

THE SIGNIFICANCE FOR BREEDING OF LINEAR REGRESSION ANALYSIS OF GENOTYPE-ENVIRONMENT INTERACTIONS

A. J. WRIGHT

Plant Breeding Institute, Trumpington, Cambridge

Received 8.i.76

SUMMARY

Methods of regression analysis of genotype-environment interaction are considered in relation to existing theory dealing with the relative efficiencies of selection for general or specific adaptation to the environment, and the choice of environments for assessment. The two alternative models involving regression on to environmental effects (model 2) or genotypic effects (model 3) are equivalent when regression lines are concurrent, but are shown to be mutually exclusive when concurrence is absent. Formulae relating the rates of advance under selection for general and specific adaptation are given, and can be used as a guide to the choice of an effective breeding strategy. When model (3) regression is important, then selection for general adaptation will be an efficient strategy but may be further enhanced by the use of environments with high regression coefficients (γ) for assessment. The advance following assessment in a single environment (k th) is expected to be better than that under n randomly chosen environments if $\gamma_k > \sqrt{n} - 1$. If model (2) regression is also important (*i.e.* regression is concurrent), then the best selector environments can be chosen on the basis of their means. If, on the other hand, model (3) regression does not hold, then selection for general adaptation will be inefficient and it is preferable to group the environments to achieve more homogeneity. When model (2) regression holds, then this grouping can be carried out on the basis of the mean expression of the environments.

1. INTRODUCTION

THE occurrence of genotype-environment interaction presents the breeder with two problems. In the first place he has to decide whether, in spite of the interaction, to try to produce a single variety with good general adaptation to the whole range of environmental and agronomic conditions of importance, or to breed varieties adapted to specific subsets of these environments (*e.g.* distinct regions or specified managements). Secondly, having chosen the range of environmental conditions within which a selection programme is to operate, he has to decide how best to evaluate his material with respect to its adaptability to this defined spectrum.

Some formulae given by Dickerson (1962) relate to these questions when environments to be used for assessment are essentially chosen at random from the population. Comstock and Moll (1963), on the other hand, recognised the importance of certain patterns of interaction in the choice of particularly useful screening environments. The development of methods of regression analysis, primarily of phenotypic values on to environmental effects (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966), but also on to genotypic effects (Wricke, 1971; Wright, 1971; Utz, 1972) has again focussed attention on the importance of these patterns for breeding methods.

The purpose of this paper is to re-examine the two questions posed above in terms of the parameters defined by these regression models, and to suggest criteria with which important decisions can be approached. Of necessity, whereas some of the arguments made are new, others represent restatements of known theory. The aim is to try to place the newer regression methods in a proper perspective within the existing framework of breeding technology.

2. THE MODELS

The basic model for the analysis of data from m genotypes grown in n environments is

$$y_{ij} = \mu + g_i + \epsilon_j + f_{ij} + e_{ij} \quad (1)$$

where μ is the overall mean, g_i the additive effect of the i th genotype, ϵ_j that of the j th environment, f_{ij} their interaction, and e_{ij} the error associated with the combination. The genotypic effect g_i may in fact pertain to a single genotype, but more generally the term is used to refer to any genetic entity with a degree of repeatability which for most crops will be a variety or some type of progeny. Both the genotypes and environments included will generally be assumed to represent only a sample from the breeding population and from the full spectrum of environments respectively, and conventional analysis of variance will therefore allow the estimation of the associated variance components: σ_g^2 , σ_e^2 , σ_f^2 and σ_e^2 .

Model (1) has been expanded in several ways in order to gain more information about the interaction component f_{ij} . The model used in essence by Finlay and Wilkinson (1963) and explicitly by Perkins and Jinks (1968) involves regression of the f_{ij} on to ϵ_j effects:

$$y_{ij} = \mu + g_i + (1 + \beta_i)\epsilon_j + d_{ij} + e_{ij} \quad (2)$$

so that $f_{ij} = \beta_i\epsilon_j + d_{ij}$.

