

THE ESTIMATION AND USE OF SELECTIVE VALUES IN PREDICTING POPULATION CHANGE *

R. W. ALLARD, JAMES HARDING and CONRAD WEHRHAHN †
University of California, Davis, California

Received 26.ix.65

1. INTRODUCTION

In recent years there have been a number of attempts to measure the relative selective (adaptive, reproductive, fitness) values of different genotypes within natural populations. Examples include the investigations of Dobzhansky (1956) on inversion polymorphisms in *Drosophila psuedoobscura*, of Lewontin and White (1960) on inversion polymorphisms of *Moraba scurra*, and the studies of Imam and Allard (1965) on polymorphisms in wild oats, *Avena fatua*. Although such field investigations have made it clear that different genotypes differ widely in reproductive capacity, precise quantitative information concerning selective values remains small and information about the manner in which selective values vary from place to place and season to season is almost non-existent. There are many practical difficulties in obtaining the required estimates of these and the other parameters which enter the equations of genetic change. Until adequate estimates are available it will not be possible to utilize fully the extensive mathematical theory of population dynamics which already exists and it is possible that pitfalls in the theory may remain undiscovered. Experiments with populations in which genetic change is solely a function of competition between homozygous genotypes have a number of advantages for the estimation of selective values and for the estimation of variances in these values due to sampling error and due to environmental fluctuations. The purpose of this paper is to report estimates of selective values in several populations composed of mixtures of pure lines.

The experimental populations to be considered have the following features in common: (1) a mixture consisting of known proportions of two or more homozygous genotypes of barley, wheat, or lima beans was seeded in a plot, grown to maturity and the seed of the plot harvested in mass; (2) a random sample of seeds was drawn from the harvest in each generation to establish another plot in the next generation; (3) the integrity of the pure lines was protected throughout each experiment by removing the hybrids which were produced by occasional intercrossing between constituent pure lines; (4) progressive changes in the populations were followed by conducting a generation-by-generation census. In experiments with this structure the various factors

* This investigation was supported in part by grants from the National Science Foundation (G14991) and the National Institutes of Health (GM10476).

† Present address: University of Saskatchewan, Saskatoon, Canada.

which affect the competitive ability of any particular genotype for any two successive generations can be summarised in a single numerical value which relates frequency in any generation n to frequency in generation $n+1$. It is convenient to express this numerical value relative to some standard. There are, however, a number of possible standards, each of which leads to a different selection model. Since the numerical results are not the same for the different models, problems of estimation will be discussed before considering numerical estimates of selective values and their implications in population change.

2. THE ESTIMATION PROCEDURE

Let f_i^n be the observed frequency of the i th genotype in generation n , let w_i^n be the selective value of the i th genotype in transition from generation n to generation $n+1$, and let N^{n+1} be the total number of individuals sampled in generation $n+1$. Suppose that the f_i^n have been estimated from census data for genotypes $i = 1, 2, \dots, k$. Assuming multinomial sampling and using previous observed frequencies as current expected frequencies, it can be shown that

$$E(f_i^{n+1}) = w_i^n f_i^n / D^n \quad (1)$$

where

$$D^n = \sum_i^k w_i^n f_i^n. \quad (2)$$

The likelihood function is

$$L = \text{const.} \prod_i^k [E(f_i^{n+1})]^{N^{n+1} f_i^{n+1}} \quad (3)$$

and

$$\frac{\partial \log L}{\partial w_i^n} = \frac{N^{n+1} f_i^{n+1}}{w_i^n} - \frac{f_i^n N^{n+1}}{D^n}. \quad (4)$$

The likelihood L , is a maximum when the values of $\hat{w}_1^n, \dots, \hat{w}_k^n$ are such that for all $i, i = 1, 2, \dots, k$,

$$\frac{\partial \log L}{\partial w_i^n} = 0, \quad (5)$$

which leads to the set of solutions

$$\hat{w}_i^n = f_i^{n+1} D^n / f_i^n, \quad i = 1, 2, \dots, k. \quad (6)$$

There are k selective values to be estimated whereas the census data for any two successive generations provide only $k-1$ degrees of freedom. Hence the equations (6) are linearly dependent and no unique set of solutions exists unless an appropriate side condition is specified. We shall consider three such side conditions which give rise to three models designated Models A, B, and C.

Model A. The condition which has most commonly been specified in similar problems is to set the selective value of some genotype, say w_1 , equal to unity, e.g., Wright and Dobzhansky (1946), Wallace (1956),

Jain and Allard (1960), Lewontin and White (1960), Allard and Workman (1963), Workman and Allard (1964). The first equation of (6) can then be solved for D^n and this value in turn substituted into the remaining $k-1$ equations. These $k-1$ equations are now linearly independent and unique solutions exist. The estimator which results is

$$\hat{w}_i^n = \frac{f_i^{n+1} f_1^n}{f_i^n f_1^{n+1}}. \quad (7)$$

Assuming that frequencies are distributed multinomially in each sample, the variance of an estimated selective value is given approximately by

$$\begin{aligned} \text{Var } w_i^n &\approx \left(\frac{\partial w_i^n}{\partial f_i^{n+1}} \right)^2 \text{Var } f_i^{n+1} + \left(\frac{\partial w_i^n}{\partial f_1^{n+1}} \right)^2 \text{Var } f_1^{n+1} + \\ &+ 2 \left(\frac{\partial w_i^n}{\partial f_i^{n+1}} \right) \left(\frac{\partial w_i^n}{\partial f_1^{n+1}} \right) \text{Cov} (f_i^{n+1}, f_1^{n+1}) + \left(\frac{\partial w_i^n}{\partial f_i^n} \right)^2 \text{Var } f_i^n + \\ &+ \left(\frac{\partial w_i^n}{\partial f_1^n} \right)^2 \text{Var } f_1^n + 2 \left(\frac{\partial w_i^n}{\partial f_i^n} \right) \left(\frac{\partial w_i^n}{\partial f_1^n} \right) \text{Cov} (f_i^n, f_1^n). \end{aligned} \quad (8)$$

