Acknowledgments.—To Mr Norman D'Arcy and to the Berkshire, Buckinghamshire and Oxfordshire Naturalists' Trust I should like to express my gratitude for their kind permission to work in various parts of Cothill Marsh.

I am most grateful to Professor E. B. Ford, F.R.S., for his generous encouragement and advice at all stages of this study. Mr P. T. Handford and Mr A. M. Shapiro helped energetically in the field and without their assistance much of the work would have been impossible. Dr E. R. Creed and Professor P. M. Sheppard, F.R.S., read this paper in typescript and I am indebted to them for many helpful suggestions.

5. REFERENCES

COOK, L. M. 1961. Food-plant specialisation in the moth *Panaxia dominula*. *Evolution*, *15*, 478-485.

DOWDESWELL, W. H., FISHER, R. A., AND FORD, E. B. 1949. The quantitative study of populations in the Lepidoptera. 2. *Maniola jurtina* L. *Heredity*, *3*, 67-84.

FISHER, R. A., AND FORD, E. B. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, *1*, 143-174.

FORD, E. B. 1960. 7th edn. *Mendelism and Evolution*. Methuen, London.

FORD, E. B. 1965. 2nd edn. *Ecological Genetics*. Methuen, London.

FORD, E. B. 1967. 2nd edn. *Moths*. Collins, London.

FORD, E. B., AND SHEPPARD, P. M. 1969. The *medionigra* polymorphism of *Panaxia dominula* L. (Lep.). *Heredity*, *24*, 561-569.

SHEPPARD, P. M. 1951. A quantitative study of two populations of the moth *Panaxia dominula* (L.). *Heredity*, *5*, 349-378.

SHEPPARD, P. M. 1953. Polymorphism and population studies, in *Symposia. Soc. Exp. Biol.*, *No. 7*, 274-289.

SHEPPARD, P. M. 1956. Ecology and its bearing on population genetics. *Proc. Roy. Soc. B.*, *145*, 308-315.

SHEPPARD, P. M., AND COOK, L. M. 1962. The manifold effects of the *medionigra* gene in the moth *Panaxia dominula* and the maintenance of a polymorphism. *Heredity*, *17*, 415-426.

WILLIAMSON, M. H. 1960. On the polymorphism of the moth *Panaxia dominula* (L.). *Heredity*, *15*, 139-151.

ENVIRONMENTAL AND GENOTYPE-ENVIRONMENTAL COMPONENTS OF VARIABILITY. VII. SIMULTANEOUS PREDICTION ACROSS ENVIRONMENTS AND GENERATIONS

J. L. JINKS and JEAN M. PERKINS

Department of Genetics, University of Birmingham

1. THE BASIS OF PREDICTION

IN this series of papers (Bucio Alanis, 1966; Bucio Alanis and Hill, 1966; Perkins and Jinks, 1968a and b; Bucio Alanis, Perkins and Jinks, 1969; Perkins, 1970) we have shown that the means of a pair of inbred lines (P_1 and P_2) and their F_1 in any environment j, can be described by the equations,

$$\bar{P}_{1j} = \mu + [d] + (1 + \beta_d)\epsilon_j$$
$$\bar{P}_{2j} = \mu - [d] + (1 - \beta_d)\epsilon_j$$
$$\text{and} \quad \bar{F}_{1j} = \mu + [h] + (1 + \beta_h)\epsilon_j$$

The conditions under which these equations hold are that

$$g_d = \beta_d \epsilon_j$$
$$\text{and} \quad g_h = \beta_h \epsilon_j$$

that is, the genotype-environment interactions (g) are linear functions (β) of the additive environmental component (ϵ).

All the parameters in the equations can be estimated by observing the inbred and F_1 families in a number of environments. Since five of these parameters, μ, $[d]$, $[h]$, β_d and β_h are constants for a particular pair of inbred lines and their F_1, we can predict the means of these families and their standard errors for any value of the variable ϵ_j, whether this is a theoretical value or one obtained by growing the two inbreds in a particular environment. Such predictions make no assumptions about the kind of gene action or interaction involved although they assume that the linear relationship may be extrapolated to the new environments, i.e., the new values of ϵ_j.

