Theor Appl Genet (1989) 77:825–831

Deployment to plantations of numbers and proportions of clones with special emphasis on maximizing gain at a constant diversity

D. Lindgren [1], W. S. Libby [2] and F. L. Bondesson [3]

[1] Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden
[2] Department of Genetics and Department of Forestry Resource Management, University of California, Berkeley, CA 94720, USA
[3] Department of Biometry and Forest Management, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden

Received November 22, 1988; Accepted January 4, 1989
Communicated by P. M. A. Tigerstedt

Summary. The value of a mixture of genetic entries present in different proportions is defined. A measure of the disadvantage of reduced diversity is defined as the sum of the squares of the proportions of the different entries. An algorithm for maximizing genetic gain of the mixture under the constraint of a constant disadvantage is developed. The optimal deployment strategy is one that lets the proportion of the genetic entries be linearly dependent on their genetic value. By use of rankits as entries for genetic values, optimal solutions for deployment were calculated for a range of values of available entries (from 10 to 5,000) and preset diversity-related disadvantage-factors (the preset values correspond to mixtures of between 2 and 100 entries in identical proportions). The values are tabulated so they can be used by breeders. The superiority of the proposed strategy increases with the proportion of the available entries which are selected. In the situation that around half would have been selected if truncation selection was applied, the improvement in genetic gain compared to classical truncation selection is up to 18%. Thus, considerable improvements in gain are possible without any sacrifice in diversity. Applications are discussed with particular reference to clonal forestry.

Key words: Diversity – Breeding – Selection

Introduction

A common tactic in artificial selection is to rank the genetic entries (parents, families, or clones) and then find some truncation point between accepted and rejected entries. The difference in genetic value between the lowest-ranking accepted entry and the highest-ranking rejected entry is frequently very small, but the difference in deployment between complete acceptance and complete rejection is very large. Among the accepted entries, the difference in genetic value between the lowest-ranked and the highest-ranked is often substantial, but all those accepted are typically expected to contribute equally to the selected population.

A basic dilemma for breeders is the balance between the high gain obtained by selecting the few very best entries and the high diversity obtained by making use of many entries. Usually there is a desire to guarantee that the diversity does not drop below a certain level, but under that constraint to make as much gain as possible.

In this paper, we propose that the highest ranking entries should contribute in the highest proportions to the selected population; and we further propose that to maintain the desired level of genetic diversity in the population, some additional entries, which would fall below a conventional truncation point, contribute in small proportions to the selected population. We have developed a rationale for determining the number and proportions of ranked clones to be deployed for clonal forestry purposes, while maximizing the genetic gain under the constraint that the diversity is kept at a preset desirable level.

Lindgren and Matheson (1986) have made a similar study for deployment of clones to seed orchards. However, their algorithm required a quantitative estimate of the drop in production caused by a certain level of disadvantage. Such an estimate is uncertain and open to criticism; furthermore, it sometimes leads to solutions generating low levels of diversity that are unlikely to be accepted or acceptable. These difficulties are circumvented by the present treatment, by using a level of diversity as an input constant rather than deriving it as a result. While Lindgren and Matheson (1986) made calculations just for a few examples, in this paper we try to

cover the whole range of situations that may occur in practice.

Intuitively, one will achieve greater gains by planting relatively more of the better clones in a deployed set, and fewer of the poorer ones. There are, however, risks associated with high proportions of a few clones, but these are usually independent of the relative advantages of the deployed clones. These risks are generally related to levels of genetic diversity in the deployed population. Our approach in this paper is to compare gains at an equal level of diversity. This is done by calculating a measure of genetic diversity in a deployed population utilizing equal proportions of a truncated number of clones and then calculating the optimum number and proportions of clones that equal that measure of diversity (risk) while simultaneously maximizing expected gain. The theoretical results presented below indicate that deploying more clones in unequal proportions will result in greater gains (higher average performance) than will deploying fewer clones in equal proportions, at equal levels of diversity. In some cases the differences between the two alternatives are small; in other cases they are substantial.