It follows that

$$\begin{aligned} \text{Var } w_i^n &\approx \frac{(w_i^n)^2}{N^{n+1}} \left(\frac{1-f_i^{n+1}}{f_i^{n+1}} + \frac{1-f_1^{n+1}}{f_1^{n+1}} + 2 \right) + \\ &+ \frac{(w_i^n)^2}{N^n} \left(\frac{1-f_i^n}{f_i^n} + \frac{1-f_1^n}{f_1^n} + 2 \right), \end{aligned} \quad (9)$$

in which N^n and N^{n+1} are the numbers of individuals sampled in generations n and $n+1$, respectively. The samples taken in successive generations will often be equal in size in which case (9) reduces to

$$\text{Var } w_i^n \approx \frac{(w_i^n)^2}{N} \left(\frac{1-f_i^{n+1}}{f_i^{n+1}} + \frac{1-f_1^{n+1}}{f_1^{n+1}} + \frac{1-f_i^n}{f_i^n} + \frac{1-f_1^n}{f_1^n} + 4 \right). \quad (10)$$

Model A is inefficient because four different estimates, each of which is subject to sampling error, enter into equation (7). Another disadvantage of Model A results from the adoption of a single genotype as a standard, which might be expected to accentuate the effects of genotype-environmental interactions on the estimates of the other selective values. An alternative procedure is to make the population mean the standard against which the selective values are compared. Models B and C provide two ways in which this can be done.

Model B. Assume there exists a set of selective values, w_i^* , $i = 1, 2, \dots, k$, such that

$$E(f_i^{n+1}) = w_i^* f_i^n / D^n \quad (11)$$

for all n , $n = 1, 2, \dots, t-1$.

D^n is now given by

$$D^n = \sum_{i=1}^k w_i^* f_i^n, \quad n = 1, 2, \dots, t-1. \quad (12)$$

If we impose the condition that one of the w_i^* , say w_1^* , equals one, we can estimate \hat{w}_2^* , \hat{w}_3^* , ... \hat{w}_k^* by iteration. We select as a trial set the values $\hat{w}_1^* = 1$, $\hat{w}_2^* = 1$, ... $\hat{w}_k^* = 1$, which are substituted into equation (12) to obtain estimates of D^n for $n = 1, \dots, t-1$. These values are substituted into (6) to obtain estimates of the \hat{w}_i^n for all i and n . We then calculate

$$\hat{w}_i^* = \frac{\sum_{n=1}^{t-1} w_i^n I_i^{n+1}}{\sum_{n=1}^{t-1} I_i^{n+1}}. \quad (13)$$

An unweighted estimate of the selective values is obtained if we let $I_i^{n+1} = 1$ for all n , $n = 1, 2, \dots, t-1$. A weighted estimate is obtained if we let $I_i^{n+1} = 1/\text{Var } w_i^n$. The process is repeated until no significant changes occur in the estimates of the w_i^n and the \hat{w}_i^* .

The sampling variance for this estimator is given by

$$\begin{aligned} \text{Var } w_i^n \approx & \left(\frac{\partial w_i^n}{\partial f_i^{n+1}} \right)^2 \text{Var } f_i^{n+1} + \left(\frac{\partial w_i^n}{\partial f_i^n} \right)^2 \text{Var } f_i^n + \left(\frac{\partial w_i^n}{\partial D^n} \right)^2 \text{Var } D^n + \\ & + 2 \left(\frac{\partial w_i^n}{\partial f_i^n} \right) \left(\frac{\partial w_i^n}{\partial D^n} \right) \text{Cov} (f_i^n, D^n). \end{aligned} \quad (14)$$

For t large the variance of the w_i^* is expected to be negligible compared to the variance of the w_i^n and it appears acceptable to treat D^n as a linear function in f_i^n with coefficients w_i^* . Assuming the w_i^* are constants, the variance of D^n is given by

$$\text{Var } D^n = \sum_{i=1}^k [(w_i^*)^2 \text{Var } f_i^n] + \sum_{j=1}^k \sum_{i \neq j}^k [w_i^* w_j^* \text{Cov} (f_i^n, f_j^n)]. \quad (15)$$

Assuming once more that the w_i^* are constants, it can be shown that

$$\text{Cov} (f_i^n, D^n) = w_i^* \text{Var } f_i^n + \sum_{\substack{j=1 \\ j \neq i}}^k w_j^* \text{Cov} (f_i^n, f_j^n). \quad (16)$$

Substituting into (14), we obtain

$$\begin{aligned} \text{Var } w_i^n \approx & \frac{(D^n)^2 (1 - f_i^{n+1})}{f_i^{n+1} N^{n+1}} + \frac{(w_i^n)^2 (1 - f_i^n)}{f_i^n N^n} + \frac{(f_i^{n+1})^2}{(f_i^n)^2 N^n} \left[\frac{2D^n}{f_i^n} \left(\sum_{j \neq i}^k w_j^* f_i^n f_j^n \right) \right. \\ & \left. - 2D^n w_i^* (1 - f_i^n) + \sum_{i=1}^k (w_i^*)^2 f_i^n (1 - f_i^n) - \sum_{j=1}^k \sum_{i \neq j}^k w_i^* w_j^* f_i^n f_j^n \right]. \end{aligned} \quad (17)$$

The average sampling error variance for the selective value, w_i , of the i th genotype can be estimated by substituting $\text{Var } w_i^n$ for w_i^n and $\text{Var } w_i$ for w_i^* in (13).

