basis of genetic similarity, the total number of zones could be reduced from 30 to 17. To be sure, the model would be required for determining to which group belong the trees growing at a particular elevation in a particular drainage.

The models are also useful in assessing the effects of environmental change on the adaptedness of populations. By knowing (*i*) the rates of genetic change along an elevational gradient and (*ii*) the rates of temperature change along the same gradient (see Baker 1944), one can readily relate genetic change to temperature change. Figure 6A locates populations that are expected to be adapted at an elevationally low (650 m) site in the Spokane drainage in the contemporary environment. If, however, the frost-free period increases by 20 days (mean temperature increases about 1.5°C), adapted genotypes would be available at only the lowest elevations (400 to 600 m) from a small geographic area surrounding the site (Fig. 6B). Quite similarly, if the frost-free period decreases by 20 days, the distribution of genotypes expected to be suited to a site at 1850 m in the Boise drainage would shift toward higher elevations and toward the southeast (Figs. 6C and 6D).

These predictions suggest that a relatively small climatic change will result in a wholesale microevolutionary reassorting of ponderosa pine populations. The predictions also suggest that a changing climate could threaten the existence of genotypes in peripheral populations. The degree of threat, however, depends on the relative balance between migration (seed transfer) and microevolution in achieving adaptedness. This balance ultimately depends on the rates of environmental change. Note, however, that the validity of the predictions requires that (*i*) ponderosa pine still is suited ecologically to the site after the climate change and (*ii*) intercorrelations among climatic variables remain the same.

Models of genetic variation are particularly suited to guiding programs in gene conservation. For example, a model might be used to identify (*i*) populations such as those in the upper Salmon River system that may be relatively unique genetically (Fig. 7) or (*ii*) genotypes that appear to be threatened because of a changing climate (Figs. 6B and 6D). Models also would be ideal for demonstrating that prescriptions, treatments, and programs are consistent with the natural system of genetic variability and, therefore, are promoting the diversity of the gene resource.

And finally, models of genetic variation can contribute to an understanding of phenotypic variation. Phenotypic variation is determined by the variance in genetic effects, variance in environmental effects, and covariance between the genetic and environmental effects. Because the covariance reflects the degree by which similar genotypes recur in similar environments, the components of phenotypic variance ordinarily are confounded. Monserud and Rehfeldt (1990) used a model of genetic variation in *P. menziesii* var. *glauca* to extract the components of phenotypic variation in site index. These authors correlated ($R^2 = 0.42$) observed site index with the 3-year height predicted by a model of genetic variation. A subsequent path analysis implied that genetic effects were one-third more important than environmental effects (topographic, physiographic, phytosociologic, and soil physical and chemical variables)

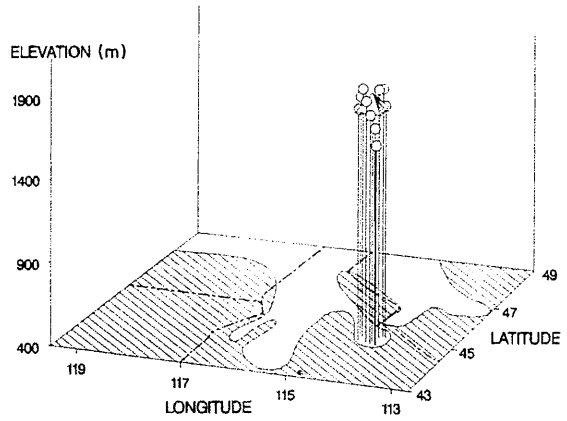

FIG. 7. Using models to locate populations that are potentially unique genetically. The pyramid locates a site at 1850 m in the upper Salmon River system, and balloons locate populations expected to be compatible in the contemporary climate.

in determining site index in that species. Although similar calculations cannot yet be made for ponderosa pine, the results suggest that models can be extremely useful for interpreting phenotypic variation and thus broadening our understanding of basic forest biology.

The degree of usefulness of a model, however, depends on its credibility. A first step is verification; and even though an exercise in validating the present model has been supportive, verification has only begun. A second step involves recognizing that genetic variation in additional variables might be occurring along clines that are independent of those already detected. This eventuality would mean that the recurrence of similar populations across the landscape is more restricted than implied. Functional models, therefore, are dynamic and require updating as additional mathematical descriptors become available. And finally, credibility requires that the appropriate species are ecologically suited to those elevations and localities for which predictions are made. Thus, data bases must be firmly coordinated with the ecological distribution of the species. As credibility increases, however, models that describe patterns of genetic variation across the landscape become directly applicable to a variety of programs dealing with the system of genetic variability.

BAKER, F.S. 1944. Mountain climates of the western United States. Ecol. Monogr. **14**: 223–254.

CAMPBELL, R.K. 1974*a*. A provenance-transfer model for boreal regions. Medd. Nor. Inst. Skogforsk. **31**: 545–564.

_____ 1974*b*. Use of phenology for examining provenance transfers in reforestation of Douglas-fir. J. Appl. Ecol. **11**: 1069–1080.

_____ 1979. Genecology of Douglas-fir in a watershed in the Oregon Cascades. Ecology, **60**: 1036–1050.

_____ 1986. Mapped genetic variation of Douglas-fir to guide seed transfer in southwest Oregon. Silvae Genet. **35**: 85–96.

