

this stand could be increased by about three percent by eliminating natural inbreeding.

Knowledge of a species' breeding system is of a practical importance because of its genetic consequences. The mixed mating system of many *Pinaceae* species (SQUILLACE, 1974) and particularly the opportunity for spatial and temporal variations in the system (BROWN *et al.*, 1975, MORAN and BROWN, 1980, and YEH and LAYTON, 1979), raises problems for provenance and progeny testing using open pollinated seeds. If one provenance sample is less vigorous than another this may reflect useful adaptive variation or merely an ephemeral difference in the proportion of inbreds in the seed crop. Similarly, an unwarranted assumption of random mating may cause over estimation of additive genetic variance (NAMKOONG, 1966).

Finally, to effectively use knowledge of a species' breeding system it is also necessary to understand the mechanisms through which the breeding system is maintained. Investigation of reproductive biology, phenology, and pollination ecology are therefore an implicit component of breeding system studies.

Literature Cited

- BROWN, A. H. D., MATHESON, A. C., and ELDRIDGE, K. G.: Estimation of the mating system of *Eucalyptus oblique* L' HERIT. using allozyme polymorphisms. *Aust. J. Bot.* 23: 931–949 (1975). — EL-KASSABY, Y. A., YEH, F. C., and SZIKLAI, O.: Inheritance of allozyme variants in coastal Douglas-fir [*Pseudotsuga menziesii* var. *menziesii* (MIRB.) FRANCO]. (in preparation). — KAHLER, A., CLEGG, M. C., and ALLARD, R. W.: Evolutionary changes in the mating system of an experimental population of barley (*Hordeum vulgare* L.). *Proc. Nat. Acad. Sci. (U.S.A.)* 72: 943–946 (1975). — KIMURA, M. and CROW, J. F.: An introduction to population genetics theory. Harper and Row, New York, Evanston and London (1970). — KRZAKOWA, M.: Variability of Glutamate-Oxalate-Transaminase (GOT-2.6.1.1) isoenzymes in open-pollinated progeny of homozygous Scots pine (*Pinus sylvestris* L.) trees. *Botanicorum Poloniae* 49: 143–147 (1980). — LANGNER, W.: Über Fehlbeurteilungen von Saatguterntebeständen nach dem Phänotyp. *Forstpflanzen-Fortsamen* 3: 25–28, 30–36 (1966). — LEWONTIN, R. C., and KRAKANER, J.: Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175–195 (1973). — LUNDKVIST, K.: Allozyme frequency distributions in four Swedish populations of Norway spruce (*Picea abies* K.). I. Estimations of genetic variance within and among populations, genetic linkage and a mating system parameter. *Hereditas* 90: 127–143 (1979). — MITTON, J. B., LINHART, Y. B., HAMRICK, J. L., and BECKMAN, J. S.: Observations on the genetic structure and mating system of ponderosa pine in the Colorado Front Range. *Theor. Appl. Genet.* 51: 5–13 (1977). — MORAN, G. F., and BROWN, A. H. D.: Temporal heterogeneity of outcrossing rates in alpine ash (*Eucalyptus delegatensis* R. T. BAK.). *Theor. Appl. Genet.* 57: 101–105 (1980). — MULLER, G.: A simple method of estimating rates of self-fertilization by analyzing isozymes in tree seeds. *Silvae Genet.* 25: 15–17 (1976). — NAMKOONG, G.: Inbreeding effects on estimation of genetic additive variance. *For. Sci.* 12: 8–13 (1966). — RUDIN, D., ERIKSSON, G., and RASMUSON, M.: Inbreeding in a seed tree stand of *Pinus sylvestris* L. in northern Sweden. A study by the aid of the isozyme technique. *Research Notes No. 25*, Department of Forest Genetics, Swedish College of Forestry. 46 pp. (1977). — SORENSON, F. C.: Frequency of seedlings from natural self-fertilization in coastal Douglas-fir. *Silvae Genet.* 22: 20–24 (1973). — SQUILLACE, A. E.: Average genetic correlations among offspring from open-pollinated forest trees. *Silvae Genet.* 23: 149–156 (1974). — STERN, K., and ROCHE, L.: Genetics of forest ecosystems. Springer-Verlag, New York. 330 pp. (1974); *Biology* 41: 97–114 (1969). — YEH, F. C. and LAYTON, C.: The organization of genetic variability in central and marginal populations of Lodgepole pine (*Pinus contorta* spp. *latifolia*). *Can. J. Genet. Cytol.* 21: 487–503 (1979). — YEH, F. C., and O'MALLEY, D. M.: Enzyme variations in natural populations of Douglas-fir [*Pseudotsuga menziesii* (MIRB.) FRANCO] from British Columbia. I. Genetic variation patterns in coastal populations. *Silvae Genet.* 29: 83–92 (1980).