This model is therefore concerned with the characterisation of each genotype in terms of a mean effect and a coefficient describing its rate of change of expression with changing environment. A second expansion of (1) is in a sense the converse of (2) in that interaction parameters are regressed on to genotypic effects:

$$y_{ij} = \mu + \epsilon_j + (1 + \gamma_j)g_i + s_{ij} + e_{ij} \quad (3)$$

Geidel, 1971; Wricke, 1971; Wright, 1971; Utz, 1972; Moav and Wohlfarth, 1974), where γ_j is the regression of genotypes in the j th environment on to genotypic effects and s_{ij} is the independent residue.

It has been shown previously that models (2) and (3) are related and that variation accounted for by heterogeneity of β_i parameters may be common with some or all of that due to the γ_j (Wright, 1971). The relationship depends on a scaling parameter c (Bliss, 1967) which can be alternatively considered as the regression of β_i on g_i or of γ_j on ϵ_j . Thus (2) and (3) can be further expanded to

$$y_{ij} = \mu + g_i + \epsilon_j + cg_i\epsilon_j + \beta'_i\epsilon_j + \gamma'_jg_i + r_{ij} + e_{ij} \quad (4)$$

(Wright, 1971; Utz, 1972), where β'_i and γ'_j are the residual regressions after removal of dependence on g_i and ϵ_j respectively. The situation where these

residual regressions are zero, and hence $\beta_i = cg_i$ or $\gamma_j = c\epsilon_j$ conforms to the graphical picture of complete concurrence of regression lines in model (2) or (3). It is also worthy of note here that the occurrence of an effect with a negative regression under one model (*e.g.* an environment with $(1 + \gamma_j) < 0$), under the alternative model conforms to an index effect with a value on the opposite side of the point of convergence from the mean.

The relationship between (2) and (3) can be further examined as follows. If c is the regression of β_i on g_i then

$$\sum_i g_i \beta_i = c \sum_i g_i^2, \quad \text{hence} \quad \epsilon_j (\sum_i g_i \beta_i) = \epsilon_j c (\sum_i g_i^2).$$

But

$$\epsilon_j (\sum_i g_i \beta_i) = \sum_i g_i \beta_i \epsilon_j = \sum_i g_i (f_{ij} - d_{ij}) = \gamma_j \sum_i g_i^2 - \sum_i g_i d_{ij},$$

so

$$\epsilon_j c \sum_i g_i^2 = \gamma_j \sum_i g_i^2 - \sum_i g_i d_{ij}, \quad \text{or} \quad c\epsilon_j = \gamma_j - b(d/g)_j, \quad \text{and} \quad \gamma'_j = b(d/g)_j,$$

where $b(d/g)_j$ is the regression of d_{ij} on g_i in the j th environment. It follows that if model (2) regression gives a perfect fit to the true values (*i.e.* all $d_{ij} = 0$), but $c = 0$, then all $\gamma_j = 0$ also. If deviations do exist, then the fit of model (3) in the absence of concurrence depends on the regression of d_{ij} on to g_i effects within environments. Exactly parallel arguments apply with models (2) and (3) interchanged. It is clear that with strong concurrence the models are equally likely to be applicable, but when concurrence is poor they are mutually exclusive.

Before leaving the subject of the models themselves, it should be noted that the unbiased estimation of c , β_i and γ_j parameters and their associated variance components is not without difficulty (Freeman and Perkins, 1971), although a variety of solutions have since been offered (Freeman, 1973; Perkins and Jinks, 1973; Mather and Caligari, 1974; Wright, 1976). This subject is beyond the scope of this paper, and for the present purpose any mention of these parameters will imply their true values as defined by models (2), (3) and (4).

A large part of the following development will deal with relative rates of response to selection carried out in different environments. Falconer (1952) showed how such correlated responses can be dealt with exactly as if they were correlated responses among different traits. The expected response in the l th environment to selection carried out in the k th is

$$CR_{l,k} = \bar{i}\sigma_l h_l h_k r_{lk} = \bar{i}\sigma_{lk} / \sigma_k \tag{5}$$

where h_k^2 is heritability in the k th environment, r_{ek} and σ_{ek} are the genotypic correlation and covariance between the environments, and σ_e and σ_k are the phenotypic standard deviations. The case of most importance to the present study is the correlated response in mean expression over all environments. The latter will be referred to throughout as general adaptation, which is thus defined in a way which bears no relation to the concept of stability.

