value that they use. For R-N, an expected value is maximized, while for R-L a safe minimum solution is derived. For any particular site, the R-L solution would guarantee a minimum relative yield, while R-N would not necessarily do so. If it is important to ensure minimum yield in every plantation, then R-L uses the better criterion of goodness. For the expected value of a regional breeding program, the R-N solution would be maximum while R-L would not necessarily be so. If average regional maximization is most important, then R-N uses the more appropriate criterion of goodness to maximize. For a predetermined number of source populations, R-N can provide an optimum solution while R-L cannot always guarantee that its minimum yield can be guarenteed. Therefore, the solution for the optimum number of source populations will not in general be the same. For a given number of populations, R-L can determine the minimum site yield that would result which might be less than desired, but they could derive minimum yield as a function of population number to guide policy. In contrast, R-N can determine the expected value of the set of populations for those population numbers, and can prescribe breeding zones as a function of population number. For R-N, the frequency distribution of sites is critical information, and even if the distribution is uniform and has strict limits, the optimization procedure will maximze the expected value. For R-L, the frequency distribution is irrelevant, since optimization is derived on a site-by-site basis.

We therefore believe that the breeder has a choice to make among several approaches to choosing source populations and demarcating their planting site zones. In the accompanying two papers, different views are offered and it is hoped that, by contrasting them, the relative merits of all approaches can be more widely discussed and that sharper debates on breeding philosophy and testing techniques will ensue.

Literature Cited

FINLAY, K. W. and WILKINSON, G. N.: The analysis of adaptation in a plant-breeding programme. Aust. J. Agric. Res. 14: 742—754 (1963). —— FREEMAN, G. H. and PERKINS, J. M.: Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. Heredity 27: 15—23 (1971). —— GREGORIUS, H.-R. and NAMKOONG, G.: Resolving the dilemmas of interaction, separability and additivity. Math. Biosci. 85: 51—69 (1987). —— NAMKOONG G.: The non-optimality of local races. In: Proc. 10th Southern Conf. on Forest Tree Breeding. Texas Forest Service, Texas A and M University Press: 149—153 (1969). —— NAMKOONG, G.: The influence of composite traits on genotype by environment relations. Theor. Appl. Genet. 71: 536—544 (1985). —— PERKINS, J. M. and JINKS, J. L.: Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity 23: 339—356 (1968). —— REHFELDT, J.: Seed transfer in the northern Rockies. In: Progeny Testing. Proc. of servicewide genetics workshop. (Charleston, South Carolina, December 5 to 9, 1983.) USDA Forest Service, TM Publ.: 34—60 (1984). —— ROBERDS, J. H. and NAMKOONG, G.: Population selection to maximize value in an environmental gradient. Theor. Appl. Genet. 77: 128—134 (1989). —— YEH, F. C. and HEAMAN, J. C. H.: Estimating genetic parameters of height growth in seven-year old coastal Douglas-fir from disconnected diallels. For. Sci. 33: 946—957 (1987).

Genetic Flexibility – A Model for Determining the Range of Suitable Environments for a Seed Source

By C. A. RAYMOND*) and D. LINDGREN**)

Abstract

A mathematical model describing the relative performance of different seed sources over a range of environments has been developed and tested using data from Scots pine in Sweden and white ash in the USA. The components of the model have a direct biological interpretation in terms of the fundamental nature of the problem : i.e. the maximal performance of a seed source, where this maximum will occur and the range of suitable sites for such a seed source. Comparisons can thus be made between seed sources as to their site preferences and the relative flexibility of each seed source calculated. Decisions regarding the optimal deployment of seed may thus be simplified and such information used to refine the allocation of breeding zones.

The model requires that a one-dimensional environmental gradient be defined. Such gradients may relate to a single overriding environmental factor (e.g. winter cold) which can be described by geographical information such as latitude, altitude and distance from the sea. Definition of this gradient is a necessary prerequisite to application of this model. Methods for establishing such gradients are discussed.

Key words: Breeding zones, Pinus sylvestris, genotype by environment interaction.

Introduction

Interactions between genotypes and environments are a potential problem in forestry in that they can lead to the choice of a 'less-than-most-productive' seed source for a particular site (MATHESON and RAYMOND, 1986). For species which exhibit clinal variation over a wide range of environments this problem is intensified, especially if variation patterns closely match environmental changes. The presence of genotype by environment (G × E) interactions serve to complicate decisions regarding establishment of breeding populations and design of breeding strategy. Once it has been established that such interactions exist, it is essential to determine whether they are of practical importance for a breeding program. Finding a significant GxE interaction does not necessarily imply that it will have a major impact on future breeding strategy (MATHESON and RAYMOND, 1984, 1986).

If such interactions are found to be real and important, there are two approaches available; either to classify and

*) CSIRO, Division of Forestry and Forest Products, "Stowell", Stowell Avenue, Hobart 7004, Australia

**) Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, S-90183 Umeå, Sweden

group environments or to group genotypes. The first-alternative aims to determine groups of environments within which GxE interactions approach zero (CAMPBELL and SORENSEN, 1978) and thus maximise accuracy of selection and precision of estimates of genetic parameters. For the second alternative, genotypes may be classified as stable or unstable (according to FINLAY and WILKINSON, 1963) and unstable ones may be removed from the breeding population, thus greatly reducing GxE interactions (MATHESON and RAYMOND, 1984). Genotypes may be grouped according to their relative performance across a range of tested environments.

To overcome the problem of GxE interactions breeding zones may be established within the natural range of clinal species and guidelines may be laid down for acceptable movement or transfer of seed. Such zonation aims to maximise yield throughout the range of available plantation sites whilst minimising the risk of using maladapted seed sources (REHFELDT, 1982a, b). Decisions regarding allocation of breeding zones are generally based on genecological or GxE studies which combine both classification of environments and of genotypes. The observed performance of many seed sources, at one or a few sites, is regressed against a set of environmental variables in an attempt to explain the causes of the patterns of variation.