Measure of proportion-dependent disadvantages

Assume that negative interactions between ramets of the same clone lead to reduced performance and that there are no positive biological interactions between ramets of the same clone. Two examples of such negative interactions are that identical genotypes make similar demands on the site at similar times, and that a pathogen genotype adapted to one ramet finds it easy to spread to other ramets of the same clone, particularly so if they are neighbors. If the likelihood and/or intensity of a biologically significant interaction between two trees in a plantation depends in part on the distance between them, and if the sign of such interactions between ramets of the same clone is always negative, then the mathematical development below may be appropriate. The probability that such a negative interaction will occur in a given plantation between two randomly developed ramets of clone i with proportion p_i in the plantation may be quantified by p_i^2. If such within-clone interactions cause a drop in performance by a factor of c compared to interactions between ramets of different clones, the loss of performance in the plantation explainable as a function of clonal proportion is $c\,\Sigma p_i^2$. We will now call Σp_i^2, d, a "disadvantage" term. (We suggest that d is a good measure of "diversity", but d becomes smaller when diversity increases, and this may cause confusion. Therefore we call it "disadvantage". Actually $1-\Sigma p_i^2$ would be a conceptionally sound measure of diversity, as it gets its minimum value = 0, when diversity is at minimum, e.g. when only one clone is deployed. It gets its maximum = 1 when an infinite large number of clones is deployed. However, the $1-d$ would make the mathematics more obscure, thus we keep to the disadvantage concept instead.) We also employ the term c, which may be viewed as a scaling factor that makes it possible to relate a certain d value to a certain drop in performance. In this paper, however, we do not need to know the value of c.

The expression $1/\Sigma p_i^2$ (or $1/d$) has been used as a measure of genetic diversity or effective population size. For example, Muller-Starck and Gregorius (1986) call it "differentiation effective number", defining p_i as relative frequency of the ith genetic type. Thus, d is not only a measure of "disadvantage", it is the inverse of the genetic diversity or effective population size within a population

$$\Sigma p_i^2 = d = 1/N_{eff}\,. \tag{1}$$

In classical truncation selection, a number of clones is selected and deployed in equal proportions. We will call that number n_e when referring to the truncation selection situation. The disadvantage, d, is just the inverse of n_e.

$$d = 1/n_e$$

Another argument to use Σp_i^2 as a measure of disadvantage is that it actually is a measure of the expected variation of the average of the clonal mixture in different environmental regimes. The value of clone i is given by the variable X_i, where $\mathrm{Var}(X_i) = \mathrm{Var}(X)$ for all i. The variance of clonal mixture will be

$$\mathrm{Var}(\Sigma p_i X_i) = \Sigma p_i^2\,\mathrm{Var}(X)\,.$$

Mathematics for maximizing net gain

Let g_i $(i=1, 2, \ldots N)$ be the genetic values of the N components planted in a genetically diverse plantation in which the components are distributed in a random fashion. (Clones are the primary considerations in this paper, but the concept applies as well to families or other units of genetic entry.) The values g_i are ranked so that

$$g_1 >= g_2 >= \ldots >= g_N\,.$$

The net value of an individual clone is its value g_i reduced by the negative effect of within-clone neighbor interactions. This negative effect is proportional to the probability (p_i) that a neighbor is of the same clone and is modified by a coefficient c, that quantifies the actual size of the negative effect $(g_i - c\,p_i)$.

The net gain to be maximized is

$$G = \Sigma p_i(g_i - c\,p_i) = \Sigma p_i g_i - c\,\Sigma p_i^2 \tag{2}$$

under the restriction that all $p_i >= 0$. There is also a constraint $\Sigma p_i = 1$. These restrictions on p_i are of course logical necessities in a real plantation, but nevertheless important for the development of the mathematics.

This maximization problem was solved by Lindgren and Matheson (1986).

Set

$$g_a = \Sigma g_i/n \quad \text{(average } g_i \text{ for } i <= n) . \tag{3}$$

The optimal proportions are

$$p_i = (g_i - g_a)/(2c) + 1/n \quad \text{for} \quad i <= n \tag{4}$$
$$p_i = 0 \quad \text{for} \quad i > n$$

where n is the highest value for which formula (4) will not give a nonpermitted negative value of p_n. The value of n has to be found by trying some values. This search can be made by some algorithm, (see below).