3. ENVIRONMENTAL SPECIFICITY OF THE VARIETY

The first question the breeder has to ask about the interaction of genotypes and environment is whether it is of such a magnitude and type that any scheme aimed at the production of a single well adapted variety should be

abandoned in favour of the production of separate varieties for specific single environments or defined groupings of environments. The statistic on which this decision has to be based must relate the expected rate of improvement under one regime with that in the other. Comstock and Moll (1963) showed that whereas the genotypic variance available for the improvement of overall performance is σ_g^2 , that for specific adaptation to individual environments is $\sigma_g^2 + \sigma_f^2$. However, comparison of these two variances does not give a true picture of the relative expectations of the two strategies, because possible variation in the standard deviations of observed values has to be taken into account (equation 5). The upper limit to response is approached as experimental error is ignored. Thus response in a specific environment approaches $\bar{i}(\sigma_g)_j$ and the average response to independent selection within all environments becomes $\bar{i}(\overline{\sigma_g})_j$. Using a well-known relation:

$$\bar{\sigma}^2 = \overline{\sigma^2} - \text{Var}(\sigma)$$

and so

$$\bar{i}(\sigma_g)_j = \bar{i}\sqrt{\sigma_g^2 + \sigma_f^2 - \text{Var}(\sigma)}. \quad (6)$$

On the other hand, selection for a single variety with general adaptation will normally be carried out on the basis of assessment in a sample of environments. In the case of a sample of size one, the expected correlated response, assuming the same selection intensity as before (\bar{i}), and again ignoring experimental error, is the average of (5) over all environments when the l th is replaced by the overall mean. This has the expectation:

$$\text{CR} = \bar{i}\sigma_g^2/\overline{(\sigma_g)_j} = \bar{i}\sigma_g^2/\sqrt{\sigma_g^2 + \sigma_f^2 - V(\sigma)}. \quad (7)$$

Hence the relative disadvantage of selection for general adaptation is the ratio of (7) to (6):

$$\sigma_g^2/(\sigma_g^2 + \sigma_f^2 - V(\sigma)). \quad (8)$$

With selection based on a sample of n environments, the contributions of both σ_f^2 and $V(\sigma)$ to the denominator of (7) are reduced by a factor of n , and the ratio becomes: $\sigma_g^2/\sqrt{\sigma_g^2 + \sigma_f^2 - V(\sigma)}\sqrt{\sigma_g^2 + (\sigma_f^2 - V(\sigma))/n}$. An upper limit is reached when n is very large, and the response for general adaptability is $\bar{i}\sigma_g$. The relative disadvantage is then only:

$$\sqrt{\sigma_g^2/(\sigma_g^2 + \sigma_f^2 - V(\sigma))}. \quad (9)$$

The superiority of independent selection shown in (8) is of course achieved at the cost of assessing far more experimental material, whereas no such differential is involved in (9). The latter formula is not merely hypothetical, however, because even if material is evaluated in many different environments, it may still not be worthwhile to go to the expense and other difficulty of producing separate varieties. The optimum size of environmental sample will inevitably depend on factors other than absolute rates of response or rates per experimental unit, and will therefore be associated with a ratio intermediate between (8) and (9). Note that even if a specially chosen selector environment is used, the upper limit to response is still $\bar{i}\sigma_g$. It is therefore suggested that a realistic decision as to breeding strategy can be made by assuming that the true relative rate of advance must lie in the

region bounded by (8) and (9). If these ratios are low, then selection for general adaptation must be considered an inefficient procedure, and the possibility should be considered of arranging environments into more homogeneous groups with a view to their separate selection. This can be most naturally achieved on the basis of the physical relationship of environments (*e.g.* locations within a region, differential managements, etc.), but this is not possible where the environmental factors involved are unpredictable and hence unrepeatable, such as seasonal fluctuations. In any case, grouping will enhance the expectation of advance only when the intra-group genotypic correlations are raised. Some special properties of genotypic variances and correlations when the response patterns of the material conform closely to model (2) (Finlay and Wilkinson) regression will be considered later, but one important feature is that the correlation between a pair of environments is a decreasing function of their difference in terms of mean effect (Appendix). This means that when this model holds, environments can be usefully grouped according to their mean effect, and selection carried out separately for performance in good and poor environments, etc. However grouping is achieved, a recalculation of (8) and (9) on a within-group basis will reveal if significant improvements in potential advance have resulted.