Such studies have a long history, with early work being done in Scandinavia by LANGLET (1936). Many other studies have followed, including those of EICHE (1966), ERIKSSON (1975), REMRÖD (1976), CAMPBELL (1979), CAMPBELL and FRANKLIN (1981) and REHFELDT (1979, 1982a, b, 1983). Each of these studies aimed to determine which environmental factors (usually in terms of latitude, altitude and distance from the sea) were important in defining variation patterns and thus to establish a basis for defining breeding zones or transfer rules for seed. Several studies have produced models which attempted to examine this variation (e.g. MORGENSTERN and ROCHE, 1969; CAMPBELL, 1974; ERIKSSON et al., 1980; ROBERDS and NAMKOONG, 1986) or have applied various multidimensional plotting routines to study the variation patterns (e.g. KUNG, 1981; REHFELDT, 1982a; SILEN and MANDEL, 1983; KUNG and CLAUSEN, 1984; PRESCHER, 1986).

Such studies provide an important base of knowledge regarding patterns of both environmental and genetic variation. This information may be used to establish and define environmental gradients which have direct effects on how provenances will perform on a variety of sites. Such gradients are usually defined in terms of geographic factors (latitude, longitude, altitude, distance from the sea) which describe or reflect some major or overriding environmental factor such as winter cold or length of the growing season. Definition of such gradients is an important first step in examining how to best use the available genetic material.

The problem then becomes one of best matching genotypes to environments in order to maximise yields. Given a known environmental gradient, it is necessary to establish the number of seed sources required to adequately cover the range of sites and then to define where to stop planting one source and begin planting another (ROBERDS and NAMKOONG, 1986). In addition, with modern silvicultural practices and artificial regeneration of large areas, it is recognised that the local seed sources may not necessarily be optimal in terms of the current economic objectives of forestry and that transfer of seed may be ad-

vantageous (NAMKOONG, 1969; ERIKSSON et al., 1980; PRESCHER, 1986).

Given that such a situation exists, the concept of genetic flexibility of a seedlot becomes important. That is, over what range of sites should the one seedlot be utilised?

MANGOLD and LIBBY (1978) present a conceptual model for such flexibility which extends the concepts of NAMKOONG (1969) in regard to the diminishing importance of reproductive fitness when artificial regeneration is used. The model is based on the idea that each seedlot (or population) has an optimal environment for expression of vegetative vigour and that this optimum may differ from its origin. If a seedlot is planted across a range of sites around this optimum, its performance will change in a curvilinear way in relation to the distance from the optimum (see Figure 1). If many seedlots are planted across an environmental gradient a series of such curves should exist, arranged in an overlapping manner, as each seedlot finds its optimum. If such a series of curves exist in reality they would provide an objective basis for examining how to optimise the use of the available material by simplifying decisions regarding the delimiting of seed zones and the refining of transfer rules. RAYMOND and LINDGREN (1986) present such a series of curves for the change of performance, across a range of sites, for some provenances of Scots pine (Pinus sylvestris) in Sweden. Another series of curves is given by KUNG (1981) for the performance of different stands of white ash (Fraxinus americana) at 4 sites covering a latitudinal transect from Wisconsin to Louisiana in the USA.

Using the Swedish data RAYMOND and LINDGREN (1986) developed and tested a mathematical model which characterised the behaviour of each provenance by examining the form of these curves. The components of this model have a direct biological interpretation in terms of the fundamental aspects of the problem: what is the maximal performance of a seedlot (in terms of survival and growth), where does this maximum occur and over what range of sites should the seedlot be used? The model is equally applicable for examining the performance of a range of seedlots at any given site (i.e. describing the origin of the optimal seedlot and the relative harshness of the site). This paper presents the model and gives examples of its application, both in terms of examining the behaviour of a seedlot over a range of sites (using the Swedish data) and for examining the behaviour of a range seedlots at a given site (using the data of KUNG, 1981).
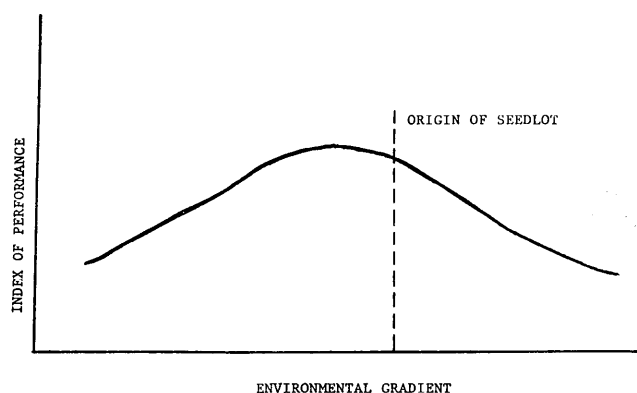

Figure 1. — Hypothetical change in performance of a seedlot when planted across an environmental gradient.

Mathematical Model for Flexibility

Figure 2 presents the performance of six provenances of Scots pine (*Pinus sylvestris*) across a range of sites in Sweden. Details of the provenance origins and of the scales used on the axes are given in RAYMOND and LINDGREN (1986) and will be outlined below. The figure indicates that these provenances behave in a curvilinear manner with the performance of each provenance dropping off in a parabolic form with increasing distance from the optimal site. From an analytical point of view, the initial approach to fitting curves is to use a polynomial. On first inspection, the shape of these curves would suggest fitting a quadratic polynomial of the form:

$$Y = \alpha + \beta X + \gamma X^2 \qquad (1)$$

where

Y = an index of the performance of each provenance at each site

X = an index reflecting the relative severity of each site

$\alpha, \beta \,\&\, \gamma$ = appropriate regression coefficients.

Such a function was fitted by KUNG (1981) to the white ash data which is used as an example here.