Model C. Another alternative is to adopt the mean selective value, D^n , as a standard to which all selective values are compared. If the condition is imposed that D^n is a specified constant, equation (6) can be solved directly. It is convenient to let $D^n = 1$, in which case

$$\hat{w}_i^n = \frac{f_i^{n+1}}{f_i^n}. \quad (18)$$

The variance of the \hat{w}_i is now the variance of the ratio of the random variables f_i^{n+1} and f_i^n , which have been estimated from different samples and are statistically independent. This variance is given by

$$\text{Var } w_i^n \approx (w_i^n)^2 \left(\frac{1-f_i^n}{f_i^n N^n} + \frac{1-f_i^{n+1}}{f_i^{n+1} N^{n+1}} \right). \quad (19)$$

Since D^n is constant and the f_i^n change over generations (unless all w_i happen to be 1), the w_i^n must change over generations. It follows that the values of the w_i^n in each generation are functions of the composition of the population in that generation. It is expected that there will be a general increase in population fitness over time. As a consequence estimated selective values should be higher in the first few generations of an experiment than in later generations. Thus, when the average selective value of a genotype is calculated over the course of an experiment, fluctuations in the first few generations may have a disproportionate effect on the average selective value. Also the observed variances of the fitness values may be inflated due to the general increase in population fitness over generations and the consequent general decrease in genotypic selection values.

Numerical evaluation of the models

In an attempt to determine the reliability of the estimators derived in the previous section, a Monte Carlo investigation was conducted on a digital computer. A hypothetical 10 genotype population was simulated in which genotypic frequencies in generations n and $n+1$ were assumed to be exactly those given in the second and third columns of table 1. The "true" selective values are therefore those given in column 4 of this table. A pair of samples, each of size 500, was drawn from generation n and from generation $n+1$ and the numbers obtained substituted into the appropriate equations of the previous section to obtain estimates of selective values. Results based on the means and standard deviations of 1000 repetitions of this process are given in columns 6 and 7 of table 1. Expected standard errors (approximate), based on equations (10) and (19) respectively, are given in column 5 of table 1. Only Models A and C were simulated because the iterative solutions required for Model B make simulation of that model prohibitively time consuming.

The results given in columns 4 and 6 of table 1 show that estimates of selective values obtained from the Model A estimator (7) were consistently higher than the "true" values. The deviations were often large, especially for genotypes which were present in low frequency. This is not surprising since the maximum likelihood estimator (7) has not been corrected for statistical bias. The estimates given by Model C were very close to the "true" values. Furthermore, some values are below and some are above the true values, indicating that this estimator is not subject to serious bias.

Columns 5 and 7 of table 1 show that the approximations of $\text{Var } \bar{w}_i^n$ given by equation 10 (Model A) and equation 19 (Model C) are in reasonable agreement with the numerical solutions for genotypes which occur with frequencies in excess of 3 or 4 per cent. If the numerical

TABLE 1

Numerical evaluation of Model A and Model C, using Monte Carlo simulation. Ten genotypes are assumed to have frequencies p_i^n and p_i^{n+1} in successive generations. The estimated selective values, \bar{w}_i^n , are calculated from (7) and (18) and are compared with empirical means, \bar{w}_i^n , from 1000 Monte Carlo runs. Population size is $N = 500$.

Assumed frequencies			Estimates		Monte-Carlo simulations	
i	p_i^n	p_i^{n+1}	\bar{w}_i^n	$\sqrt{\text{Var } \bar{w}_i^n}$	\bar{w}_i^n	$\sqrt{\text{Var } \bar{w}_i^n}$
Model A						
1	0.12	0.10	1.00	...	1.00	...
2	0.19	0.16	1.01	0.247	1.05	0.270
3	0.24	0.30	1.50	0.341	1.54	0.346
4	0.02	0.01	0.60	0.348	0.69	0.505
5	0.08	0.09	1.35	0.391	1.40	0.415
6	0.15	0.16	1.28	0.320	1.32	0.352
7	0.03	0.01	0.40	0.220	0.44	0.257
8	0.01	0.01	1.20	0.793	1.59	1.466
9	0.06	0.07	1.40	0.440	1.47	0.461
10	0.10	0.09	1.08	0.303	1.12	0.326
Model C						
1	0.12	0.10	0.83	0.168	0.85	0.155
2	0.19	0.16	0.84	0.129	0.86	0.121
3	0.24	0.30	1.25	0.112	1.26	0.129
4	0.02	0.01	0.50	0.472	0.57	0.395
5	0.08	0.09	1.13	0.223	1.14	0.243
6	0.15	0.16	1.07	0.153	1.08	0.161
7	0.03	0.01	0.33	0.453	0.35	0.192
8	0.01	0.01	1.00	0.629	1.30	1.159
9	0.06	0.07	1.17	0.264	1.20	0.297
10	0.10	0.09	0.90	0.187	0.91	0.182

solutions are accepted as close estimates of the true variance, this indicates that the assumptions which were made in deriving equations (10) and (19) do not lead to serious error if frequencies are not low. It will be noted that the sampling variances which attend Model A are substantially larger than those for Model C. Thus Model A appears to lead to not only biased but also less efficient estimators of selective values than Model C.

3. APPLICATIONS TO EXPERIMENTAL DATA

The estimators of selective values derived earlier were applied to data from 16 separate experimental populations, among which nine

have previously been reported in the literature (Harlan and Martini, 1938; Laude and Swanson, 1943; Suneson and Wiebe, 1942; Suneson, 1949). More than 1000 estimates of selective values were made from the data, but only part of the results need be considered since a detailed account of a few experiments serves to illustrate the main findings.