An alternative way of expressing formula (4) is

$$p_i = (g_i - g_0)/(2c), \quad \text{where} \quad g_0 = (g_a - 2c/n) . \tag{5}$$

Formulae (4) and/or (5) show that the optimal proportions are linearly related to the genetic values.

In this paper we solve the maximization problem, adding an additional constraint that

$$\Sigma p_i^2 = d = 1/n_e$$

is constant. In practical terms, some desired or required level of diversity or risk is decided upon prior to allocating clones to the plantation. [The solution to this problem was mentioned by Lindgren (1986), referring to this paper, then in preparation.]

To maximize G under the constraints

$$\Sigma p_i = 1 \quad \text{and} \quad \Sigma p_i^2 = d \tag{6}$$

in the region $D = \{\boldsymbol{p}; p_i >= 0, \ i = 1, \ldots, N, \ (\boldsymbol{p} = (p_1, \ldots, p_N)\}$ we may use the Lagrange multiplier technique; cf. e.g., Luenberger (1973, pp. 224–225). First we ignore that $\boldsymbol{p}$ should be in D. Let

$$L = G - \Theta(\Sigma p_i - 1) - \mu(\Sigma p_i^2 - d) ,$$

where Θ and μ are the multipliers. If $\boldsymbol{p} = (p_1, \ldots, p_N)$ is a local maximum point or minimum point of G under the constraints (6), then

$$\frac{\partial L}{\partial p_i} = 0 , \quad (i = 1, \ldots, N) , \quad \frac{\partial L}{\partial \Theta} = 0 , \quad \frac{\partial L}{\partial \mu} = 0$$

(The Lagrange equations).

In our case the equations will be

$$g_i - \Theta - 2\mu p_i = 0 , \quad i = 1, \ldots, N, \tag{7}$$
$$\Sigma p_i = 1 \tag{8}$$
$$\Sigma p_i^2 = d . \tag{9}$$

Here (9) together with (8), may be rewritten as

$$\Sigma (p_i - 1/n)^2 = d - 1/N . \tag{10}$$

By summing (7) and using (8), we get $\bar{g} - \Theta - 2\mu/N = 0$, and hence by eliminating Θ in (7),

$$p_i - 1/N = (g_i - \bar{g})/(2\mu) , \quad i = 1, \ldots, N . \tag{11}$$

From (10) we then get $\mu = \pm \mu_0$, where

$$\mu_0 = \sqrt{\Sigma (g_i - \bar{g})^2/(4(d - 1/N))} . \tag{12}$$

Thus, there are only two solutions of the Lagrange equations:

$$p_i = 1/N + (g_i - \bar{g})/(2\mu_0) \quad \text{and} \quad p_i' = 1/N - (g_i - \bar{g})/(2\mu_0)$$
$$(i = 1, \ldots, N) .$$

Certainly, G is largest at the first solution. Since the domain for $\boldsymbol{p}$ is compact, there is a global maximum point as well as a global minimum point. Thus, the solutions above are these two points. We may neglect the second solution. If all p_i are positive (i.e., if $\boldsymbol{p}$ belongs to the interior of D), the maximization problem is solved. If not, in absence of any other local extrema than those found, the global maximum in the region D must be attained on the boundary of D, i.e. at a point where some p_i equals 0. As $g_1 >= g_2 >= \ldots \geq = g_N$, some reflection shows that $p_N = 0$. The procedure above can be repeated for $p_1, \ldots, p_{N-1}$ instead of $p_1, \ldots, p_N$. If the solution is not feasible, we set $p_{N-1} = 0$ etc. The iterative procedure stops at an $n = N - k$ such that $p_1, \ldots, p_n$ are all positive. The final solution will be the global maximum point of G in D under the constraints (6).

Thus the optimal solution is

$$p_i = 1/n + (g_i - g_a)/(2\mu_0) \quad \text{for} \quad i <= n \tag{13}$$

$p_i = 0$ for $i > n$, where μ_0 is given by expression (12) with $\bar{g}$ replaced by g_a and N by n.