Some formulae related to those used in this Section were given by Dickerson (1962) who used a method given by Robertson (1959) to derive the average genotypic correlation between pairs of environments. This was found to equal (8), and a modification of Robertson's method can be used to show that (9) is equal to the average genotypic correlation between a single environment and the mean of all environments. One important qualification about all these formulae is necessary, however. This is that the ratio of two means has been used instead of the mean of a ratio, two quantities which may differ. This is true of the derivation used by Dickerson as well as the simpler one used here, although the nature of this approximation is clearer in the present form.

4. METHODS OF ASSESSMENT

(i) *Randomly chosen environments*

Having decided on the spectrum of environmental variation within which he can effectively operate, the breeder can opt for one of several means of assessment of his material. He can use a sample of environments chosen at random from the population of interest, or environments deliberately selected in an attempt to maximise response. He can also decide on the size of such a sample to optimise the rate of advance in relation to the resources he has available.

The expected response to selection based on a randomly chosen set of environments is, from (7), equal to $\bar{i}\sigma_g^2/\bar{\sigma}_j$. These parameters are defined with respect to the appropriate population of environments which may be the whole or a subset of those considered in the previous section. It should also be noted that this formula is again an approximation for the same reasons as those previously given. Relation (6) can again be used to find the expectation of $\bar{\sigma}_j$, but this time allowing for error variation. Thus, for single environments

$$\bar{\sigma}_j = \sqrt{\sigma_g^2 + \sigma_f^2 + \sigma_e^2 - V(\sigma)}$$

where $V(\sigma)$ is now the variance of *phenotypic* standard deviations. Hence, for a random sample of n environments each represented by r replications:

$$CR = \bar{i}\sigma_g^2/\sqrt{\sigma_g^2 + (\sigma_f^2 - V(\sigma))/n + \sigma_e^2}/rn \quad (10)$$

where σ_e^2 is the error attached to an individual plot or plant observation. This formula is similar to that derived by Dickerson (1962) who examined the effect of increasing the number of environments used for testing (n) while holding the total amount of experimental material per genetic unit (nr) constant. He concluded that unless the correlation among environments (equal to expression (8) of this paper) is low, then no increase in n above ten is justified. It has already been argued here that if (8) is low, then an incorrect breeding strategy is being employed. The rate of increase of response with increasing n and constant nr can be found by differentiation of (10) to be

$$\frac{CR}{2n^2} (\sigma_f^2 - V(\sigma))/(\sigma_g^2 + (\sigma_f^2 - V(\sigma))/n + \sigma_e^2/rn)$$

which shows that the advantage of further increases in n diminishes to the order of n^3 , suggesting that much lower numbers (two or three) are likely to be optimum.

(ii) *The use of regression models*

The above argument deals with the responses to be expected from environments taken at random from the range of interest. When data are available on the performance of the breeding population in a sample of environments from the range bred for, then application of (3) and (4) can lead to deliberate choices of good screening environments, which, insofar as they are controllable and repeatable, can be used for future screening of the same or similar material. This last point is important, because the patterns of interaction of two different populations of genotypes with a single set of environments may be quite dissimilar, and although it is unlikely that fixed entities such as varieties will be assessed more than once in the same environments, material slowly evolving under some form of recurrent selection scheme will be.