An experiment conducted at Davis, California, by Suneson and

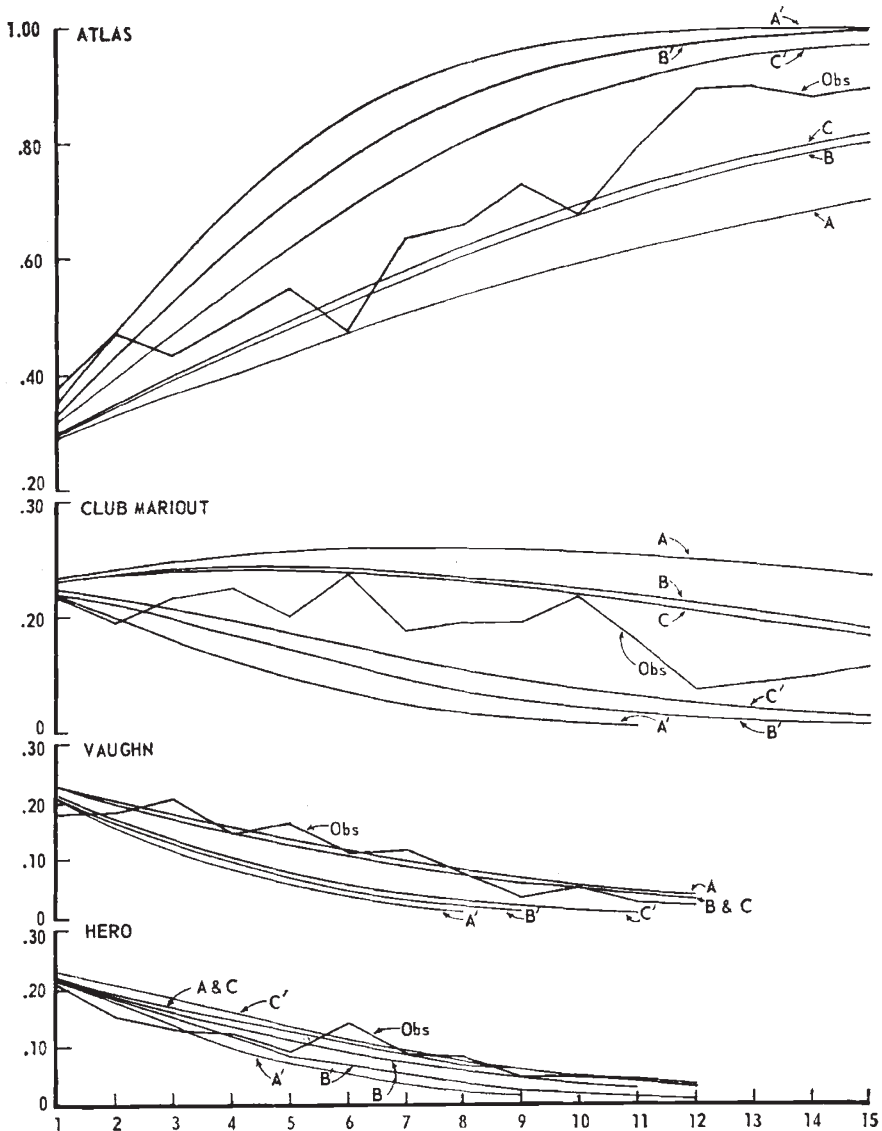


FIG. 1.—Observed frequencies (broken lines) of Atlas, Club Mariout, Vaughn, and Hero in the barley population of Suneson and Wiebe (1942) and Suneson (1949). A, B, and C represent deterministic runs using arithmetic means for w_i^p in formula (1) for Models A, B, and C respectively. A', B', and C' are based on the weighted means which appear in table 2.

Wiebe with barley will be considered in greatest detail because only four genotypes were in competition and presentation of the data is therefore simple. In this experiment Atlas soon became the dominant component of the mixture, primarily at the expense of Vaughn and Hero, which were promptly reduced to low frequency (fig. 1). The decline of Club Mariout from its initial frequency of 25 per cent. to about half that frequency was slow and erratic. Individual estimates of selective values \hat{w}_i^n for each genotype in transition from generation n to $n+1$, computed according to the appropriate estimators for Models A, B and C, are given in table 2. The means (weighted) of these values over the $t-1 = 15$ transition generations are also given. The standard errors given in the column labelled Exp (S.E.) are the means of the expected sampling errors (unweighted) of the individual w_i^n computed according to equations (9), (17), or (19). The empirical standard errors given in the column labelled Obs (S.E.) provide a measure of the dispersion of individual w_i^n about their means (\bar{w}_i).

The individual selective values given in table 2 show that Atlas had a higher selective value than the other genotypes. Averaged over generations the selective advantage of Atlas varied from 20 per cent. to slightly more than 40 per cent., depending on the estimator. Model A consistently gave the lowest and Model C consistently the highest mean estimates for the other three genotypes relative to Atlas. However, in single transition generations the Model B or C estimates of the selective value of Atlas were often lower than the Model A estimate.

The observed standard errors of these selective values vary from 0.079 to 0.307. In each of the 11 possible comparisons the observed standard error was larger than the corresponding standard error attributable to sampling. This indicates that the selective values are not constant but fluctuate from generation to generation. There was little apparent relationship between selective values and genotypic frequencies; hence, the main cause of these fluctuations is almost certainly season-to-season change in environmental conditions. Averaged over all genotypes and methods of estimation the observed standard error is 0.165 units larger than the sampling standard error. Averaged over methods of estimation the observed standard errors for Atlas, Club Mariout, Vaughn and Hero are, respectively, 0.058, and 0.206, 0.132, and 0.087 larger than expected. Averaged over genotypes the values for Models A, B and C are 0.162, 0.107, and 0.124. These results therefore indicate that the selective values of Atlas are the least and those of Club Mariout the most sensitive to fluctuations in the environment. This is biologically reasonable because Atlas is known to be an exceptionally good competitor under a variety of conditions and Vaughn and Hero are known to be poor competitors unless conditions are exceptionally favourable. Consequently, the selective values of Atlas are expected to be consistently high and those of Vaughn and Hero rather consistently low. Club Mariout, on the other hand, is a well adapted type which can compete favourably with Atlas in many

TABLE 2
 Individual estimates of w_i^* with means, \bar{w}_i (weighted), average sampling error, $Exp(S.E.)$, and observed standard error, $Obs(S.E.)$, from the barley population of Suneson and Wiebe (1942) and Suneson (1949). Calculations were made using Model A, Model B, and Model C estimators.