However, the expected form of these curves may be anticipated by applying some biological information to this relationship. Several characteristics were considered desirable in the fitted function:

1. Parameters should be interpretable in a biologically meaningful way.

2. The maximum should be relatively smooth.

3. The function should asymptotically approach zero for milder climates; i.e. for milder climates survival should not be greatly affected, although growth may be reduced. Total production should not reach zero.

4. For harsher climates, the function should approach zero asymptotically or linearly but without too steep a slope; i.e. both survival and growth will be affected and for harsh sites all trees may die, leading to zero production.

5. The appearance of the curve should be more sensitive to values close to the optimum and less sensitive to lower values (which result from maladapted provenances of little practical interest).
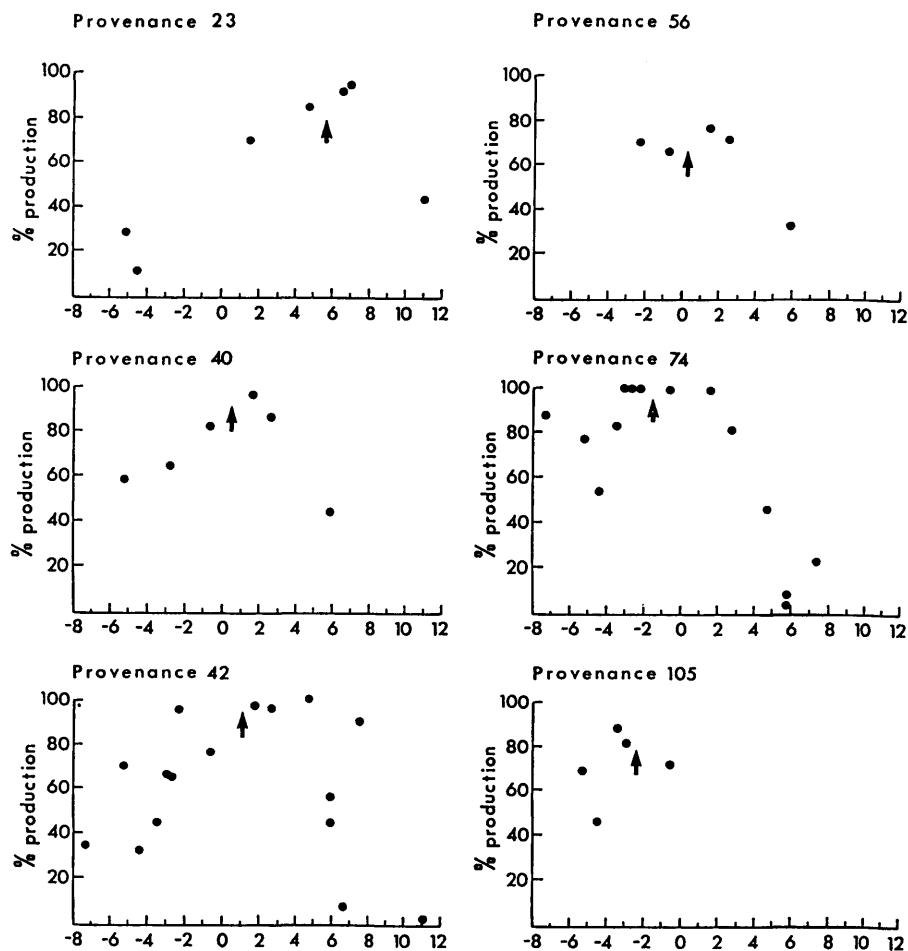
↟ Calculated optimum site value

Figure 2. — Relative performance (in terms of volume per hectare in 1981) of the six Swedish provenances of Scots pine.

Under these constraints the following alternative function was suggested:

$$Y = a/[1 + ((X - b)/c)^2] \quad (2)$$

where Y and X are as above, but here the coefficients have a direct biological interpretation (*Figure 1*):

a = parameter representing maximal performance (height of the curve)

b = parameter locating the optimum (i.e. optimal site for a provenance)

c = parameter describing the width of the curve (i.e. between (b — c) and (b + c), Y > 0.5a).

The interpretation of the parameters of this function requires further explanation. The function may be used in two different ways. Firstly, it may examine the performance of each provenance across a range of sites. The second option is to examine all provenances planted at a single site. The interpretation of the parameters are discussed separately for each of these options.

1.) Provenance behaviour across many sites

The Y values represent an index of performance for each provenance at each site. To examine the behaviour of provenances in this manner it is necessary to combine data across sites. As differences in sites are a major source of variation in provenance trials the data for each site was transformed to a relative performance. A maximal value of 100 was assigned to represent the most productive (well adapted) provenance for each site and the performance of all other provenances was calculated relative to this maximum. This scaling system was chosen to make sites as comparable as possible and also to stabilise the variances across sites.

An index for the relative severity of each site (X value) is also required. Such an index should be an one-dimensional gradient covering the range of interest. Definition of this gradient is vital for this type of study and relies on genecological or provenance studies relating variation patterns to environmental factors. Such gradients are generally defined using multiple regression techniques and regressing environmental variables (e.g. latitude, altitude and distance from the sea) on seedlot performance (CAMPBELL, 1979; REHFELDT, 1979, 1982a, b). Defining such gradients relies on the presence of a single major or overriding environmental factor (e.g. winter cold) which is closely associated with geographic variables (e.g. latitude and altitude). If such an overriding factor is not present, the definition of such a gradient may not be possible.

The a, b and c parameters relate to the performance of each provenance as follows. The a value represents the maximal relative performance for the provenance (i.e. the height of the apex of the curve). As the Y axis is scaled against the best-performing provenances at each site the a value indicates the merit of this provenance when compared to others tested across the range of sites studied. An a value close to 100 indicates that the provenance is performing amongst the best of those tested at the range of sites evaluated and thus deserves further attention. Conversely, provenances which do not perform amongst the best at any of the sites may also be identified.