To predict the means of generations other than parents and F_1's, however, it is necessary to assume that the additive-dominance model of gene action is adequate (Bucio Alanis, Perkins and Jinks, 1969). If this assumption holds the expected mean of any generation that can be derived following a cross between two inbred lines can be written in terms of the parental and F_1 parameters and hence we can predict these means from the values of the parameters obtained from observations on the parental and F_1 families alone.

In the cross between varieties 1 and 5 of *Nicotiana rustica* the assumptions underlying the adequacy of the model have been shown to hold for the character plant height and the means of F_2, B_1 and B_2 families have been satisfactorily predicted in each of 16 environments from estimates of the parameters obtained from parental and F_1 families grown in these environments. To test the powers of prediction further, 21 generations derived from this cross have been grown in two environments which were two locations in the season 1965. The structure of the experiment, the means and standard errors of the generations in the two environments and their expectations in terms of the parameters of the model are given in table 1. Two replicate blocks were grown in each environment and all plants were individually randomised.

2. PREDICTION IN TWO ENVIRONMENTS

The estimates of the five constants of the model obtained from the parental and F_1 generations of the cross 1×5 grown in 16 environments (which do not include the two environments of the present experiment) are

$$\hat{\mu} = 44 \cdot 20$$
$$[\hat{d}] = 2 \cdot 89$$
$$[\hat{h}] = 5 \cdot 49$$
$$\hat{\beta}_d = 0 \cdot 35$$
$$\hat{\beta}_h = 0 \cdot 24 \quad \text{(Bucio Alanis, Perkins and Jinks, 1969)}.$$

To predict the means of the 21 generations in table 1 we require an estimate of the variable ϵ_j for each of the two environments. This is obtained from the means of the parental families (P_1 and P_2) in environments 1 and 2 as,

$$\hat{\epsilon}_1 = \tfrac{1}{2}(60 \cdot 46 + 49 \cdot 92) - \mu = 10 \cdot 99$$
$$\hat{\epsilon}_2 = \tfrac{1}{2}(49 \cdot 20 + 40 \cdot 96) - \mu = 0 \cdot 88$$

Environment 2 is clearly very close to the average of the 16 environments. Environment 1, on the other hand, is well above average, indeed

TABLE 1

The mean plant height (o) and standard errors (So) of the 21 generations when grown in two environments, the structure of the experiments and the model for predicting their expected means. (NF = number of families and NS = number of sibs per family in each of the two replicate blocks)