Variety	Model	Years														\bar{w}_i	Exp (S.E.)	Obs (S.E.)					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14				15				
Atlas	A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.070	...
	B	1.14	1.00	0.76	0.92	0.93	0.75	1.11	0.92	0.99	0.85	1.06	1.06	0.97	0.95	1.00	0.97	0.95	0.97	1.00	0.97	0.024	0.079
	C	1.51	1.25	0.91	1.13	1.11	0.87	1.33	1.04	1.10	0.92	1.17	1.13	1.01	0.98	1.03	1.01	0.98	1.01	1.03	1.01	0.023	0.084
Club Mariout	A	0.62	0.64	1.36	0.93	0.75	1.57	0.47	1.05	0.93	1.31	0.56	0.37	1.23	1.15	1.17	1.15	1.15	1.17	0.66	0.64	0.084	0.307
	B	0.71	0.64	1.03	0.86	0.69	1.17	0.53	0.96	0.92	1.11	0.59	0.39	1.20	1.09	1.14	1.09	1.14	0.73	0.73	0.70	0.070	0.258
	C	0.94	0.80	1.24	1.05	0.83	1.37	0.63	1.09	1.03	1.21	0.66	0.42	1.24	1.13	1.18	1.13	1.13	0.80	0.80	0.77	0.077	0.285
Vaughn	A	0.47	0.81	1.23	0.63	1.02	0.79	0.77	0.64	0.41	1.73	0.40	0.74	0.62	1.37	0.22	0.62	1.37	0.59	0.59	0.109	0.242	
	B	0.54	0.81	0.93	0.58	0.95	0.59	0.85	0.59	0.40	1.46	0.42	0.79	0.60	1.31	0.21	0.62	1.31	0.62	0.62	0.093	0.209	
	C	0.71	1.01	1.12	0.70	1.13	0.69	1.02	0.67	0.45	1.59	0.47	0.84	0.62	1.35	0.22	0.68	1.35	0.68	0.68	0.105	0.251	
Hero	A	0.55	0.59	0.92	0.84	0.68	1.71	0.46	0.89	0.54	1.04	0.71	0.67	0.54	1.02	0.48	0.63	1.02	0.63	0.63	0.116	0.224	
	B	0.63	0.60	0.69	0.78	0.63	1.27	0.51	0.82	0.53	0.88	0.75	0.71	0.53	0.97	0.46	0.67	0.97	0.67	0.67	0.104	0.175	
	C	0.83	0.75	0.84	0.95	0.76	1.49	0.61	0.93	0.59	0.96	0.83	0.76	0.55	1.00	0.48	0.76	1.00	0.76	0.76	0.117	0.200	

seasons but not in others. Hence its selective values are expected to be quite variable. The above results also indicate that Model A estimates are influenced more by genotype-environment interactions than Model B or C estimates.

To test the usefulness of the estimates of selective values obtained by the three models in predicting the course of population change, the values of the \bar{w}_i given in table 2 were substituted into the recursion equation (1), which was used to obtain the smooth curves (A', B' and C') plotted in fig. 1. Comparable values obtained from unweighted estimates are also plotted in these figures (A, B, and C). Chi-square

TABLE 3

Chi-square test of "goodness of fit" of the observed data to 6 deterministic models based on the recurrence formula (1). The weighted models are based on mean selective values weighted inversely to sampling variances.

Model	χ^2				
	Atlas	C. Mariout	Vaughn	Hero	Total
A Arithmetic	769.4	1280.6	111.7	58.6	2220.3
Weighted	915.3	26897.2	3233.8	1498.2	32544.5
B Arithmetic	235.2	594.4	82.8	68.2	980.7
Weighted	645.6	6366.5	2048.3	549.8	9610.1
C Arithmetic	183.2	498.6	80.3	72.1	834.2
Weighted	338.6	2187.3	678.8	49.5	3254.2

values for goodness of fit between observed and theoretical frequencies are given in table 3. Two features of these results stand out. First, although the estimates of selective values obtained by all three models predict the general course of change in genotypic frequencies, Model C estimates give the best and Model A estimates give the poorest fit to observed values. Apparently, therefore, mean population fitness is an operationally more reliable standard than the fitness of any single genotype, as has been argued by Li (1963*a, b*), Sanghvi (1963), and Dobzhansky (1964). Second, weighted estimates of the selective value of the "best" genotype relative to other genotypes tend to be too high and the opposite is the case for unweighted estimates. This result suggests that the weighting of selective values by the inverse of their variances overcorrects estimates of selective values. In general, unweighted estimates led to better agreement between predicted and observed changes than weighted estimates in the experiments analysed in this study. We were unable to find a method of weighting that gave means which led to better fits than those obtained using arithmetic means.

Another experiment that will be considered in some detail involved a mixture of eleven barley genotypes which was grown at ten locations

TABLE 4
 Summary of \bar{w}_t , $\text{Exp}(\text{S.E.})$, and $\text{Obs}(\text{S.E.})$ from the New York and Idaho populations of *Harlan and Martini (1938)*.