The b values represent the location of the apex of the curve on the X axis, or the site location description for where this provenance should perform at its best. The optimal site may be identified even though the provenance may not have been planted at such a site.

The width of the curve is described by the c values which represent the relative flexibility of the provenance. A low c value indicates that the provenance will only perform well over a very limited range of sites or that it is finely adapted to conditions at its original location. A larger c value reflects a more flexible provenance whose performance is less sensitive to environmental changes.

2.) Many provenances at a single site

This function can equally well be applied to examining the behaviour of various seedlots at a single site. For such a study the interpretation of the components of the function differs slightly. As each site is examined separately, the scale for the Y values may vary: the actual scale of measurement may be used or the data may be transformed to a relative scale as used above. The X values would correspond to an index reflecting the severity of the site of origin for each seedlot. The a values will reflect the scale used for the Y axis and should approximate the maximum for that site. The b values represent the optimal seedlot origin for that site, whilst the c values indicate the discriminating ability of the site. Low c values reflect a more demanding site where discrimination of seedlots is strong and efficient; whereas a higher c value indicates that most of the seedlots tested have performed equally well.

As examples of how this function may be used, it is applied to the data from Swedish Scots pine provenances presented in *Figure 2* and to data for each site from KUNG (1981).

Materials and Methods

1.) Swedish provenance data

Data from a Swedish experiment with Scots pine is used to illustrate how the function may be used to examine the performance of a provenance which has been planted at a range of sites across the environmental gradient. Details of the Swedish Scots pine data were presented in RAYMOND and LINDGREN (1986). Thus only an outline is given here. The six provenances form part of the "EICHE series" of trials which is a nationwide series planted 1952 to 1955 to evaluate the effects of transferring provenances (see EICHE, 1966, for details of the design). Data for total volume production (on a per hectare basis) until and including 1981 were extracted for all sites for which complete information was available. Eighteen sites were included with 75 provenances represented, but with the distribution of provenances across sites being extremely unbalanced. The six provenances used here were selected on the basis that they appeared at five or more sites and cover the majority of the range of the provenances sampled in the trial series. Details of the origin of these provenances and of the number of trial sites where they appear are given in *Table 1*.

As it was necessary to combine data across sites, the total volume production per hectare was transformed to a relative performance as described above. The indices for the relative severity of each site was based on previous work done in Sweden which aimed to quantify site severity in terms of latitude and altitude for both plantation sites and seed sources. The indices used here are those officially approved by the Swedish Board of Forestry in the Forestry Act (SKOGSSTYRELSEN, 1979) and were based on a combination of ideas and work of ERIKSSON et al. (1980) and REMRÖD (1976):

Table 1. — Descriptive information for the Swedish Scots pine provenances.

Provenance No.	Name	No. of sites	Latitude	Longitude	Altitude	Index for origin
23	Malå, Stromfors	7	65°08'	18°53'	305	6.03
40	Bråmön, Sanna	6	62°12'	17°42'	5	2.10
42	Sveg, Malmbäcken	16	62°03'	14°19'	385	2.95
56	Vinje, Prästgården	5	59°35'	7°50'	550	1.15
74	Eckersholm	14	57°36'	14°13'	225	-1.83
105	Kungsbacka	5	57°30'	11°55'	10	-2.60

Site Severity Index = Lat − 74.2609 + Alt/300 +

0.000006875·Alt² + Lat·Alt/40768 + Lat²/282.82

where Lat = Latitude in degrees, converted to decimal values

where Alt = Altitude as metres above sea level.

Provenance Origin Severity Index = Lat − 60.1 + Alt/300

The scale for the site severity values was defined based on latitudes north of 61°, thus negative values may appear (as in *Figure 2*). Although these indices form part of a legal document they are biologically sound and their use for predicting optimal provenance transfers, based on the present set of data, was shown in a separate analysis (LINDGREN and RAYMOND, 1987).

2.) White ash data

The other set of data examined is that of KUNG (1981) which is used to illustrate the fitting of the function to performance of many provenances at a single site. Height and survival at 3 years of age were recorded for 4 plantation sites covering a north-south transect across the central USA (*Table 2*). Nineteen provenances covering this same range of latitudes were planted at all 4 sites. Only 2 of the sites (Illinois and Ohio) had provenances transferred from both north and south. The remaining 2 sites (Louisiana and Wisconsin) were at the extremes of the range of provenances tested and thus only had northward or southward transfers represented. KUNG presents a series of curves relating both height and survival for all provenances at each site to their respective latitudes of origin. For the most southern site (Louisiana) the curve relating height growth to provenance had no maximum but instead was inverted with a minimum at about 40°. As this site was at the extreme of the range examined and displayed an unusual curve, it was not included in further analyses.

The height data were transformed in the same manner as that used for the Swedish data. For survival, the scaling system imposed an automatic restriction of 100 thus this data was used without further transformation. The index of environmental severity for the provenance origin was the same as used by KUNG; i.e. latitude of origin.

3.) Analyses

In order to evaluate the merit of the suggested function relative to a quadratic polynomial, both functions 1 and 2 were fitted to both sets of data. All analyses were done using an iterative least-squares fitting procedure (Procedure NLIN in SAS; SAS, 1982). In contrast to the situation for linear modelling, there are no explicit guarantees of fit for nonlinear modelling techniques. The adequacy of a model is usually assessed by the occurrence of convergence and, for predictive models, by the minimising of the sum of the squared residuals or R^2 values (CHAMBERS, 1973).

Table 2. — Descriptive information for the white ash sites.

Site	Latitude	Longitude
Louisiana	30.4°	92.0°
Illinois	37.5°	89.3°
Ohio	40.0°	82.0°
Wisconsin	46.6°	89.5°

The acceptance of a nonlinear model is generally based on theoretical or empirical evidence plus the evaluation of the model over many data sets (RATKOWSKY, 1983). For these data R^2 was calculated and the distribution of residuals examined for each analysis. As the curve fitting was done by least squares, the R^2 values do not necessarily indicate the potentially best-fit curve in both dimensions, but serve as an indication of the value of the function tested.