(a) *Model (3)*

In the first place, equation (5) can be expanded in terms of model (1) as was done by Comstock and Moll (1963) to give, for the correlated mean response over a group of environments to selection in the k th:

$$CR_k = \bar{i}(\sigma_g^2 + (\sigma_{gf})_k)/\sqrt{\sigma_g^2 + 2(\sigma_{gf})_k + (\sigma_f^2)_k + (\sigma_e^2)_k} \quad (11)$$

where $(\sigma_f^2)_k$ is the variance of interaction effects in the k th environment and $(\sigma_{gf})_k$ their covariance with genotypic effects. Clearly, this expression is increased by positive increments in $(\sigma_{gf})_k$ because the numerator is raised more than is the denominator, especially since concomitant increases in $(\sigma_f^2)_k$ will usually be involved. Because terms of the form $\sum_i g_i s_{ij}$ are zero, model (3) can be used to further expand (11). (This is not true of model (2) because $\sum_i g_i d_{ij} \neq 0$.) Thus:

$$CR_k = \bar{i}\sigma_g^2(1 + \gamma_k)/\sqrt{\sigma_g^2(1 + \gamma_k)^2 + (\sigma_s^2)_k + (\sigma_e^2)_k} \quad (12)$$

where $(\sigma_s^2)_k$ and $(\sigma_e^2)_k$ are the variances of s and e effects in the k th environment. Note that if γ_k is less than -1 , then the correlated response will be negative. The beneficial influence of increasing γ_k can be quantified by means of the derivative of (12) with respect to γ_k . Provided that σ_s^2 and σ_e^2 are homogeneous over environments, this is equal to

$$\bar{i}\sigma_g^2(\sigma_s^2 + \sigma_e^2)/(\sigma_g^2(1 + \gamma_k)^2 + \sigma_s^2 + \sigma_e^2)^{1\frac{1}{2}}.$$

Hence increases in γ_k not associated with increases in $(\sigma_s^2)_k$ and $(\sigma_e^2)_k$ are always beneficial, but to a diminishing extent with higher levels of γ_k .

If there are no deviations from the k th regression line ($(\sigma_s^2)_k = 0$), then (12) can be written as $CR_k = \bar{i}\sigma_k h_k^2/(1 + \gamma_k)$ because the regression of general adaptation on to genotype in the k th environment is $1/(1 + \gamma_k)$. Thus when model (3) gives a good account of the data, and $(\sigma_s^2)_k$ are independent of γ_k , although both the direct (response in the selector environment) and correlated (response for general adaptation) responses will be greater for environments with high γ_k , the correlated responses will have a smaller range and will in fact be greater than the direct responses for environments with negative γ_k . In the extreme situation where σ_s^2 and σ_e^2 are zero, correlated response is always $\bar{i}\sigma_g$ irrespective of the environment used for selection. It is important to remember, therefore, that if regression accounts for a large proportion of the variation, then there is little to be gained either from careful choice of selectors or from the use of more than one environment, since the response from any single environment, given adequate replication, is close to the optimum.

However, for the case where σ_s^2 and σ_e^2 remain substantial quantities, it is of interest to examine the advantage to be gained from the deliberate rather than the random choice of environments for use as selectors. To return first to the point made earlier that equations (7) and (10) are derived as the ratio of means instead of the mean of a ratio, it is now clear from (12) that the numerator and denominator cannot be assumed to be independent, and so the formula can be regarded as approximate only. However, the discrepancy becomes less serious as the size of environmental sample is increased. In this case the correlated response expected to follow assessment in a sample of environments is therefore approximately:

$$CR = \bar{i}\sigma_g^2(1 + \bar{\gamma})/\sqrt{\sigma_g^2(1 + \bar{\gamma})^2 + \sigma_s^2/n + \sigma_e^2/n} \tag{13}$$

and so (12) and (13) are equal when $(1 + \gamma_k)/(1 + \bar{\gamma}) = \sqrt{n}$. Assuming that with randomly chosen environments, $\bar{\gamma}$ will be very close to zero, it is seen that a single environment must have γ in excess of $(\sqrt{n} - 1)$ to be expected to give a higher response than a random sample of environments (this again assumes that $(\sigma_s^2)_k$ and $(\sigma_e^2)_k$ are independent of γ_k). Thus to be superior to random samples of two, three, or four environments, γ must exceed 0.41, 0.73 and 1.00 respectively. Insufficient examples of this type of analysis are yet available to know how often, and under what conditions, values of this magnitude are approached in practice.