Performance Level – Standardized Score for Progeny Test Performance

By A. V. HATCHER, F. E. BRIDGWATER and R. J. WEIR

Respectively, Data Processing Supervisor, North Carolina State-Industry Tree Improvement Cooperative, Research Geneticist, U.S.F.S., N. C. State University, Director N.C.S.U.-Industry Tree Improvement Cooperative.

(Received 6th July 1981)

Summary

Performance levels are standardized scores for families in progeny tests of the North Carolina State University-Industry Tree Improvement Cooperative. These scores are one method of combining data from genetic tests with unbalanced experimental designs and unbalanced mating designs. Mean performance levels across tests are used to rank specific cross (full-sib) and clonal (half-sib) values to determine their relative worth in tree improvement programs. With appropriate transformation, one may determine what proportion of a normally distributed population individual full-sib or half-sib average performance levels outrank.

Key words: Progeny testing, standard scores, performance level

Zusammenfassung

In den Nachkommenschaftsprüfungen des kooperativen Waldbaumzüchtungsprogramms zwischen der Universität

North Carolina und der Industrie sind die Leistungsniveaus von Familien standardisierte Punktzahlen. Diese Punktzahlen sind eine Methode, um Daten von genetischen Tests von unbalancierten Versuchsplänen mit unvollständigen Kreuzungsplänen zu kombinieren. Um bestimmte Kreuzungen (Vollgeschwister) und Klone (Halbgeschwister) in eine Reihenfolge zu bringen, werden Leistungsniveaus über mehrere Tests verwendet, und auf diese Weise der relative Wert der Versuchsglieder in Züchtungsprogrammen bestimmt. Mit einer geeigneten Transformation kann man bestimmen, welcher Anteil einer normal verteilten Population über der durchschnittlichen individuellen Voll- oder Halbgeschwisterleistung liegt.

Introduction

A major reason for progeny testing is to rank parents based on the average performance of their offspring. Progeny tests of forest trees are replicated in time and space

to quantify genotype \times environment interactions associated with different site types and year-to-year environmental variation. If each parent is represented by a sufficient number of specific crosses for accurate evaluation, and each of these crosses is represented in every test, genetic, environmental, and genotype \times environment effects can be easily estimated by analysis of variance. This degree of balance rarely occurs in progeny tests of forest trees because of pressures to initiate tests quickly (the first step into the next generation of improvement) have been great. Customarily when sufficient seeds were produced from about 25 specific crosses, a field test was planted. The result has been that all crosses were not planted in the same year and/or location. Development of a procedure which would place all the available information on a comparable basis was necessary before a summary analysis of the worth of a parent was possible.

The North Carolina State University-Industry Pine Tree Improvement Cooperative used a tester mating scheme to produce progeny tests of first-generation parents selected from natural populations. In this mating design, four to six male parents were used to control-pollinate the other trees selected by an individual cooperator. The same set of specific crosses (a half-diallel among the tester parents) was to be planted in every test. These six crosses were referred to as genetic check seedlots and were to be the basis for adjusting the performance of crosses in different tests to a comparable basis. The shortcomings of this system became apparent over time. First and foremost, since the same set of genetic checks was not included in every test, adjustment to a common basis was impossible. Second, if the average performance of genetic checks was influenced by genotype \times environment interactions, adjusted values for crosses would be biased. Finally, since variances among tests were not homogeneous, pooling information over tests would bias the evaluation in the direction dictated by the tests with the greatest variances. Similar situations occur with other kinds of data, particularly educational testing scores, and procedures for adjusting scores to a common base for ease of comparison are well known (FAGAN and HORST 1976). This paper describes application of such procedures to progeny test data.

Standard Scores

Scores expressed in terms of the standard deviation of the distribution are called standard scores. These indicate how far above or below the mean of the test population a score lies in units of the standard deviation. If the units of the raw data, say feet of height, are equal-interval (i.e., one unit of change means the same thing everywhere in the distribution), then the standard scores are also equal interval. This is important, since it means standard scores can be meaningfully averaged across tests.