Results

1.) Functions 1 and 2

Table 3 presents results for fitting both functions to the Swedish Scots pine data and similar results for the white ash data are given in *Table 4*. Both functions removed over 90% of the variation from all data sets but, in general, function 2 resulted in a slightly better fit. Examination of the residuals resulting from fitting both functions revealed an essentially random pattern. Both functions have three degrees of freedom. As function 2 gave an equal or better fit to the data, the curves have a more reasonable appearance in a biological sense (i.e. the tails will asymptote instead of going to zero) and the parameters of this function have a biological interpretation, it seems reasonable to accept it as a valid model. The interpretation of the components of this function deserve further attention.

2.) Swedish Scots pine data

This example presents the case where the function is fitted to a provenance's performance across many sites. *Table 3* indicates that the a, b and c values differed for each provenance. The a values may be used to identify which provenances perform well over the range of sites. For four provenances (23, 40, 42 and 74) these a values were close to 100 indicating that they were performing well. In contrast, for provenances 56 and 105, the a values were lower, indicating that they will not perform amongst the best anywhere over the range of test sites.

The b values allow for the identification of an optimum, even when this optimum is not actually represented in the field. The optimal site indices for each provenance (as given by the b values) follow a similar pattern to the indices for the origin of each provenance (correlation 0.968 and significant $p < 0.05$) indicating that the optimal site

Table 3. — Results from fitting the two functions to the Swedish provenances.

Provenance

	23	40	42	56	74	105
Index for Origin	6.03	2.10	2.95	1.15	-1.83	-2.60
Function 1:$\alpha+\beta X+\gamma X^2$						
α	73.5	89.0	83.6	75.7	89.3	66.0
β	6.86	0.45	1.68	1.19	-4.46	12.97
γ	-0.79	-1.24	-0.83	-1.42	-1.00	-2.76
$R^2(\%)$	97.6	98.7	89.4	99.5	95.0	97.1
Function 2:$a/(1+((X-b)/c)^2)$						
a	95.5	93.2	101.6	78.1	104.7	85.2
b	5.47	0.58	1.15	0.40	-1.50	-2.34
c	5.65	6.02	5.57	5.46	5.35	4.04
$R^2(\%)$	99.0	98.9	90.1	99.0	94.0	97.4

Table 4. — Results from fitting the two functions to the white ash sites.

Site

	Illinois		Ohio		Wisconsin	
	Height	Survival	Height	Survival	Height	Survival
Latitude of site	37.5°		40.0°		46.6°	
Function 1:$\alpha+\beta X+\gamma X^2$						
α	78.3	-235.3	-319.6	-506.2	-275.8	-735.9
β	3.37	16.34	20.30	29.38	16.26	39.24
γ	-0.09	-0.20	-0.25	-0.37	-0.18	-0.47
$R^2(\%)$	97.7	99.4	98.5	96.4	97.9	97.7
Function 2:$a/(1+((X-b)/c)^2)$						
a	92.3	99.3	86.9	85.8	84.6	89.8
b	32.01	40.89	40.02	40.17	43.16	41.56
c	10.66	19.83	15.07	11.00	15.44	8.19
$R^2(\%)$	98.0	99.4	98.6	96.6	97.8	97.4
Estimated optimal provenance origin from KUNG (1981)	18°	41°	40°	40°	44°	42°

is close to the provenance origin. The b values were also different for each provenance indicating that their site preferences differed.

The width of the curves are described by the c values. For each provenance these values give an idea of the relative flexibility or genetic plasticity relative to the index of environmental severity. For the data examined here the c values proved to be remarkably stable, given the diversity of the provenance origins.

3.) White ash data

The performance of many provenances at a single site is examined here. Table 4 presents results for fitting both functions to the 3 northern sites of KUNG (1981). KUNG also fitted function 1 to both survival and untransformed height. The coefficients for survival are similar to those of KUNG, but as height was transformed in this study these coefficients differ. However, a comparison of results is valid, especially in regard to assessing the origin of the optimal provenance for each site.

The a values resulting from fitting function 2 to each site were all above 85%. The b values follow the same pattern as the latitude of the sites, indicating that each site requires a different provenance for maximal production. The agreement between the b values and the estimated origin of the optimal provenance (as defined by KUNG's algorithm) is excellent for survival at all sites and for height at the Ohio and Wisconsin sites. For the Illinois site the estimated optimum, using the algorithm applied by KUNG, was 18° north, which is outside the natural range

of the species. In contrast, the estimated b value from the current analysis (32⁰) is within the range of the species.

When the function is fitted to data for a range of provenances at a particular site the c values may be interpreted as describing the discriminating ability of the site, with lower c values reflecting more demanding sites. For the two northern sites (Ohio and Wisconsin) the c values for survival were lower than those for height, whereas for the Illinois site the reverse was true. For height growth the c values appear unrelated to latitude of the site, whilst for survival there is a negative relationship, indicating that the range of suitable provenances decreases with increasing latitude.

Potential uses for the function

By using the suggested function the shape of the curve for performance of a provenance across a range of sites can be examined mathematically as well as visually. Once the a, b and c values are known the function may be solved for any level of relative performance (i.e. any percentage of Y) and thus the range of suitable sites for each provenance determined. If seed zones are to be allocated based on a predetermined level of acceptable performance (such as 90% or 95% of the maximum as defined by the a values from function 2) then this function is of great value. By solving the equations for each provenance for a range of acceptable levels of performance the relative widths of the range of suitable sites may be examined. *Figure 3* presents results for the Swedish Scots pine provenances. For all 6 provenances the width of the ranges are similar, reflecting their similar c values. On average, moving one of these provenances by 1.23 units of site index (or 2.46/2 from *Figure 3*) in either direction from its optimum would result in a loss of 5% at the extremes of the range.