(b) *Model (2)*

It has already been pointed out that model (2) is of limited use for the formulation of selection responses because genotypic variances and

covariances contain unwieldy cross product terms between deviations (d_{ij}) and genotypic effects. However, evidence is accumulating to show that deviations from this model are commonly small or absent (Hill, 1976), and so the expectations for the case when regression gives an exact fit to the true values is not without interest.

An important immediate property of note is that the average expression of any genotype over all environments is equal to its expression in any environment which has average effect (*i.e.* $\epsilon_j = 0$). The distinction between the "mean of environments" and a "mean environment" disappears. From (2), the variance of genotypic values in the j th environment is

$$\sigma_j^2 = \sigma_g^2 + 2\sigma_{g\beta}\epsilon_j + \sigma_{\beta}^2\epsilon_j^2,$$

where $\sigma_{g\beta}$ is the covariance of β_i with g_i effects, and σ_{β}^2 their variance. The mean variance within environments is $\sigma_g^2 + \sigma_{\beta}^2\overline{\epsilon_j^2}$ or $\sigma_g^2 + \sigma_f^2$, as expected. The variance in an average environment is σ_g^2 , and so is always less than the average variance, a fact which may not be intuitively apparent but which can be demonstrated graphically. Using model (4)

$$\sigma_j^2 = \sigma_g^2(1 + c\epsilon_j)^2 + \sigma_{\beta'}^2\epsilon_j^2 \quad (14)$$

where $\sigma_{\beta'}^2$ is the variance of β'_i parameters. This leads to the cases of complete or zero concurrence as $\sigma_g^2(1 + c\epsilon_j)^2$ and $\sigma_g^2 + \sigma_{\beta'}^2\epsilon_j^2$ respectively. With no concurrence, it is apparent that genotypic variance is increased as any environment departs from the average in either direction, whereas with complete concurrence the increase is symmetrical above and below an environment with $\epsilon_j = -1/c$. For the general situation of partial concurrence (14), the minimal variance occurs where $\epsilon_j = -c\sigma_g^2/(c^2\sigma_g^2 + \sigma_{\beta'}^2)$. It is therefore another useful general property of the model which is not apparent from its graphical representation, that the within environmental variance always increases symmetrically above and below a certain critical value of ϵ_j . It is likely that the rate of response to direct selection within environments will follow a similar pattern, unless the error variance is related to ϵ_j .

A similar procedure can be used to give the covariance between the k th environment and the overall mean, and hence the correlated response of general adaptation to selection in the k th as

$$CR_k = \bar{i}\sigma_g^2(1 + c\epsilon_k)/\sqrt{\sigma_g^2(1 + c\epsilon_k)^2 + \sigma_{\beta',\epsilon_k}^2 + (\sigma_e^2)_k}. \quad (15)$$

This expression is identical to (12) for the case of $\sigma_d^2 = 0$, because now $\gamma'_k = 0$, as already shown, and $\gamma_k = c\epsilon_k$. In this form, however, it can be differentiated with respect to ϵ_k in order to find whether extreme or intermediate environments are likely to be the most effective for the improvement of general adaptation. It turns out that the sign of this derivative is the same as that of $(c\sigma_e^2 - \sigma_{\beta',\epsilon_k}^2)$. The balance of these two terms again reflects the conflict between concurrent and independent regression. Because concurrence increases the numerator as well as the denominator of (15), extreme environments with the highest variance are the best selectors because the detrimental effect of error (again provided that σ_e^2 is independent of ϵ_k) is thereby reduced. Without concurrence, on the other hand, increased variance is itself detrimental, and intermediate environments are superior.

In practice a mixture of concurrent and independent regression is likely, and will result in environments somewhat higher (c positive) or lower (c negative) than the average being the best selectors.