Genotypes	Model A			Model B			Model C			Final frequency
	\bar{w}_t	$\text{Exp}(\text{S.E.})$	$\text{Obs}(\text{S.E.})$	\bar{w}_t	$\text{Exp}(\text{S.E.})$	$\text{Obs}(\text{S.E.})$	\bar{w}_t	$\text{Exp}(\text{S.E.})$	$\text{Obs}(\text{S.E.})$	
	New York	1.00	0.234	...	1.00	0.160	0.332	0.87	0.137	
Trebi (+ Coast)	0.80	0.242	0.290	0.96	0.217	0.311	0.78	0.178	0.264	0.018
Gatami	0.90	0.248	0.306	1.14	0.224	0.293	0.92	0.180	0.230	0.104
Smooth Awn	0.70	0.258	0.339	0.90	0.264	0.346	0.72	0.214	0.309	0.006
Lion	0.45	0.257	0.167	0.29	0.168	0.164	0.30	0.171	0.155	...
Meloy	0.50	0.352	0.294	0.57	0.345	0.241	0.50	0.308	0.246	...
White Smyrna	0.62	0.201	0.394	0.84	0.179	0.384	0.69	0.150	0.339	0.068
Hanchen	0.52	0.249	0.329	0.64	0.244	0.376	0.54	0.212	0.341	0.004
Svanhals	0.26	0.138	0.128	0.42	0.161	0.227	0.38	0.150	0.228	...
Deficients	1.00	0.207	0.360	1.40	0.098	0.187	1.07	0.070	0.140	0.686
Manchuria										
Idaho										
Coast (+ Trebi)	1.00	0.153	...	1.00	0.089	0.146	1.02	0.089	0.164	0.420
Gatami	0.67	0.234	0.219	0.70	0.214	0.191	0.77	0.236	0.199	0.020
Smooth Awn	0.52	0.230	0.197	0.57	0.213	0.150	0.63	0.240	0.204	...
Lion	0.39	0.185	0.277	0.50	0.192	0.270	0.54	0.214	0.312	0.004
Meloy	0.52	0.246	0.279	0.58	0.238	0.247	0.64	0.265	0.288	0.016
White Smyrna	0.94	0.163	0.209	1.00	0.114	0.180	1.02	0.115	0.164	0.314
Hanchen	0.83	0.165	0.267	0.92	0.132	0.197	0.96	0.137	0.217	0.100
Svanhals	0.65	0.212	0.295	0.75	0.207	0.262	0.81	0.224	0.271	0.036
Deficients	0.42	0.236	0.236	0.44	0.223	0.314	0.47	0.245	0.256	0.004
Manchuria	0.36	0.187	0.299	0.49	0.201	0.297	0.52	0.219	0.346	0.006

across the northern United States (Harlan and Martini, 1938). Summaries of mean selective values and standard errors for two representative locations, Ithaca, New York, and Aberdeen, Idaho, are given in table 4. At Ithaca, Manchuria was the dominant variety and this was reflected in estimates of selective values obtained by the application of Model B and C. However, the estimated Model A selective value of Trebi, a variety which was reduced to low frequency in the population, was the same as that of Manchuria. This indicates that Model A estimates can lead to serious errors. At Aberdeen, Idaho, Coast and White Smyrna were co-dominant varieties and Hannchen also persisted in substantial frequency into late generations. This result is reflected by the selective values estimated by all three Models. Space does not permit presentation of the curves (120) showing the possible comparisons between observed and predicted changes in genotypic frequencies. However, the results were similar to those for the Suneson and Wiebe experiment in that Model B and C gave better fits than Model A, especially for the population grown at Ithaca. At Aberdeen the fit given by Model C was substantially superior to that given by Model B.

Averaged over all genotypes and methods of estimation the mean observed standard error was 0.069 larger than the mean sampling standard error at Ithaca, and 0.039 larger at Aberdeen. Thus the environment is apparently more stable at Aberdeen than at Ithaca. The comparable value for the Suneson and Wiebe experiment conducted in California was 0.165. The genotypes involved were different in the California experiment, so that direct comparisons cannot be made. However, the higher standard error observed in the latter experiment suggests that selective values fluctuate more widely in California as a result of season-to-season changes in environment than in the other locations.

Mean selective values and standard errors for four experiments involving species other than barley are summarised in table 5. The results follow the same pattern as those of the barley experiments and hence need not be considered in detail. One feature of the three experiments that were conducted with lima beans should be pointed out, however. A single genotype became predominant in two of the populations (Populations 52 and 74) within five generations. This was reflected in large differences between the selective values of the dominant genotypes (L 48 and L 106) and those of all other genotypes in the population. The differences were so large as to suggest general and overwhelming competitive superiority of the dominant genotype in each case. When these experiments were repeated in a different series of years the results of the second run were similar to those of the first run. However, in a third population (Population 14, fig. 2) the predominant genotype of the first run was rapidly reduced to low frequency in the second run and a genotype (L 92) which was reduced to very low frequency in the first run, became the most frequent genotype

TABLE 5
Summary of \bar{w}_i , Exp (S.E.), and Obs (S.E.) from the Samson and Wiebe (1942) experiment with wheat, and from unpublished data from two lima bean populations.

Genotypes	Model A		Model B		Model C			Final frequency		
	\bar{w}_i	Exp (S.E.)	Obs (S.E.)	\bar{w}_i	Exp (S.E.)	Obs (S.E.)	\bar{w}_i		Exp (S.E.)	Obs (S.E.)
Wheat:										
Poso	1.00	1.00	0.052	0.097	1.05	0.053	0.118	0.317
Baart	0.89	0.097	0.262	0.99	0.074	0.187	1.04	0.077	0.199	0.319
Pacific Bluestem	0.79	0.095	0.176	0.85	0.078	0.125	0.90	0.082	0.143	0.243
White Federation	0.67	0.086	0.246	0.75	0.071	0.214	0.80	0.076	0.223	0.001
Ramona	0.46	0.079	0.219	0.57	0.076	0.164	0.62	0.084	0.162	0.027
Lima Beans:										
Population 52										
L 48	1.00	1.00	0.014	0.048	1.05	0.013	0.110	0.868
L 5	0.46	0.042	0.116	0.48	0.035	0.102	0.60	0.044	0.204	0.010
L 37	0.11	0.028	0.021	0.10	0.024	0.009	0.16	0.038	0.024	...
L 88	0.31	0.040	0.198	0.35	0.037	0.190	0.40	0.045	0.287	0.004
L 126	0.48	0.046	0.294	0.51	0.041	0.273	0.70	0.054	0.312	0.116
Population 74										
L 106	1.00	1.00	0.012	0.062	1.12	0.011	0.091	0.912
L 59	0.36	0.028	0.200	0.40	0.025	0.177	0.47	0.031	0.260	0.006
L 96	0.21	0.030	0.107	0.20	0.027	0.099	0.28	0.037	0.124	0.007
L 119	0.09	0.019	0.124	0.09	0.017	0.119	0.10	0.021	0.153	...
L 121	0.54	0.034	0.234	0.57	0.029	0.206	0.69	0.036	0.273	0.073

in the second run. It is clear that fluctuations in selective values are sufficiently large so that even the existence of large mean differences in selective values does not guarantee determinism in the course of genetic change in populations exposed to variable environments.