Such diagrams provide a sound basis examining the behaviour of many provenances relative to their maximum. Decisions about the optimal deployment of seedlots may thus be simplified and such questions as the tradeoff between the number of seed zones and the loss in performance can be further examined.
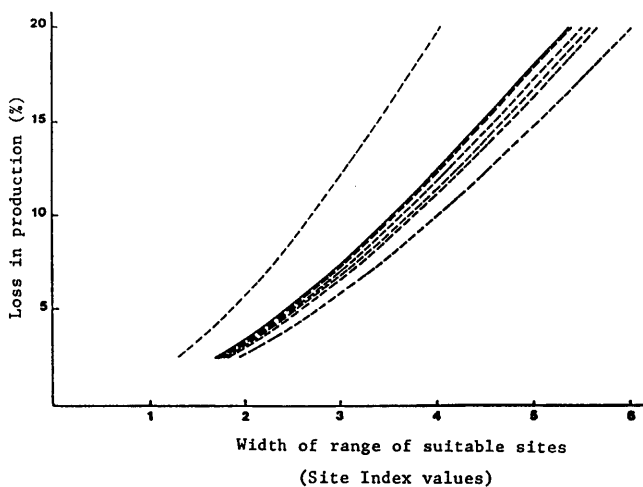

Figure 3. — Relationship between the relative loss in production (at the extremes of the planting range) and the width of the range of planting sites for the six Swedish provenances of Scots pine. Average loss is shown by the solid line.

Discussion

The need to develop breeding zones has long been recognised (CAMPBELL and SORENSEN, 1978). However, the delineation of such zones has not always been based on biological information. Factors such as political and managerial considerations together with economic factors including costs and administration of seed collection, nursery operations and area of established plantations have generally been combined with biological intuition to design breeding zones (REHFELDT, 1979; ROBERTS and NAMKOONG, 1986). Several approaches have been developed to produce models and assist in defining zones based on the biology of the species involved.

When developing such models it is essential that a large amount of variation exists between the seed sources tested (PRESCHER, 1986) and that the environmental gradient may be defined. As the aim of allocating seed zones is to minimise GxE interactions it is assumed that such interactions are present and economically important. An additional problem arises in that not all traits in a population will follow an identical cline (CAMPBELL and SORENSEN, 1978). Thus a decision must be made regarding which trait or traits are to be used as the index of growth and adaptation. The optimum environment for one trait may differ to that for other traits (KUNG and CLAUSEN, 1984) or the ranking of traits may change within the species distribution. For example, in northern Sweden, due to the harshness of the environment, the most economically important trait for Scots pine is survival (PRESCHER, 1986). Further south, where the environment is milder, the variation in survival is smaller. Growth rate becomes more important. Thus, it may be necessary to use different bases for defining breeding zones in different parts of the species range or, alternatively, to combine survival and growth rate into a single index (e.g. total volume production per hectare as used for the Scots pine data presented here).

Most studies of seed source variation have been based on seed, collected from many widely scattered stands throughout the natural distribution of the species, which is grown a one or more locations (KUNG and CLAUSEN, 1984). Results of such studies are then either related back to factors in the environment of origin of each seed source (e.g. LANGLET, 1936; CAMPBELL, 1979) or a genotype-environment interaction study is undertaken (e.g. REHFELDT, 1979, 1982b). The latter type of study aims to establish whether GxE interactions exist and to investigate their possible causes. Much statistical theory exists for such interactions (see reviews by FREEMAN, 1973 and HILL, 1975) and many such interactions have been found in forestry (see review by MATHESON and RAYMOND, 1986).

How these two types of studies interrelate is not clear. REHFELDT (1979, 1982a, b) has undertaken both types of studies for Douglas fir (*Pseudotsuga menziesii*) in two areas of the US (north Idaho and north east Washington 1979; western Montana 1982b) and for *Larix occidentalis* (1982a). For Douglas fir in western Montana and for *Larix* GxE interactions were not found for most of the traits studied, and the ranking of the populations tested were essentially the same at all sites. Thus the observed GxE interactions were not large and were mainly due to differences in means or variances. Stepwise multiple regression models including environmental variables were then fitted to the data, and significant proportions of the variation in seedlot performance could be removed by such models. For Douglas fir in Idaho and Washington (1979) the GxE inter-

actions were significant for all traits studied and the data were further examined using a joint regression analysis (PERKINS and JINKS, 1968) where the performance of each seedlot is regressed against the mean value of all seedlots at each environment. The GxE interaction may then be partitioned into variance due to the heterogeneity of regression lines and that due to deviations from the regressions. For this data, most of the interaction was due to the heterogeneity of regressions with regression lines converging within the range of environments tested.

REHFELDT (1979) considered that the mean values for traits were significantly correlated with the regression coeff.cients and that the joint regression analysis masked variation which is necessary for assessing differentiation of populations. This lack of independence of the measure of the environment has been criticised by other authors including FREEMAN and PERKINS (1971), WRIGHT (1971, 1976) and FREEMAN (1973). Alternative methods for specifying the environmental gradient by using other seedlots or additional replications at each test site have been suggested. Both of these involve additional effort in establishing trials and only provide an estimate which is subject to error. Thus, they may not add much information or truly reflect the environmental gradient.

In contrast, multiple regression on independent environmental variables (latitude, longitude, altitude and habitat type) was able to differentiate the seedlots, even though some of these environmental variables were intercorrelated. RHEFELDT (1979) concluded that joint regression analyses should not be used for selection of genotypes but may be used for comparing the response of populations to an environmental gradient.

Given that two very different definitions of environment were used, such a conclusion is questionable. No attempt is made to examine how these definitions relate to each other or which definition is more valid. However, results of the joint regression analysis were used to calculate mean differences among populations which are associated with least significant differences (lsd) at given levels of probability. The multiple regression models are then used to define the appropriate changes in environmental variables required to alter the traits of interest by an amount related to the lsd. The breeding zones are then allocated based on these calculated differences. Whether such zones relate to maximising yield is unclear (ROBERDS and NAMKOONG, 1986) and to allocate zones when there is no strong evidence of GxE interactions (e.g. REHFELDT, 1982a, b) appears questionable.