Generations	Environment 1				Environment 2				Model
	o	So	NF	NS	o	So	NF	NS	
P_1	60·46	0·44	1	80	49·20	0·82	1	40	$\mu + [d] + (1 + \beta_d)\epsilon_j$
P_2	49·92	0·41	1	50	40·96	0·71	1	25	$\mu - [d] + (1 - \beta_d)\epsilon_j$
F_1	61·30	0·33	1	70	50·74	0·80	1	36	$\mu + [h] + (1 + \beta_h)\epsilon_j$
F_2	57·24	0·34	1	150	48·56	0·67	1	70	$\mu + \frac{1}{2}[h] + (1 + \frac{1}{2}\beta_h)\epsilon_j$
B_1	61·53	0·42	1	80	48·96	0·90	1	35	$\mu + \frac{1}{2}[d] + \frac{1}{2}[h] + (1 + \frac{1}{2}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
B_2	55·96	0·39	1	80	44·76	0·84	1	35	$\mu - \frac{1}{2}[d] + \frac{1}{2}[h] + (1 - \frac{1}{2}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
F_3	57·55	0·37	40	5	45·83	0·64	15	5	$\mu + \frac{1}{4}[h] + (1 + \frac{1}{4}\beta_h)\epsilon_j$
$F_2 \times P_1$	62·10	0·27	40	5	50·89	0·57	15	5	$\mu + \frac{1}{2}[d] + \frac{1}{2}[h] + (1 + \frac{1}{2}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
$F_2 \times P_2$	56·10	0·27	40	5	45·53	0·53	15	5	$\mu - \frac{1}{2}[d] + \frac{1}{2}[h] + (1 - \frac{1}{2}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
$F_2 \times F_1$	58·93	0·31	40	5	48·55	0·60	15	5	$\mu + \frac{1}{2}[h] + (1 + \frac{1}{2}\beta_h)\epsilon_j$
$F_2 bip$	59·74	0·34	20	10	48·04	0·74	10	5	$\mu + \frac{1}{2}[h] + (1 + \frac{1}{2}\beta_h)\epsilon_j$
B_{11}	62·00	0·45	14	5	49·65	0·33	5	5	$\mu + \frac{3}{4}[d] + \frac{1}{4}[h] + (1 + \frac{3}{4}\beta_d + \frac{1}{4}\beta_h)\epsilon_j$
B_{12}	58·08	0·48	14	5	48·19	1·06	5	5	$\mu - \frac{1}{4}[d] + \frac{3}{4}[h] + (1 - \frac{1}{4}\beta_d + \frac{3}{4}\beta_h)\epsilon_j$
B_{22}	53·20	0·43	14	5	43·32	0·92	5	5	$\mu - \frac{3}{4}[d] + \frac{1}{4}[h] + (1 - \frac{3}{4}\beta_d + \frac{1}{4}\beta_h)\epsilon_j$
B_{21}	61·16	0·48	14	5	51·80	1·06	5	5	$\mu + \frac{1}{4}[d] + \frac{3}{4}[h] + (1 + \frac{1}{4}\beta_d + \frac{3}{4}\beta_h)\epsilon_j$
$B_1 \times F_1$	59·67	0·37	14	10	49·36	0·62	10	5	$\mu + \frac{1}{4}[d] + \frac{1}{2}[h] + (1 + \frac{1}{4}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
$B_2 \times F_1$	56·92	0·40	14	10	47·87	0·63	10	5	$\mu - \frac{1}{4}[d] + \frac{1}{2}[h] + (1 - \frac{1}{4}\beta_d + \frac{1}{2}\beta_h)\epsilon_j$
$B_1 bip$	61·30	0·37	7	15	50·25	1·00	5	5	$\mu + \frac{1}{2}[d] + \frac{3}{8}[h] + (1 + \frac{1}{2}\beta_d + \frac{3}{8}\beta_h)\epsilon_j$
$B_2 bip$	56·69	0·45	7	15	44·54	0·89	5	5	$\mu - \frac{1}{2}[d] + \frac{3}{8}[h] + (1 - \frac{1}{2}\beta_d + \frac{3}{8}\beta_h)\epsilon_j$
$B_1 S$	59·91	0·35	14	10	48·84	0·73	10	5	$\mu + \frac{1}{2}[d] + \frac{1}{4}[h] + (1 + \frac{1}{2}\beta_d + \frac{1}{4}\beta_h)\epsilon_j$
$B_2 S$	55·82	0·41	14	10	45·07	0·80	10	5	$\mu - \frac{1}{2}[d] + \frac{1}{4}[h] + (1 - \frac{1}{2}\beta_d + \frac{1}{4}\beta_h)\epsilon_j$

TABLE 2

Comparisons of the expected (E) and observed (O) means of 21 generations grown in two environments. The observed means and the model for predicting the expected means are given in table 1

Generations	Environment 1			Environment 2		
	E	(O-E)	V(O-E)	E	(O-E)	V(O-E)
P_1	61·93	− 1·47	0·3364	48·28	+ 0·92	0·8220
P_2	48·45	+ 1·47	0·3082	41·88	− 0·92	0·6411
F_1	63·32	− 2·02	0·1704	50·78	− 0·04	0·6923
F_2	59·25	− 2·01	0·1289	47·93	+ 0·63	0·4593
B_1	62·62	− 1·09	0·0274	49·53	− 0·57	0·8675
B_2	55·89	+ 0·07	0·1998	46·33	− 1·57	0·7517
F_3	57·22	+ 0·33	0·1380	46·51	− 0·68	0·4102
$F_2 \times P_1$	62·62	− 0·52	0·1236	49·53	+ 1·36	0·3701
$F_2 \times P_2$	55·89	+ 0·21	0·1240	46·33	− 0·80	0·3277
$F_2 \times F_1$	59·25	− 0·32	0·1094	47·93	+ 0·62	0·3724
$F_2 bip$	59·25	+ 0·49	0·1282	47·93	+ 0·11	0·5698
B_{11}	62·27	− 0·27	0·2894	48·90	+ 0·75	0·1935
B_{12}	59·60	− 1·52	0·2691	48·56	− 0·37	1·1709
B_{22}	52·17	+ 1·03	0·2666	44·11	− 0·79	0·9322
B_{21}	62·97	− 1·81	0·2757	50·16	+ 1·64	1·1719
$B_1 \times F_1$	60·94	− 1·27	0·1606	48·73	+ 0·63	0·4121
$B_2 \times F_1$	57·57	− 0·65	0·1823	47·13	+ 0·74	0·4214
$B_1 bip$	61·61	− 0·31	0·1811	48·82	+ 1·43	1·0477
$B_2 bip$	54·87	+ 1·82	0·2470	45·62	− 1·08	0·8366
$B_1 S$	60·59	− 0·68	0·1652	48·10	+ 0·74	0·5692
$B_2 S$	53·85	+ 1·97	0·2068	44·91	+ 0·16	0·6843