5. BREEDING IMPLICATIONS

The formulae relating advance under selection for specific or general adaptation do not lend themselves to a straightforward expansion in terms of models (2) or (3). The significant parameter is $V(\sigma)$, which was first identified by Robertson (1959) as being that portion of the observed interaction which has no effect on genotypic correlations among environments. It should be noted that $V(\sigma)$ does not conform exactly to any of the components of variance associated with models (2), (3) or (4), but approximates sufficiently closely to allow an interpretation of the models in this light. Moav and Wohlfarth (1974) suggested that the scaling effect removed by $V(\sigma)$ is identical to that associated with the scale parameter c in model (4). This is true only if concurrent regression gives a perfect fit to the data, and would be equally true for all the variance associated with regression in model (3), including that due to concurrence, if this fitted the data. This is apparent because, provided $(1 + \gamma_j)$ was positive, γ_j would then be proportional to σ_j . But when σ_s^2 is not zero, the expectation of $V(\sigma)$ is complex. However, it seems that $V(\sigma)$ is fairly closely related to the variance accounted for by γ , especially when this is large, and that the "true interaction", $(\sigma_f^2 - V(\sigma))$, is related to σ_s^2 .

On this basis, it can be predicted that when (8) and (9) are close to unity, and general adaptation within the environmental spectrum is a worthwhile breeding objective, than σ_s^2 will be small and model (3) regression will give a good account of the data. It has already been noted that under these circumstances, the assessment of general adaptation can be efficiently achieved in any single environment. If (8) and (9) are lower, then the improvement of general adaptation is more difficult, and the choice of the means of assessment is more critical for success. The heterogeneity of γ parameters will give scope for the choice of efficient selectors and may obviate the necessity for large environmental samples. If concurrence is important, then the best selectors will also be recognisable by their high (or low) means.

If the expected rate of advancement of general adaptation is low relative to that for specific environments, then the full potential of the crop will not be realised unless the environmental spectrum is subdivided into more homogeneous portions. Furthermore, under these conditions, the fit to model (3) regression will be poor, and there will be little scope for the choice of good selectors. Model (2) may account for a large part of the interaction, and then environments may be advantageously grouped simply according to their mean values, and the production of separate varieties for good, intermediate and poor environments be a more efficient procedure. However, to the extent that the environmental variation depends on unpredictable, uncontrollable factors such as seasonal effects, then this approach will not be possible. The variation in rates of genotypic response to the environment implied by the predominance of model (2) regression naturally suggests selection for improved stability (Perkins and Jinks, 1968, 1971), and if

environmental variation is of the uncontrollable type which cannot be dealt with by any meaningful subdivision of the environmental spectrum, then this may be the only approach left to the breeder. However, it is important to remember that improvement of stability represents increased performance in poor at the expense of good environments in order to provide acceptable levels throughout, and has little or no effect on mean performance overall. It is therefore no substitute for the appropriate limitation of breeding objectives to defined environmental conditions whenever possible, and is otherwise likely to be costly in terms of overall advance.