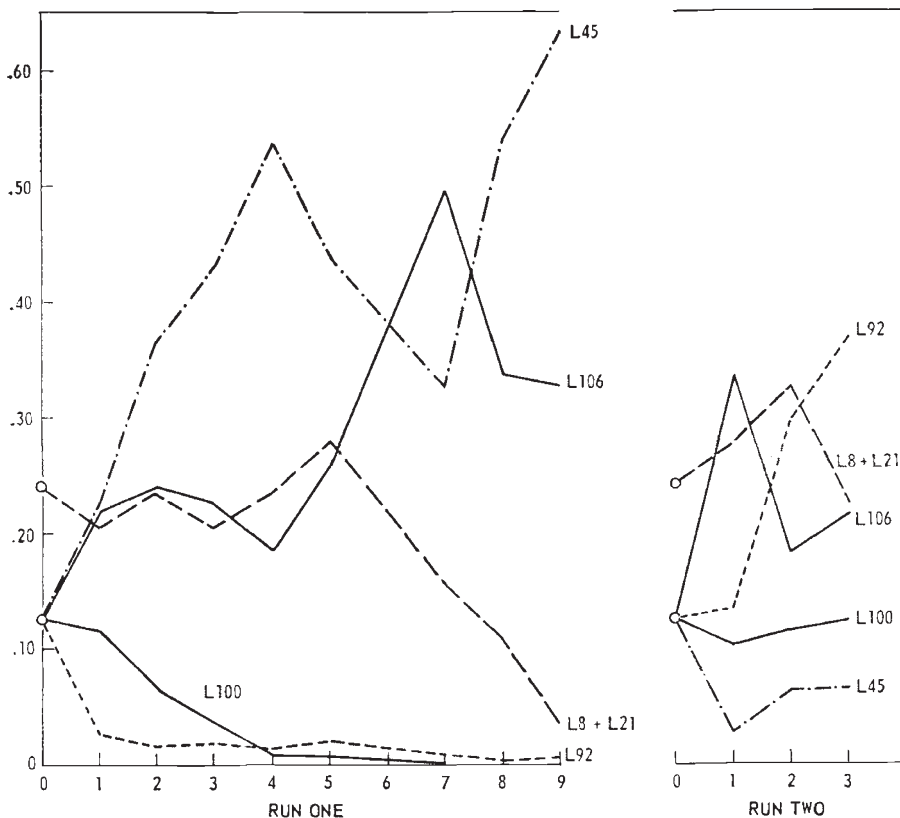


FIG. 2.—Observed frequency curves for two runs of lima bean population 14. One genotype is not shown in the figure because it was rapidly eliminated in both cases.

4. DISCUSSION

There are a number of advantages in estimating selective values from populations with the structural simplicity of those which are the concern of this discussion. Reproduction is by complete self-fertilisation and no assumptions need therefore be made concerning the mating system. The competing types remain the same throughout the experiment so that assumptions concerning constancy of the genetic entities under investigation are unnecessary. Complications imposed by overlapping of generations, fluctuations in population size and migration from other populations are also avoided. Finally, the environments in which the experiments are conducted are the natural ones for the species involved and they presumably vary in space and time in response to many of the forces which affect the environments

encountered by natural populations of plants. Thus experiments of this type circumvent many of the obstacles that complicate the estimation of selective values in naturally occurring populations.

The results obtained when the selective values of various genotypes were estimated, generation by generation, followed a similar pattern for 16 different populations consisting of mixtures of pure lines. In the majority of the populations the average selective value of the most successful genotype was more than 10 per cent. greater than that of the next most successful type and in nearly every population some genotypes had selective values indicating near lethality under population conditions. The numerical values obtained indicate, therefore, that the genotypes in these populations have a wide range of selective values under population conditions, even though all of the genotypes produce vigorous individuals when they are grown in pure stands.

The estimates of selective values were not the same for the three different estimators. One of these estimators is based on a model in which selective values of other genotypes in a population are compared with the selective value of a standard genotype (Model A). The other two estimators are based on models (Models B and C) in which the population mean is the standard against which all selective values are compared. Model B treats the selective values as constants while in Model C they are treated as variables. These two models have theoretical advantages, both statistical and biological, over Model A and in practice the numerical estimates of selective values obtained from them proved to be superior for purposes of prediction.

It is widely accepted that every species and population inevitably lives in an environment which is neither constant in time nor uniform in space and that the "relative fitness of different genotypes varies in different environments" (Dobzhansky, 1964). However, the magnitude of variations in selective values and the manner in which they are distributed about their means is far from well understood. The present results show that the selective value of a single genotype can vary over a wide range numerically, *e.g.*, the values for Vaughn in the Suneson and Wiebe experiment (table 2, Model C) varied from a low of 0.22 to a high of 1.59 in 15 transition generations. If individual selective values from Suneson and Wiebe's experiment are expressed as deviations from their arithmetic means and their means are all scaled to unity, the selective values are distributed as shown in fig. 3. In this case, therefore, the selective values appear to be more or less normally distributed.