Note that g and μ_0 are functions of n, cf (11) and (12).

Alternatively, the formula may be expressed:

$$p_i = b(g_i - g_0), \quad \text{where} \quad g_0 = g_a - 2\mu_0/n$$
$$\text{and} \quad b = 1/(2\mu_0) . \tag{14}$$

The relationship is graphically demonstrated in Fig. 1. There is a linear relationship between optimal proportion and genetic value. This is a useful and powerful finding in itself. The slope of the line is $b = 1/(2\mu_0)$.

Note the close similarity between formula (4) and (13). The alternative formulations (5) and (14) are, of course, also similar.

For actual computations of n, it is usually easiest to start with $n = 1$ and then successively try one step higher n. For each n, equation (14) is solved, till an n making $p_n < 0$ is found. The optimal n is one step higher (the highest n making $p_n > 0$). The proof presented above actually just covers the situation that n values are tested starting with the biggest and successively trying one step lower values. However it is possible to prove that both procedures give equivalent results.

Scaling

For our demonstration, we will use g-values from an idealized case, with normal distribution and standard

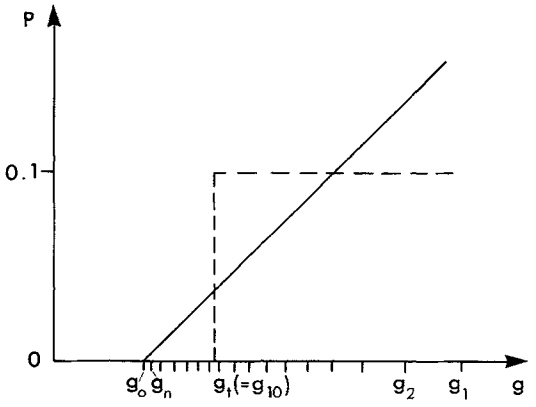

Fig. 1. Graphic illustration. On the x-axis the genetic value (g) is found, and on the y-axis the proportion (p) corresponding to the g-value. The classical approach has been to deploy by truncation selection, thus for the entries with g-values above a certain truncation point (g_t), all are deployed in equal proportions. This is demonstrated by a hatched line in the figure. The optimal proportions lie on a straight line as indicated in the figure. g_0 is the genetic value at which p becomes 0. For $g_i < g_0$ the optimal proportion is 0. The slope of the line is $b = 1/(2\mu_0)$. For a set of g_i-values and a chosen value for d (corresponding to the desired diversity level), such an optimizing line can be found. Examples of such optimizing lines are found in Table 2. The figure demonstrates a numerical example: If d is set to 0.1 (corresponding to a mixture of 10 clones in equal proportions) and there are 50 candidate clones whose genetic values are expressed in rankits, cf Table 1, the optimizing line (formula 13) will give the equation:

$$p = 0.9981 \cdot (g - 0.4726)$$

For some g_i-values, the same p_i-values can alternatively be obtained using Table 3.

Note that the best clones get a higher representation by the optimal solutions, while those on the bottom of the list of those deployed after truncation selection get lower representation. On the other hand around 40% clones more enter the list of deployed

deviation 1. Ranked values can always be parameterized by assigning corresponding expected ordered values from the standardized normal distribution (rankits). In practice, a breeder could take into account larger or smaller gaps between adjacent ranked entries, and adjust the values accordingly. Our standardized rankit values will be designated x instead of g. A change of scale with a scaling factor s, the standard deviation, so that $g = s\,x$, will not change the optimal proportions obtained by (14). If (5) is applied, it has to be considered that the scale of c will also change.

Calculations and results

In this paper we calculate optimal deployment for a number of cases which range over most situations of practical interest.