reference to Bucio Alanis (1966) shows that only one of the 16 earlier environments (season 1960) was as extreme as this environment.

We can now predict the expected means of the 21 generations in each of the two environments by substituting the values of μ, $[d]$, $[h]$, β_d, β_h and ϵ_j in the formulae in table 1. The errors of the expectations are derived from the errors of their constituent components by the usual method for calculating the error of any value of the dependent variable predicted by a linear regression equation (Mather, 1943). The expected means (E) and the deviations of the observed from expected means (O-E) are given in table 2. The error variances of the deviations (V(O-E)) have been obtained as the sum of the error variances of the corresponding observed and expected means.

The sum of the products of the squared deviations and the reciprocals of their variances is a χ^2 for 20 degrees of freedom which tests the goodness of fit of the 21 observed and expected means, one degree of freedom having been used to estimate ϵ_j from the observed means of the parental families. The χ^2 for environment 1 is highly significant (152·63, P < 0·001) while that for environment 2 is non-significant (2·86, P > 0·99). Thus the predictions show an exccellent fit with the observations in environment 2 but an unsatisfactory fit in environment 1. It should be noted that the test of goodness of fit is more sensitive for environment 1 because of the smaller errors of the generation means in this environment. This difference in sensitivity, however, would account for only a small fraction of the difference between the χ^2's for the two environments.

3. INTERPRETATION

The generation means in environment 2, the average environment, show the good fit with the predictions that was expected from the earlier, more limited experiments (Bucio Alanis, Perkins and Jinks, 1969). The failure to predict the observations from environment 1 can only result from the failure of one or both of the underlying assumptions.

From the investigations of the original 16 environments we know that the linear relationship holds for a range of ϵ_j values from −5·9 to 11·00. Furthermore, we know that the variation in these values is largely attributable to seasonal variation in one location, the location used for environment 2. (Three of the 16 environments were from a different location namely, environments 1, 2 and 3 of Bucio Alanis (1966) but these all had average ϵ_j values). Since the ϵ_j value of environment 2 falls in the centre of the distribution and its value is largely determined by seasonal factors in 1965 in the same location the linear relationship would be expected to hold for this environment. Environment 1, however, not only has an ϵ_j value that is at the upper extreme of the distribution, but its value is determined by locational differences that do not apply to any of the earlier values of ϵ_j and seasonal differences which it has in common with environment 2. Hence there may well be unique environmental agencies relatable to edaphic factors operating in environment 1.

The adequacy of an additive-dominance model of gene action is readily tested by fitting an m, $[d]$ and $[h]$ model to the family means in each environment. This test has already been carried out on the data from environment 1 (Jinks and Perkins, 1969) and it showed a highly significant failure of the

model ($\chi^2_{18} = 68 \cdot 27$, P < 0·001) which was traced to the presence of linked, interacting genes. The comparable analysis of the data from environment 2, on the other hand, shows a satisfactory fit with the model ($\chi^2_{18} = 7 \cdot 34$, P = 0·98). There are, therefore, good reasons why the predictions of the generation means in environment 1 are unsatisfactory and those in environment 2, satisfactory.