CAMPBELL, R.K., PAWUK, W.A., and HARRIS, A.S. 1989. Microgeographic genetic variation of Sitka spruce in southeastern Alaska. Can. J. For. Res. **19**: 1004–1013.

CONKLE, M.T., and CRITCHFIELD, W.B. 1988. Genetic variation and hybridization of ponderosa pine. *In* Ponderosa pine: the species and its management. Washington State University, Cooperative Extension, Pullman.

DRAPER, N.R., and SMITH, H. 1981. Applied regression analysis. John Wiley & Sons, New York.

ERIKSSON, G.S., ANDERSSON, S., EICHE, V., IFVER, J., and PERSSON, A. 1980. Severity index and transfer effects on survival and volume production of *Pinus sylvestris* in northern Sweden. Stud. For. Suec. 156.

HOFF, R.J. 1986. Susceptibility of ponderosa pine to the needle cast fungus *Lophodermium baculiferum*. USDA For. Serv. Res. Pap. INT-386.

_____ 1989. Differential susceptibility of ponderosa pine to the gouty pitch midge (*Cecidomyia piniinopis*). USDA For. Serv. Res. Pap. INT-399.

JOHNSON, R.A., and WICHERN, D.W. 1982. Applied multivariate statistics. Prentice-Hall, Englewood Cliffs, NJ.

KUNG, F.H., and CLAUSEN, K.E. 1984. Graphic solution in relating seed sources and planting sites for white ash plantations. Silvae Genet. **33**: 46–53.

LINDGREN, D., KRUTZCH, P., TWETMAN, J., and KIELLANDER, C.L. 1976. Survival and early growth of *Pinus contorta* in northern Sweden. Department of Forestry, Royal College of Forestry, Stockholm, Sweden. Gen. Res. Note 20.

LITTLE, E.J. 1971. Atlas of United States trees. Vol. 1. Conifers and important hardwoods. Misc. Publ. U.S. Dep. Agric. No. 1146.

MONSERUD, R.A., and REHFELDT, G.E. 1990. Genetic and environmental components of site index in Inland Douglas-fir. For. Sci. **36**: 1–9.

MORGENSTERN, E.K. 1978. Range-wide genetic variation of black spruce. Can. J. For. Res. **8**: 463–473.

MORGENSTERN, E.K., and ROCHE, L. 1969. Using concepts of selection to delimit seed zones. *In* Second World Consultation on Forest Tree Breeding. FAO-FO-FTB-69-2/16.

RAYMOND, C.A., and LINDGREN, D. 1989. A model of genetic flexibility. *In* Provenances and Forest Tree Breeding for High Latitudes, Frans Kempe Symposium, 10–11 June 1986, Umeå, Sweden. *Edited by* D. Lindgren. Swedish University of Agricultural Sciences, Umeå. Rep. 6. pp. 159–178.

REHFELDT, G.E. 1984. Veracity of short-term testing. *In* Progeny Testing, Proceedings of Servicewide Genetics Workshop, 5–9 Dec. 1983, Charleston, SC. U.S. Department of Agriculture, Washington, DC. pp. 118–148.

_____ 1986*a*. Adaptive variation in *Pinus ponderosa* from Intermountain regions. I. Snake and Salmon River Basins. For. Sci. **32**: 79–92.

_____ 1986*b*. Adaptive variation in *Pinus ponderosa* from Intermountain regions. II. Middle Columbia River Systems. USDA For. Serv. Res. Pap. INT-373.

_____ 1986*c*. Development and verification of models of freezing tolerance for Douglas-fir populations in the Inland Northwest. USDA For. Serv. Res. Pap. INT-369.

_____ 1988. Ecological genetics of *Pinus contorta* from the Rocky Mountains (USA): a synthesis. Silvae Genet. **37**: 131–135.

_____ 1989. Ecological adaptations in Douglas-fir (*Pseudotsuga menziesii* var. *glauca*): a synthesis. For. Ecol. Manage. **28**: 203–215.

_____ 1990. Genetic differentiation among populations of *Pinus ponderosa* from the upper Colorado River Basin. Bot. Gaz. **151**: 125–137.

ROBERDS, J.H., and NAMKOONG, G. 1989. Population selection to maximize value in an environmental gradient. Theor. Appl. Genet. **77**: 128–134.

ROBERDS, J.H., HYUN, J.O., NAMKOONG, G., and RINK, G. 1990. Height response functions for white ash provenances grown at different latitudes. Silvae Genet. **39**: 121–129.

SAS INSTITUTE INC. 1982. SAS user's guide. Vol. 2. SAS Institute, Inc., Cary, NC.

STEEL, R.G.D., and TORRIE, J.H. 1960. Principles and procedures of statistics. McGraw-Hill, New York.

UNITED STATES DEPARTMENT OF COMMERCE. 1968. Climatic atlas of the United States. U.S. Department of Commerce, Environmental Data Service, Washington, DC.

YING, C.C., THOMPSON, C., and HERRING, L. 1989. Geographic variation, nursery effects, and early selection in lodgepole pine. Can. J. For. Res. **19**: 832–841.