Although it is possible to use estimated g_i values obtained from real experiments, we make calculations using rankit values. The reasons for this are the same as the use of "selection intensity" in genetic gain calculations. They serve as predictions of the typical outcome of a selection procedure rather than making the best possible predictions of the outcome in a specific experimental material, which may be of less general interest. The use of the standardized rankit values makes the outcome of calculations for different cases more comparable. The rankits are expected normal order statistics, which we designate x_i. The rankits were generated using the algorithm NSCOR1 (Royston 1982). Rankit values for some situations are presented in Table 1 for illustrative purposes, and to encourage calculations like ours. Rankit values for some $N <= 400$ have been published, e.g., Harter (1970); for other situations it is necessary to rely on one's

Table 1. Rankit values, x_i. Expected value of the ith largest of a sample of size N from a normal distribution with standard deviation 1. Thus, the values are expressed as standard deviations. Rankits for $i > N/2$ may be calculated using $x_i = -x_{N+1-i}$. Even these below-average clones may be included in an optimal mixture as in Table 3. Intermediary rankits may be obtained by linear interpolation

N	$i=1$	2	3	4	5	6	7	8	9	10	12	15	20
10	1.5388	1.0014	0.6561	0.3758	0.1227	−0.1227	−0.3758	−0.6561	−1.0014	−1.5388			
20	1.8675	1.4076	1.1309	0.9210	0.7454	0.5903	0.4483	0.3149	0.1870	0.0620	−0.1870	−0.5903	−1.8675
50	2.2491	1.8549	1.6286	1.4637	1.3311	1.2185	1.1195	1.0304	0.9489	0.8732	0.7351	0.5508	0.2781

N	$i=1$	2	3	4	5	10	15	20	30	50	70	90
100	2.5076	2.1481	1.9464	1.8018	1.6872	1.3062	1.0551	0.8574	0.5376	0.0125	−0.5089	−1.2495
200	2.7460	2.4137	2.2300	2.0999	1.9978	1.6658	1.4547	1.2938	1.0457	0.6816	0.3916	0.1318
500	3.0367	2.7323	2.5667	2.4506	2.3603	2.0719	1.8936	1.7607	1.5620	1.2864	1.0842	0.9187
1,000	3.2414	2.9541	2.7992	2.6912	2.6076	2.3431	2.1818	2.0628	1.8872	1.6490	1.4790	1.3434
2,000	3.4353	3.1626	3.0165	2.9151	2.8370	2.5915	2.4432	2.3347	2.1759	1.9637	1.8147	1.6977
5,000	3.6776	3.4210	3.2847	31.906	3.1182	2.8927	2.7578	2.6597	2.5174	2.3297	2.1997	2.0990

own calculations. Our calculated cases comprise the N-values 10, 20, 50, 100, 200, 500, 1000, 2000, and 5000.

Recall that d is our "disadvantage" term. We used values of $d = 0.01$, 0.02, 0.05, 0.1, 0.2, and 0.5, which correspond to diversity in random mixtures made up of equal proportions of $n_e = 100$, 50, 20, 10, 5, and 2 deployed clones, respectively.

Formula (13) (or 14) was applied for all possible combinations of the N and d values given above, using rankits as genetic values. Table 2 presents results for combinations of N and d. This table presents the number of clones deployed (n), the slope of the optimizing line (b), and the rankit value (x_0) of the zero-proportion intercept (Fig. 1). Note that n and x_0 are a way to express the solution for the maximum gain; they constitute the two parameters required to describe the linear solution. By inserting the b and x_0-values from Table 2 and the rankit values from Table 1, the optimal proportions for the clones can be calculated according to their rank. The procedure is graphically illustrated in Fig. 1. The maximum gain is a function of the entries N and d. The lowest-ranking deployed clone is the clone whose rankit value (x_n) is equal to or is the closest value larger than x_0.

For some combinations of N and d, and for some x_i, disproportion coefficients (f_i) for the ranked clones were calculated, and are presented in Table 3. For example, a d of 0.2 results if 5 clones ($n_e = 5$) are deployed in equal proportions (each in proportion 0.2). The top row of Table 3 indicates that the top-ranked clone is deployed in df_i proportion $= (0.2)(1.508) = 0.3016$, and the 7th-ranked clone (not included at all in truncation selection) is deployed in the proportion $(0.2)(0.1036) = 0.02072$. Note that Σf_i for any given complete line in Table 3 should add up to $1/d$. In the lower part of Table 3, it was not feasible to present f_i values for all i. Neither is it feasible to print solutions for many d- and N-values. But the grid of values is dense enough to make it reasonably safe to make linear interpolations for non-tabulated values of d, N, and i if required, excepting i-values close to n. The latter class of values is small, and the genetic values are close, so small interpolation errors will be insignificant. One reason that we tabulate f_i rather than p_i is that the values do not change drastically with variation in d and N, thus making interpolations safer.