4. PRACTICAL CONSIDERATIONS

Although the predictions of the generation means in environment 1 must be formally regarded as unsatisfactory on the basis of the χ^2 test of goodness of fit, reference to table 2 shows that the deviations between the observed and expected means are relatively small. Indeed, predictions of this order of reliability could well be acceptable and of considerable practical value particularly in circumstances where genotype-environment interactions are present. This is underlined by the high correlation over the 21 generations in environment 1 (r = 0·97) between the observed and predicted means. Hence we can predict the relative magnitudes if not the absolute magnitudes of the means in this environment with a high degree of reliability. If, therefore, information on the relative performances of different generations, populations or genotypes is sufficient, and this would be sufficient for many purposes, for example, for selection, the present approach is adequate even in environment 1. Once, however, we admit that predictions of relative performances or rankings are sufficient a much simpler approach to prediction is available in certain circumstances. Thus where the regression lines for the different generations relating the genotype-environment interaction component of the generation means and the environmental values do not intersect, the ranking of the means of the generations remains constant over the environments even though their absolute values may change considerably. Similarly, if there are intersections and these are clustered within a certain range of environmental values, rankings will remain constant for environmental values above and below this range but they will change across it.

In the generations derived from the cross 1×5 the intersections of the linear regression lines are clustered around low values of ϵ_j (negative ϵ_js). Hence for the remainder of the range, that is, average to high environmental values ($\epsilon_j = 0$ to 11·0) the expected ranking should remain constant. The best estimate of the ranking over the 16 environments is, of course, the mean over all environments which is also the expected ranking for $\epsilon_j = 0$. This ranking should, therefore, hold for all values of ϵ_j greater than 0 and hence it is the expected ranking in environment 1. An equivalent approach is to derive the expected means of the 21 generations for $\epsilon_j = 0$ by substituting the values of μ, $[d]$ and $[h]$, obtained from the parents and F_1's in the 16 environments, in the model in table 1 (β_d and β_h will make no contribution to the expectations for $\epsilon_j = 0$).

The value of this prediction of the relative performances of the 21 generations in environment 1 may be tested by calculating the correlation between these expectations and the observed means. This has a value of $r = 0 \cdot 94$. The simpler approach, therefore, falls little short of the first in its power to predict the relative performances of the generations in the special circumstances that apply in the 1×5 cross.

5. Summary

1. The assumptions that must be satisfied for making predictions over environments and over generations when genotype-environment interactions are present are defined.

2. Predictions have been made for the character plant height of the expected means of 21 generations derived from a cross between inbred varieties 1 and 5 of *Nicotiana rustica* when grown in two environments from estimates of parameters obtained from the inbred varieties and their F_1 grown in 16 environments.

3. The predictions were satisfactory for one of the environments which was of the same kind as the previous 16 environments but not for the other environment which introduced new environmental factors.

4. The failure of the predictions in one of the environments can be traced to the failure of the assumption that an additive-dominance model of gene action is adequate in that environment.

5. In contrast, the relative performances of the 21 generations in the new environments can be reliably predicted in the present data.

6. It is suggested that reliable predictions of relative performances are sufficient for many purposes and would be of practical value particularly where genotype-environment interactions are present.

6. References

BUCIO ALANIS, L. 1966. Environmental and genotype-environmental components of variability. I. Inbred lines. *Heredity, 21,* 387-397.

BUCIO ALANIS, L., AND HILL, J. 1966. Environmental and genotype-environmental components of variability. II. Heterozygotes. *Heredity, 21,* 565-579.

BUCIO ALANIS, L., PERKINS, JEAN M., AND JINKS, J. L. 1969. Environmental and genotype-environmental components of variability. V. Segregating generations. *Heredity, 24,* 115-127.

JINKS, J. L., AND PERKINS, JEAN M. 1969. The detection of linked epistatic genes for a metrical trait. *Heredity, 24,* 465-475.

MATHER, K. 1943. *Statistical Analysis in Biology.* Methuen, London.

PERKINS, JEAN M. 1970. Environmental and genotype-environmental components of variability. VI. Diallel sets of crosses. *Heredity, 25,* 29-40.

PERKINS, JEAN M., AND JINKS, J. L. 1968*a*. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity, 23,* 339-356.

PERKINS, JEAN M., AND JINKS, J. L. 1968*b*. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity, 23,* 525-535.