The present analysis permitted the partitioning of the variance in selective values into a component due to sampling error and one due to environmental fluctuations. This partitioning cannot, of course, be applied when Model C estimates are used because the magnitude of these estimates tends to decrease with increasing population fitness and thus the estimated variance due to environmental fluctuations is

inflated. The present results show that in the case of the Suneson and Wiebe experiment (table 2) sampling error was of relatively little importance. For example, the selective value of Club Mariout in table 2, Model B, has a sampling variance of $0.07^2 = 0.0049$ and an observed variance of $0.258^2 = 0.0665$. The estimated variance due to environmental fluctuations of $0.0665 - 0.0049 = 0.0616$ is therefore 12.6 times as large as the sampling variance. However, in the case of the Harlan and Martini experiment (table 4) sampling errors are the

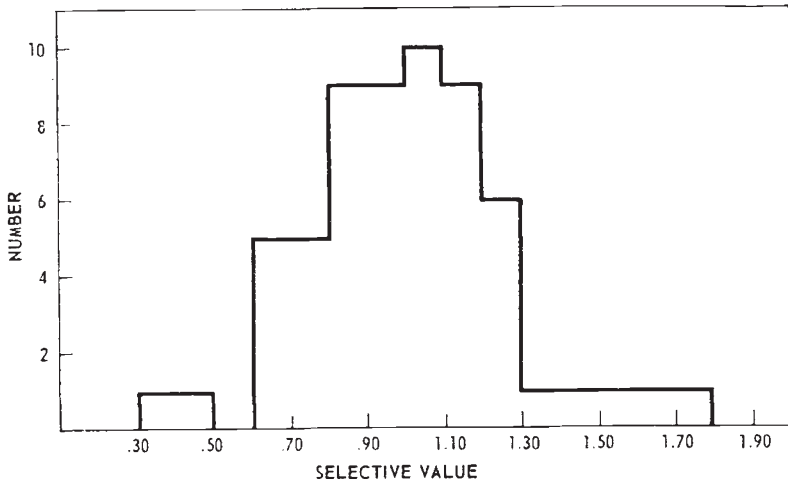


FIG. 3.—Distribution of individual selective values (Model C) from the population of Suneson and Wiebe (1942) and Suneson (1949). Model C estimates from table 2 are expressed as deviations from their respective means and the means are all scaled to unity.

main source of variation in selective values. This points out the need for adequate sized samples in populations composed of numerous genotypes.

Because the populations under consideration in this investigation were structurally simple, it might be expected that the estimation of parameters affecting genotypic change would also have been simple and that the subsequently predicted course of genotypic change would be in close agreement with observed changes. It was found that, although the general course of genotypic change could be predicted, the accuracy of such prediction left much to be desired. Fluctuations in selective values appeared to introduce a large stochastic element into the process of genotypic change. The fact that difficulties were encountered in studies involving simple populations suggests that prediction will be difficult in populations in which many factors other than selection and chance play a role in population change.

5. SUMMARY

Methods were derived for the estimation of selective values in populations composed of mixtures of pure lines and were used to make

more than 1,000 estimates of selective values in 16 different populations. The main findings were that:

- (1) average selective values differ widely for different genotypes in a population,
- (2) selective values fluctuate widely from year to year within individual populations,
- (3) the general course of genotypic change can be predicted from estimated selective values,
- (4) the accuracy of prediction was not altogether satisfactory because of the marked season-to-season variation in selective values,
- (5) two of the estimators derived were found to give results that were superior for purposes of prediction to those obtained from the third estimator.

6. REFERENCES

- ALLARD, R. W., AND WORKMAN, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution*, 17, 470-480.
- DOBZHANSKY, TH. 1956. XXV. Genetic changes in populations of *Drosophila pseudoobscura* and *Drosophila persimilis* in some localities in California. *Evolution*, 10, 82-92.
- DOBZHANSKY, TH. 1964. How do the genetic loads affect the fitness of their carriers in *Drosophila* populations? *Am. Nat.*, 98, 151-166.
- HARLAN, H. V., AND MARTINI, M. L. 1938. The effect of natural selection in a mixture of barley varieties. *Jour. Agric. Res.*, 57, 189-199.
- IMAM, A. G., AND ALLARD, R. W. 1964. Population studies in predominantly self-pollinated species. VI. Genetic variability between and within natural populations of wild oats, *Avena fatua* L., from differing habitats in California. *Genetics*, 51, 49-62.
- JAIN, S. K., AND ALLARD, R. W. 1960. Population studies in predominantly self-pollinated species. I. Evidence for heterozygote advantage in a closed population of barley. *Proc. Nat. Acad. Sci.*, 46, 1371-1377.
- LAUDE, H. H., AND SWANSON, A. F. 1943. Natural selection in varietal mixtures of winter wheat. *Jour. Amer. Soc. Agron.*, 34, 270-274.
- LEWONTIN, R. C., AND WHITE, M. J. D. 1960. Interaction between inversion polymorphisms of two chromosome pairs in the grasshopper, *Moraba scurra*. *Evolution*, 14, 116-129.
- LI, C. C. 1963a. Decrease of fitness upon inbreeding. *Proc. Nat. Acad. Sci.*, 49, 439-445.
- LI, C. C. 1963b. The way the load ratio works. *Am. Jour. Human Gen.*, 15, 316-231.
- SANGHVI, L. D. 1963. The concept of the genetic load: A critique. *Am. Jour. Human Gen.*, 15, 298-309.
- SUNESON, C. A. 1949. Survival of four barley varieties in a mixture. *Agron. Jour.*, 41, 459-461.
- SUNESON, C. A., AND WIEBE, G. A. 1942. Survival of barley and wheat varieties in mixtures. *Jour. Amer. Soc. Agron.*, 34, 1052-1056.
- WALLACE, BRUCE. 1956. Studies on irradiated populations of *Drosophila melanogaster*. *Jour. Genetics*, 54, 280-293.
- WORKMAN, P. L., AND ALLARD, R. W. 1964. Population studies in predominantly self-pollinated species. V. Analysis of differential and random viabilities in mixtures of competing pure lines. *Heredity*, 19, 181-189.
- WRIGHT, S., AND DOBZHANSKY, TH. 1946. Genetics of natural populations. XII. Experimental reproduction of some changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics*, 31, 125-156.