The simplest standard score is a "Z" score. This commonly used transformation, described in most basic statistics textbooks, changes the observed data so that the sample mean is zero and the sample variance is one. The transformation is simply:

$$Z = \frac{x - \bar{x}}{s} \quad (1)$$

where: x = an observation on the untransformed random variable

\bar{x} = the sample mean (test population)

s = the estimated standard deviation

If x is normally distributed, and the sample size is sufficiently large, then Z has an approximate standard normal distribution. If x is not known to be normally distributed, a transformation to normalize the variable may be necessary. We assume that most measured characteristics of forest trees are normally distributed. However, each progeny test represents a subsample of the essentially infinite population for a species and therefore may not have a normal distribution. In this case scores would not be on an equal-interval scale and it would not be appropriate to average information over tests. For example, if there were many tall trees in a test and relatively few short ones, an increase of one unit of height for a tall family would increase the rank more than the same change at the low end of the scale. Scores can be normalized by determining the cumulative relative frequency from the distribution of observed values and converting these to standard normal deviates by using a table of areas under a normal curve.

Standardized values above the mean have positive values and those below have negative values. For ease of calculation, some kind of transformation is desirable to make all values positive. If such a transformation is linear, (i.e. each value has a constant added to or subtracted from it and/or is multiplied by a constant), the new set of values will remain on an equal-interval scale.

Performance Levels

The calculation of "performance levels" is a procedure recently adopted by the North Carolina State University-Industry Tree Improvement Cooperative to standardize observed values of family means. The procedure for calculating performance level, (PL) is as follows:

$$PL = (100) \cdot \frac{(x - x \text{ min})}{\text{Range}} \quad (2)$$

where:

x = the mean value for a cross

$x \text{ min}$ = the minimum value for the poorest cross mean

Range = the maximum value for the best cross mean minus $x \text{ min}$.

Cross means falling outside two standard deviations from the test mean are arbitrarily assigned values of two times the standard deviation. Thus, $x \text{ min} = \bar{x} - ((2) \cdot (\text{standard deviation}))$ and the Range = $(4) \cdot (\text{standard deviation})$. The standard deviation of family means, s_f , is used to calculate performance level. Thus (2) becomes:

$$\begin{aligned} PL &= (100) \cdot (x - (\bar{x} - 2s_f)) \\ &= (100) \cdot \frac{(x - \bar{x}) + \frac{4s_f}{2}}{4s_f} \\ &= (100) \cdot \frac{Z + \frac{1}{2}}{\frac{4}{2}} \\ &= 25Z + 50 \end{aligned} \quad (3)$$

which is a linear transformation of a Z-value. Performance level is a linearly transformed standardized score and has an equal-interval scale if family means are normally distributed. A numerical example is given below. The within-test performance level for cross 38-11 \times 38-2 using values from Figure 1 is:

$$\begin{aligned}
 \text{Performance level} &= (100) \cdot \frac{(\text{Cross value} - \text{Minimum value})}{\text{Range in values}} \\
 &= (100) \cdot \frac{(22 \text{ ft} - 12 \text{ ft})}{16 \text{ ft}} \\
 &= 62.5
 \end{aligned}$$

Using Performance Level

Performance level is useful only as a measure of relative performance. It has no meaning other than by association, i.e. a PL = 50 is average, a value above 50 is better than average, etc. Performance level *does not* indicate the proportion of a population a particular family ranks above. For example, a cross whose performance level is 70 does not rank higher than 70% of the crosses in the test population. However, this useful piece of information may be simply determined by recalling (4) and solving for Z,

$$Z = \frac{PL - 50}{25} \quad (5)$$

Then, for PL = 70, Z = 0.80, and from a table of areas under the normal curve (e.g. SNEDECOR and COCHRAN 1967, p. 548) we find that a family with PL = 70 outperforms 79% of the families in that test. This value may be easily determined for any performance level using Figure 2. Simply find the performance level on the x-axis and determine the percentage of entries that ranked lower on the Y-axis. The example for PL = 70 is repeated in Figure 2. It should be noted that these percentile figures are not equal-interval scale and therefore should not be averaged.

Averaging Performance Levels

Thus far, examples have shown how to calculate performance levels for specific crosses (full-sib families) within a test. Since performance level is a standardized score, we can meaningfully average these values for a family from different tests.