Both the joint regression and multiple regression analyses have one major problem in relation to attempting to maximise yield over a range of environments. Both analyses assume a linear relationship between genotype performance and changes in the environment. It is impossible to distinguish a sub-optimal environment from a supra-optimal one. Both the sub- and supra-optimal environments for a seedlot will be classed as producing a lower mean performance (i.e. the tails of the curve in *Figure 1* would be classed together). To maximise yield it is necessary to determine the location of the optimum for each seed source and to select seed sources with a high yield over a range of environments. By using the linear regression analyses this is not possible.

Several alternative forms of analyses based on multidimensional curvilinear surfaces have been proposed. The multidimensional plotting routines proposed by KUNG

(1981) and KUNG and CLAUSEN (1984) provide one method of integrating relative seedlot performance with environmental gradients. A multidimensional response surface is fitted to the data and plotted as a shaded contour graph. Such plots illustrate the range of suitable seed sources for a given plantation site but give no indication of the relative flexibility of each seed source and when to change from one seed source to another. However, these plots give some idea as to the relative importance of site variation and seed source variation and an indiction of whether it is beneficial to transfer seed across the environmental gradient. This technique relies on the optimum site-seedlot combination being within the range examined. If this is not the case (e.g. PRESCHER, 1986) then interpretation of the plot is unclear and no recommendations regarding use of available seed are possible.

ROBERDS and NAMKOONG (1986; 1988) present an alternative approach based on curvilinear relationships. Their approach is somewhat similar to that presented in this paper in that they aim to present a method for establishing breeding zones that will maximise yield based on biological information. By developing a response function for seedlot performance across an environmental gradient and using optimisation theory they determine the number of seed sources required for maximising yield and the ranges of environments suitable for each seed source. The response functions are based on assuming Gaussian distributions for seedlot performance, with all seedlots assumed to have equal variance. Such an assumption appears valid, especially given the results for the Scots pine data presented here. Their other main assumption relates to developing a probability density function for the distribution of planting sites. Sites are assumed to be more frequent where the species is well adapted and less frequent where it is poorly adapted (ROBERDS and NAMKOONG, 1988). Based on these assumptions a normal distribution is used. For the Swedish Scots Pine data presented here such an assumed distribution is questionable due to the strict regulations governing forestry. In this case the distribution of planting sites may in fact be uniform (see ROBERDS and NAMKOONG, 1988, for discussion of this point).

This assumption is required (ROBERDS and NAMKOONG, 1986, 1988) as their model is based on using Gaussian distributions whose mathematical properties are well known. In contrast, the function presented in this paper is a Cauchy distribution, which resembles a Gaussian distribution but asymptotes much more slowly. For a Cauchy distribution the curve approaches the axis so slowly that an expectation of the density function does not exist (FELLER, 1966). In addition the characteristic equation of this function is not differentiable. Thus, none of the moments of the Cauchy distribution exist (FISZ, 1963).

While the function presented in this paper is mathematically more difficult to handle (from some points of view), for application to the types of data presented, the infinite variance of a Cauchy distribution is not a serious problem. As the X axis.is an environmental gradient, the distribution of the function is effectively truncated by the extremes of this gradient (i.e. for the Scots pine data the gradient commences at the tree line in the north and ends at the coast in the south). Thus the tails are truncated and only the central part of the distribution is of interest.

The components of the function presented have a clear interpretation in terms of the biological aspects of match-

ing seed sources to environments. The three main factors of defining the maximal performance of each seedlot, the location of this maximum and over how wide a range of environments each seedlot should be used may be examined mathematically and graphically.

Seed sources with high yields over a range of environments are easily detected and decisions regarding when to stop planting one source and begin planting another are simplified. Boundaries for breeding zones may be allocated based on a predetermined level of acceptable performance. Breeding material may be distributed across a defined area, provided that the choice of material is such that the performance does not fall below the preset minimum acceptable level.

The model relies on the accurate specification of the environmental gradient. This component is potentially the most difficult part of the model to specify as the environment is a complex of many factors. It is essential to define which of these factors directly affect seedlot performance and then to combine such factors in a meaningful way to provide a one-dimensional gradient. Although it is desirable to be able to describe environment in this manner it is often difficult in practice (FREEMAN and PERKINS, 1971). Thus, before it is possible to apply such a model, it is necessary to undertake genecological studies so that the effects of environmental factors are understood.