The expected performance of the population, in terms of gain, is $\Sigma p_i g_i$. After truncation selection with the requirement of d, the gain would be $\Sigma g_i/n_e = d\,\Sigma g_i$ (sum to $i = n_e$). If g_i are rankits, this is actually the selection intensity. A ratio of $\Sigma p_i g_i$ where p_i is unequal, and the corresponding, where the p_i are equal, provides a measure of advantage or "efficiency" of weighted deployment compared to equal deployment. Table 4 provides such efficiencies as percentage improvement, by ratio

$$Q = [(\text{unequal}/\text{equal}) - 1] \cdot 100\,,$$

Table 2. Maximum-gain solution at various combinations of sets of N normally-distributed clones and d levels of diversity-related disadvantage, providing number of clones deployed (n), slope (b) of the regression of proportion deployed on rankit value, and rankit value (x_0) of the zero-proportion intercept

N	n	b	x_0	n	b	x_0
	$d = 0.01$, $1/d = 100$			$d = 0.02$, $1/d = 50$		
100	–	–	–	72	0.01311	−0.5957
–						
200	146	0.006510	−0.6012	80	0.01764	0.2534
–						
500	165	0.009356	0.4471	85	0.02218	0.9549
–						
1,000	172	0.010993	0.9518	87	0.02508	1.3588
–						
2,000	176	0.01242	1.3560	89	0.02768	1.7025
–						
5,000	180	0.01410	1.8014	90	0.03081	2.0965
	$d = 0.05$, $1/d = 20$			$d = 0.1$, $1/d = 10$		
–						
20	–	–	–	14	0.06896	−0.5598
–	–	–	–	–		
50	30	0.03763	−0.2521	16	0.09981	0.4726
–				–		
100	32	0.04820	0.4598	16	0.1178	0.9729
–				–		
200	34	0.05680	0.9626	17	0.1336	1.3757
–				–		
500	35	0.06652	1.4817	17	0.1520	1.8194
–				–		
1,000	35	0.07306	1.8105	17	0.1646	2.1110
–				–		
2,000	36	0.07907	2.1025	17	0.1763	2.3755
–				–		
5,000	36	0.08643	2.4483	18	0.1908	2.694
	$d = 0.2$, $1/d = 5$			$d = 0.5$, $1/d = 2$		
+						
10	7	0.1467	−0.5170	3	0.6490	0.5518
–						
20	8	0.1983	0.2979	3	0.7759	1.0391
–						
50	8	0.2525	0.9919	3	0.9195	1.5483
–						
100	8	0.2867	1.3921	3	1.0154	1.8724
–						
200	8	0.3174	1.7332	3	1.1035	2.1612
–						
500	8	0.3540	2.1245	3	1.2109	2.5033
–						
1,000	8	0.3794	2.3883	3	1.2863	2.7391
–						
2,000	8	0.4033	2.6313	3	1.3580	2.9593
–						
5,000	9	0.4331	2.9272	3	1.4472	3.2308

for combinations of N and d. It should be noted that this is a measure of efficiency for gross gain. To get net gain (G), as in formula (2), a term $c\,d$ should be deducted for both the unequal and the equal case; thus, the quotient between net gains will be higher, and thus the advantage