6. REFERENCES

- BLISS, C. I. 1967. *Statistics in Biology*. McGraw-Hill, New York.
- COMSTOCK, R. E., AND MOLL, R. H. 1963. Genotype-environment interactions. In *Statistical Genetics and Plant Breeding*, NAS-NRC982, Washington, D.C., pp. 164-196.
- DICKERSON, G. E. 1962. Implications of genetic-environmental interactions in animal breeding. *An. Prod.*, 4, 47-64.
- EBERHART, S. A., AND RUSSELL, W. A. 1966. Stability parameters for comparing varieties. *Crop Sci.*, 6, 36-40.
- FALCONER, D. S. 1952. The problem of environment and selection. *Am. Nat.*, 86, 293.
- FINLAY, K. W., AND WILKINSON, G. N. 1963. The analysis of adaptation in a plant breeding programme. *Austr. J. Agric. Res.*, 14, 742-754.
- FREEMAN, G. H. 1973. Statistical methods for the analysis of genotype-environment interactions. *Heredity*, 31, 339-354.
- FREEMAN, G. H., AND PERKINS, J. M. 1971. Environmental and genotype-environmental components of variability. VIII. Relations between genotypes grown in different environments and measures of these environments. *Heredity*, 27, 15-23.
- GEIDEL, H. 1971. Zur Interpretation von Interaktionen bei Versuchsserien. *EDV in Medizin und Biologie*, 2, 13-18.
- HILL, J. 1976. Genotype-environment interactions—a challenge for plant breeders. *J. Agric. Sci., Camb.*, 85, 477-494.
- MATHER, K., AND CALIGARI, P. D. S. 1974. Genotype and environment interactions. I. Regression of interaction on overall effect of the environment. *Heredity*, 33, 43-59.
- MOAV, R., AND WOHLFARTH, G. W. 1974. Magnification through competition of genetic differences in Carp. *Heredity*, 33, 181-202.
- PERKINS, J. M., AND JINKS, J. L. 1968. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity*, 23, 525-535.
- PERKINS, J. M., AND JINKS, J. L. 1971. Specificity of the interaction of genotypes with contrasting environments. *Heredity*, 26, 463-474.
- PERKINS, J. M., AND JINKS, J. L. 1973. The assessment and specificity of environmental and genotype-environmental components of variability. *Heredity*, 30, 111-126.
- ROBERTSON, A. 1959. The sampling variance of the genetic correlation coefficient. *Biometrics*, 15, 469-485.
- UTZ, H. F. 1972. Die Zerlegung der Genotyp \times Umwelt-Interaktionen. *EDV in Medizin und Biologie*, 3, 52-59.
- WRICKE, G. 1971. Eine orthogonale Aufteilung der Interaktion für ein eingeschränktes Modell. In *Rdsehr. Arb.-Gem. Biom. DLG1*.
- WRIGHT, A. J. 1971. The analysis and prediction of some two factor interactions in grass breeding. *J. agric. Sci., Camb.*, 76, 301-306.
- WRIGHT, A. J. 1976. Bias in the estimation of regression coefficients in the analysis of genotype-environmental interaction. *Heredity*, in press.

7. APPENDIX

A proof follows that when genotypic values are completely specified by regression on to environmental effects, the genotypic correlation between

two environments is decreased as the difference between their mean effects is increased. From (2) and (4):

$$y_{ij} = \mu + \epsilon_j + g_i(1 + c\epsilon_j) + \beta'_i\epsilon_j.$$

Writing $\epsilon_k = \epsilon$, and $\epsilon_l = \epsilon + x$,

$$y_{ik} = \mu + \epsilon + g_i(1 + c\epsilon) + \beta'_i\epsilon$$

and

$$y_{il} = \mu + \epsilon + x + g_i(1 + c(\epsilon + x)) + \beta'_i(\epsilon + x).$$

Dealing with the square of the correlation (r_{kl}^2) for simplicity, this is

$$r^2 = \frac{[\sigma_g^2(1 + c\epsilon)(1 + c(\epsilon + x)) + \sigma_{\beta'}^2\epsilon(\epsilon + x)]^2}{[\sigma_g^2(1 + c\epsilon)^2 + \sigma_{\beta'}^2\epsilon^2][\sigma_g^2(1 + c(\epsilon + x))^2 + \sigma_{\beta'}^2(\epsilon + x)^2]} = \frac{N^2}{D_1 D_2}$$

and differentiating with respect to x :

$$\frac{d(r^2)}{dx} = \frac{[c\sigma_g^2(1 + c\epsilon) + \sigma_{\beta'}^2\epsilon]2ND_1D_2 - [2c\sigma_g^2(1 + c(\epsilon + x)) + 2\sigma_{\beta'}^2(\epsilon + x)]D_1N^2}{D_1^2D_2^2}.$$

For the present purpose, only the sign of this derivative is important, and this is dependent on:

$$[c\sigma_g^2(1 + c\epsilon) + \sigma_{\beta'}^2\epsilon]D_2 - [c\sigma_g^2(1 + c(\epsilon + x)) + \sigma_{\beta'}^2(\epsilon + x)]N$$

which after substitution for D_2 and N reduces to $-\sigma_g^2\sigma_{\beta'}^2x(1 - cx) \approx -\sigma_g^2\sigma_{\beta'}^2x$. As expected, this is zero if concurrence is complete ($\sigma_{\beta'}^2 = 0$), but otherwise is always negative, as required.