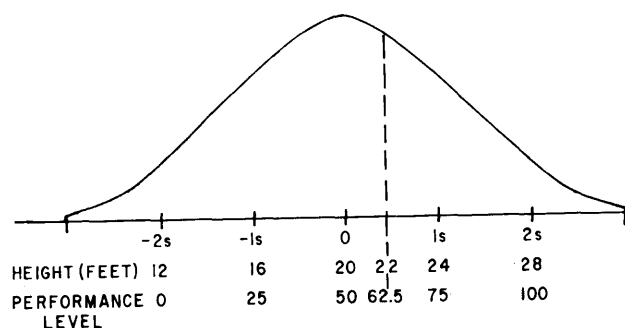
For example, family 38-11 x 38-12 appears in 6 tests, and thus has 6 "within-test" performance level values:

Test Number	Performance Level
1	62
2	56
3	60
4	58
5	64
6	60
\bar{x}	60

The overall performance level is simply the mean of these 6 values ($\bar{x} = 60$). If we wish to calculate a performance level for clone 38-11 we average the performance level estimates for all specific crosses with 38-11 as a parent:

Cross	Performance Level
38-11 x 38-12	60
38-11 x 38-4	55
38-11 x 38-15	65
38-11 x 38-17	76
\bar{x}	64

We recognize that averages of performance levels over tests are providing information on some functions of genetic factors in tests adjusted for site variability. Therefore,



CROSS 38-11 x 38-12 = 22 FT.
 POPULATION MEAN = 20 FT.
 STANDARD DEVIATION = 4 FT.

Fig. 1. — Illustration of performance level for a full-sib family in one genetic test.

PL is an indicator of relative performance and not a precise measure of performance. However, this information is useful for tree breeders who must have some measure of relative average performance on which to base selections.

Note that Figure 2 may not be used to determine the percentage of either full-sib or half-sib family means an entry outranks without further calculation. The range of performance level averages will be reduced from 100 since it is unlikely that any cross or group of crosses will have the same relative rank in all tests (e.g., seldom will any cross always be poorest or best in all tests). Since the standard normal distribution of individual values has been truncated, a new distribution based on average values must be used to find the proportion of values an individual value outranks. For example, one industry-owned seed orchard contains 37 clones for which test information is available

Table 1. — Form of clonal performance report provided annually to cooperators in the North Carolina State University — Industry Tree Improvement Cooperative.

CLONE	PERFORMANCE LEVELS FOR CLONES				NUMBER OF CROSSES
	HEIGHT	VOLUME	FORM	RUST RESISTANCE	
1	12	18	39	60	1
2	25	43	8	63	1
CC	28	39	13	59	2
3	29	32	33	74	1
4	31	33	61	48	6
5	35	36	48	59	26
6	36	38	19	58	1
7	37	42	19	53	2
8	39	41	52	41	4
9	40	39	32	62	6
10	40	37	56	78	2
11	40	51	44	41	2
12	41	50	57	59	5
13	42	45	18	50	5
14	43	48	44	51	4
15	43	43	52	63	4
16	44	41	43	71	1
17	44	48	20	51	1
18	45	50	48	69	6
19	45	36	62	21	4
20	46	48	42	48	5
21	46	46	66	66	25
22	51	54	52	38	4
23	51	52	51	48	9
24	52	54	67	54	4
25	53	45	63	37	4
26	53	48	45	74	4
27	57	57	31	48	1
28	57	59	50	31	26
29	59	62	55	51	4
30	60	54	37	47	4
31	60	57	51	48	3
32	62	57	45	50	26
33	62	51	26	59	1
34	65	66	41	60	5
35	66	67	60	50	5
36	73	74	44	61	4
37	76	69	66	61	6

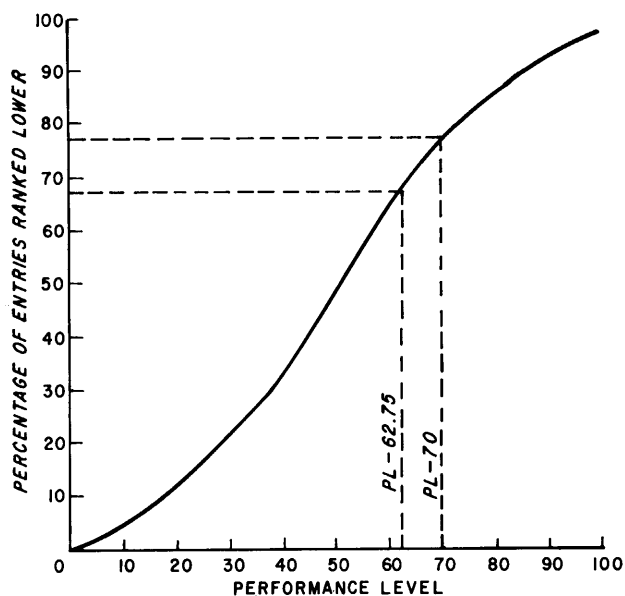


Fig. 2. — Relationship between performance level and percentage entries ranked lower.