Table 3. Disproportion coefficients (f_i). Optimal proportions related to d ($d=\Sigma p_i^2$). $f_i = p_i/d$. $f_i = 1$ corresponds to the case when all entries are equally represented

d	N												
		$i=1$	2	3	4	5	6	7	8	9			
0.2	10	1.508	1.114	0.8605	0.6550	0.4693	0.2893	0.1036	0	0			
0.2	20	1.556	1.100	0.8260	0.6178	0.4437	0.2899	0.1492	0.0169	0			
0.2	50	1.587	1.089	0.8038	0.5957	0.4282	0.2860	0.1611	0.0487	0			
0.2	500	1.614	1.076	0.7825	0.5771	0.4172	0.2855	0.1730	0.0744	0			
0.2	5,000	1.625	1.070	0.7742	0.5704	0.4137	0.2859	0.1776	0.0836	0.0003			
		$i=1$	2	3	4	5	6	8	10	12	14		
0.1	20	1.674	1.357	1.166	1.021	0.9001	0.7932	0.6033	0.4288	0.2571	0.0769		
0.1	50	1.773	1.380	1.154	0.9893	0.8569	0.7444	0.5568	0.3999	0.2620	0.1370		
0.1	500	1.850	1.387	1.136	0.9590	0.8219	0.7088	0.5276	0.3837	0.2636	0.1598		
0.1	5,000	1.877	1.388	1.127	0.9477	0.8097	0.6970	0.5187	0.3793	0.2642	0.1660		
		$i=1$	2	3	4	5	10	15	20	30	50	70	90
0.05	50	1.882	1.586	1.415	1.291	1.192	0.847	0.605	0.399	0.019			
0.05	100	1.974	1.628	1.433	1.294	1.183	0.816	0.574	0.383	0.075			
0.05	500	2.068	1.664	1.444	1.290	1.168	0.786	0.548	0.372	0.106			
0.05	5,000	2.125	1.682	1.446	1.283	1.158	0.768	0.535	0.365	0.113			
0.02	100	2.035	1.799	1.667	1.572	1.497	1.247	1.082	0.953	0.743	0.340	0.086	
0.02	500	2.309	1.971	1.788	1.659	1.559	1.239	1.041	0.894	0.673	0.368	0.143	
0.02	5,000	2.436	2.040	1.830	1.685	1.574	1.226	1.018	0.868	0.648	0.360	0.159	
0.01	200	2.179	1.963	1.843	1.758	1.692	1.476	1.338	1.234	1.072	0.835	0.646	0.477
0.01	500	2.420	2.136	2.981	1.873	1.788	1.519	1.352	1.228	1.042	0.784	0.595	0.441
0.01	5,000	2.646	2.284	2.092	1.959	1.857	1.539	1.349	1.210	1.010	0.745	0.562	0.420

Table 4. Relative advantage Q of unequal proportion presented as a percentage greater than equality of the two methods

N	$\Sigma p_i^2 = d$					
	0.5	0.2	0.1	0.05	0.02	0.01
10	4.1	14.5	–	–	–	–
20	2.8	7.6	16.0	–	–	–
50	2.0	4.6	7.5	13.3	–	–
100	1.6	3.5	5.3	8.0	17.4	–
200	1.3	2.9	4.1	5.8	9.6	17.6
500	1.1	2.3	3.2	4.2	6.1	8.6
1,000	1.0	1.9	2.7	3.5	4.8	6.2
2,000	0.9	1.8	2.4	3.0	3.9	4.8
5,000	0.8	1.5	2.0	2.5	3.2	3.8

of weighted deployment will actually be higher than predicted by the values in the table.

Discussion

The discussion in this paper about proportion-dependent disadvantages connects to clonal forestry. However, exactly the same formulations can be used for many other situations of interest to the forest tree breeder like selfing in clonal orchards (Lindgren and Matheson 1986), effective family number in the breeding population (Lindgren 1986), inbreeding in a seedling seed orchard, inbreeding in natural regeneration or seed collection from a stand created from a clonal plantation, or from plantations with a family structure, etc.

We have solved the problem of optimizing the genetic gain at a given value of diversity-related "disadvantage". Thus, it is possible to state that the solutions give the highest possible gain under certain politically or environmentally or otherwise formulated diversity-restrictions. It is possible to see the problem in a reversed way: an optimal solution can be found which produces as much gain as the truncation selection, but at the same time minimizes the disadvantage connected to reduction in diversity. Our optimal solutions have this characteristic also. Actually, compared to the truncation case, there are a number of deployment solutions which combine a higher genetic gain with a wider diversity.