Acknowledgements

References

CAMPBELL, R. K.: A provenance transfer model for boreal regions. Norsk Institutt for Skogsforskning. 31, 10, 544—566 (1974). —— CAMPBELL, R. K.: Genecology of Douglas-fir in a watershed in the Oregon cascades. Ecology 60, 1036—1050 (1979). —— CAMPBELL, R. K. and FRANKLIN, J. F.: A comparison of habitat type and elevation for seed-zone classification of Douglas fir in western Oregon. Forest Science 27, 49—59 (1981). —— CAMPBELL, R. K. and SORENSEN, F. C.: Effect of test environment on expression of clines and on delimination of seed zones in Douglas-Fir. Theoretical and Applied Genetics 51, 233—246 (1978). —— CHAMBERS, J.: Fitting non-linear models: Numerical techniques. Biometrika 60 1—13 (1973). —— EICHE, V.: Cold damage and plant mortality in experimental plantations with Scots pine in Northern Sweden. Studia Forestalia Suecica 36, 1—218 (1966). —— ERIKSSON, G.: Beräkning av hårdhetsindex och dess praktiska applikationer. Skogshögskolan, Institutionen för Skogsgenetik, Rapporter och Uppsatser 17, 29—38 (1975). —— ERIKSSON, G., ANDERSSON, S., EICHE, V., IFVER, J. and PERSSON, A.: Severity index and transfer effects on survival and volume production of *Pinus sylvestris* in Northern Sweden. Studia Forestalia Suecica 156, 1—32 (1980). —— FELLER, W.: An introduction to probability theory and its applications. Volume II. John Wiley and Sons, New York. (1966). —— FINLAY, K. W. and WILKINSON, G. N.: The analysis of adaptation in a plant breeding programme. Australian Journal Agricultural Research 14, 742—54 (1963). —— FISZ, M.: Probability theory and mathematical statistics. John Wiley and Sons, New York (1963). —— FREEMAN, G. H.: Statistical methods for the analysis of genotype-environment interactions. Heredity 31, 339—54 (1973). —— FREEMAN, G. H. and PERKINS, J. M.: Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. Heredity 27, 15—23 (1971). —— HILL, J.: Genotype - environment interactions - a challenge for plant breeding. J. Agric. Sci., Camb. 85, 477—493 (1975). —— KUNG, F. H.: Delineating seed collection zones based on multiplantation provenance tests. Proceedings 16th Southern Forest Tree Improvement Conference May 27—28, 1981. Virginia, USA (1981). —— KUNG, F. H. and CLAUSEN, K. E.: Graphic solution in relating seed sources and planting sites for white ash plantations. Silvae Genetica 33, 46—53 (1984). —— LANGLET, O.: Studien über die physiologische Variabilität der Kiefer und deren Zusammenhang mit dem Klima. Beiträge zur Kenntnis der Ökotypen von *Pinus sylvestris* L. Medd. Statens Skogsförs Anst. 29, 219—470 (1936). —— LINDGREN, D. and RAYMOND, C. A.: Proveniens för flyttning av tall i Sverige. Swedish University of Agricultural Sciences, Department of Forest Genetics and Plant Physiology, Arbetsrapport 22 (1987). —— MANGOLD, R. D. and LIBBY, W. J.: A model for reforestation with optimal and suboptimal tree populations. Silvae Genetica 27, 66—68 (1978). —— MATHESON, A. C. and RAYMOND, CAROLYN A.: The impact of genotype × environment interaction on Australian *Pinus radiata* breeding programs. Australian Forest Research 14, 11—25 (1984). —— MATHESON, A. C. and RAYMOND, CAROLYN A.: A review of provenance × environment interaction: its practical importance and use with particular reference to the tropics. Commonwealth Forestry Review 65, 283—302 (1986). —— MORGENSTERN, E. K. and ROCHE, L.: Using concepts of selection to delimit seed zones. Proceedings Second World Consultation on Forest Tree Breeding August 7—16, 1969. Washington, USA 1, 204—215 (1969). —— NAMKOONG, G.: Nonoptimality of local races. Proceedings 10th Conference Forest Tree Improvement. pp 149—153 (1969). —— PERKINS, J. M. and JINKS, J. L.: Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity 23, 339—356 (1968). —— PRESCHER, F.: Transfer effects on volume production of *Pinus sylvestris* L.: a response surface model. Scandinavian Journal of Forest Research 1, 285—292 (1986). —— RATKOWSKY, D. A.: Nonlinear Regression Modelling. Marcel Dekker Inc., New York and Basel. 276 p (1983). —— RAYMOND, C. A. and LINDGREN, D.: A model for genetic flexibility. Proceedings Frans Kempe Symposium. June 10 to 11 1986. Umeå, Sweden. pp 159—177 (1986). —— REHFELDT, G. E.: Ecological adaptations in Douglas-fir (*Pseudotsuga menziesii* var. *glauca*) populations. I. North Idaho and north-east Washington. Heredity 43, 383—397 (1979). —— REHFELDT, G. E.: Differentiation of *Larix occidentalis* populations from the northern Rocky Mountains. Silvae Genetica 31, 13—18 (1982a). —— REHFELDT, G. E.: Ecological adaptations in Douglas-fir populations. II. Western Montana. U.S. Forest Service Research Paper INT-295, Intermountain Forest and Range Experiment Station, Ogden, Utah. 8 pp. (1982b). —— REHFELDT, G. E.: Seed transfer guidelines for Douglas-fir in Western Montana. US Forest Service Research Note INT-329, Intermountain Forest and Range Experiment Station, Ogden, Utah, 3 pp. (1983). —— REMRÖD, J.: Val av tall provenienser i norra Sverige — en analys av överlevnad, tillväxt och kvalitet i 1951 års tallproviensförsök. Skogshögskolan, Institutionen för Skogsgenetik, Rapporter och Uppsatser 19, 1—132 (1976). —— ROBERDS, J. H. and NAMKOONG, G.: Maximization of expected value for a trait in an environmental gradient. Proceedings 9th North American Forest Biology Workshop. June 1986. Oklahoma State University (1986). —— ROBERDS, J. H. and NAMKOONG, G.: Population selection to maximise value in an environmental gradient. Theoretical and Applied Genetics 77, 128—134 (1988). —— ROBERDS, J. H., HYUN, J. O., NAMKOONG, G. and RINK, G.: Analysis of height growth response functions for White Ash provenances. Silvae Genetica, (In Press), (1990). —— SAS Institute Inc.: SAS User's Guide, Statistics. (1982). —— SILEN, R. R. and MANDEL, N. L.: Clinal genetic growth variation within two Douglas-fir breeding zones. Journal of Forestry 81, 216—227 (1983). —— Skogsstyrelsen: Föreskrifter m.m. till Skogsvårdslagen (1979:429) (1979). —— WRIGHT, A. J.: The analysis and prediction of some two factor interactions in grass breeding. J. Agric. Sci., Cambridge 76, 301—306 (1971). —— WRIGHT, A. J.: Bias in the estimation of regression coefficients in the analysis of genotype-environment interaction. Heredity 37, 299—303 (1976).