(Table 1). The average performance levels for these clones range from 18 to 69 for volume. Substituting these values into (2) we assign a PL of "0" to the poorest clone (PL = 18) and "100" for the best clone (PL = 69). These newly transformed values may now be used to enter Figure 2. A clone whose nontransformed PL was 50 becomes 62.75 when transformed using (2) above. Using Figure 2, we find that this clone would outrank approximately 70% of entries on the average in a normally distributed population.

The procedure of transforming each time performance levels are averaged to arrive at new values must be followed to determine the proportion of entries a given family outranks. However, if only ranks, *per se*, among specific cross means or clone means are desired, this procedure is unnecessary.

Comparison of Procedures

The ranks of family means based on performance levels were compared with those based on values adjusted using genetic checks common to several tests. Results of 30 progeny tests which had at least five common full-sib crosses planted in at least three different tests were used to compare the two ranking procedures. Descriptive information

Table 2. — Weighted Pearson's and Spearman's Rank correlation coefficients among half-sib family mean values from two procedures for pooling means from several tests.

Comparison Number	Numbers of:			$r_{(weighted)}$	r_s
	Families	Tests	Common Full-Sib crosses		
1	17	4	6	0.92	0.92
2	15	4	6	0.97	0.89
3	34	4	7	0.98	0.97
4	29	3	4	0.97	0.95
5	34	4	7	0.94	0.94
6	25	4	5	0.96	0.99
7	27	7	21	0.97	0.99

for the tests and correlation coefficients between the family values derived from the two procedures are presented in Table 2.

All correlations presented in Table 2 are significantly different from zero at the 0.0001 confidence level. Weighted rather than unweighted Pearson's correlation coefficients were calculated since the numbers of full-sib crosses used to calculate half-sib means varied widely within a comparison group. Spearman's correlation coefficient indicates the extent to which ranks of means agree and is perhaps a more appropriate statistic for comparing the two methods of pooling values over tests. Values for both correlation coefficients are uniformly high for all comparison groups, particularly for those groups with at least 25 families. Both methods of ranking half-sib family performance give essentially the same result. The method of performance levels is preferred since the great majority of tests, unlike those presented in Table 2, do not have substantial numbers of common crosses to use in adjusting for test effects. Other methods of adjusting for test effects, e.g. regression procedures, were not considered because they are more complex and require more expensive computing time than the two procedures considered. These more complex procedures may be required to analyze data from advanced-generation tests where objectives other than progeny testing, e.g. estimation of genetic parameters, become more important.

Reporting Performance

This paper was written to document the reasons and procedures for calculating performance levels and to discuss various ways to use these values to derive useful information about relative performance. In practice, all cooperators in the North Carolina State University-Industry Tree Improvement Cooperative are routinely provided the information in Table 1 for their program.

Average clonal performance levels are given for height, volume (when diameter is measured), tree form and rust resistance. The performance level calculated for tree form is a combined assessment for crown and straightness and the performance level for rust resistance is a combined assessment using cronartium score and percent infection. Tree form score is calculated arbitrarily weighting straightness twice as heavily as crown. The performance level for rust resistance is calculated by multiplying the percent infection by the cronartium score which is essentially the level of infection weighted by the severity of the infection. When pooling information from age 4 and age 8 tests, age 8 tests are arbitrarily weighted twice as heavily as age 4 tests.

Performance levels are a convenient way to combine scores for families in genetic tests with unbalanced experimental and mating designs. This use simplifies the reporting and the interpretation of average full-sib or half-sib family performance over several experiments.

Literature Cited

FAGAN, B. M. and D. P. HORST: Types of test scores, *In: ESEA Title I Evaluation and Reporting System*, RMC Research Corporation, Mountain View, California, Tech. Paper No. 8, October, 1976, 8 pp. (1976). — SNEDCOR, G. W. and W. G. COCHRAN: *Statistical Methods*. 6th Edition. Iowa State University Press, Ames, Iowa (1976).