From a practical point of view, very low proportions are difficult to handle and may be uneconomical. The lowest ranked of the deployed clones are expected to be within a reasonably narrow range of genetic values. If interpolations in the tables we have supplied are used rather than computer programs, there are also difficulties in calculating the optimal proportions for about the lowest ranking 15% of the deployed clones. Equal deployment for these clones at the bottom of the included

list could be a reasonable simplification. If the optimization is actually made by a computer, it may be useful to avoid the lowest values by setting the n value of Eq. (13) at some number of entries above what is necessary to get the lowest $g_i > 0$.

The efficiency of the procedure to utilize unequal deployment proportional to genetic values is highest for small N and d (Table 4). The efficiency is high and rather constant in the range 14%–18%, when $1/Nd$ is constant 0.5, i.e. when a large share of the available clones are going to be deployed.

Clonal forestry often starts with many thousands of clones and aims at using only a small percentage of them in a large scale. The proposed algorithms are not much superior to truncation selection if only a small fraction of the clones are deployed. However, the philosophy of clonal testing is usually stepwise, with many clones being discarded during early rounds of testing before final deployment. Of those remaining and tested in later-founded field tests, a rather large fraction may be deployed. It may be argued that the normality may be skewed if there are several cycles of selection, but truncation selection makes a skew distribution only if it is very accurate, and it is not likely that it is economical to get such accurate test results in the first cycles of testing.

Probably the optimum is a rather smooth one, thus minor deviations from the solutions suggested here may not mean large losses in efficiency. It is difficult to carry out the exact calculations for a real case, and it is also likely that there are practical constraints, e.g., the physical availability of some of the high-ranking clones may be low. In such a situation, a practical way to handle the situation may be that a breeder makes a suggestion of what p_i to use, which is similar to what our tables suggest, and then calculates the $\Sigma p_i g_i$, Σp_i^2, $\Sigma g_i/n_e$ and the efficiency measure Q for the suggestion. If d is reasonably high and if the Q is not much below the maximum one according to Table 4, the suggestion can be regarded as satisfactorily efficient. If the efficiency is considerably below the maximum, it is recommended to adjust the representation of the clones to improve efficiency.

There are some noteworthy features of the relationships between values, which may have a deeper meaning, and indicate that more generalized expressions may be found. The f_i-values (disproportion coefficients) do not change much by N (cf Table 3). Study the f_i values for $i = n_e$. They are always close to 0.4. The range found in the Table is $0.47 > f_{n_e} > 0.35$. Thus, the last clone on the list in a truncation selection will get its proportion reduced to around 40% almost irrespective of the values of d and N (cf Table 2). n is always in the range $0.8\, n_e \geqq n \geqq 0.4\, n_e$, thus the list of deployed clones will increase approximately 60%, if optimal selection is applied over a wide range of N and d.

Acknowledgements. Part of the work has been supported by funds from the Swedish Council of Agricultural Sciences.

References

Harter HL (1970) Order statistics and their use for testing and estimation, vol 2. Estimates based on order statistics of samples from various populations. Aerospace Research Labs, US Air Force, Washington

Lindgren D (1986) How should breeders respond to breeding values? Conference proceeding from a joint meeting of IUFRO WP on breeding theory, progeny testing and seed orchards. October 13–17. Virginia, pp 361–372

Lindgren D, Matheson AC (1986) Increasing the genetic quality of seed from seed orchards by using the better clones in higher proportions. Silvae Genet 35:173–177

Luenberger DG (1973) Introduction to linear and nonlinear programming. Addison Wesley. Reading

Muller-Starck G, Gregorius HR (1986) Monitoring genetic variation in forest tree populations. Proceedings from 18th IUFRO World Congress, Division 2, vol. II. Yugoslav IUFRO World Congress Organizing Committee, Yugoslavia, pp 589–599

Royston JP (1982) Expected normal order statistics (exact and approximate). Appl Stat 31:161–165