

## STATISTICAL METHODS FOR THE ANALYSIS OF GENOTYPE-ENVIRONMENT INTERACTIONS\*

G. H. FREEMAN

*National Vegetable Research Station, Wellesbourne, Warwick*

Received 7.xii.72

### SUMMARY

This is largely a review paper, describing various statistical methods for analysing interactions in general and genotype-environment interactions in particular, and giving nearly 100 references to previous work. The joint regression analysis approach introduced by Yates and Cochran (1938) is considered in some detail; alternatives to regression are discussed, as are various stability parameters. Much work has been done on statistical methods for testing for interactions in general, and this also is reviewed, from Tukey's (1949) one degree of freedom for non-additivity to Milliken and Graybill's (1970) generalisation to testing for various types of possible interaction. The difficulties of testing and inference in the presence of interaction are discussed. Data from a two-way table may be regarded as a multivariate set, as first shown by Williams (1952) and later extended by others, particularly Mandel (1969*b* and 1971). These methods are only just beginning to be used in studies of genotype-environment interactions, and several recent references are given. External measurements may be used to measure the environment and these may be either physical or biological. Again, the appropriate methods of analysis are fairly new. The interpretation of interactions is considered in relation to the use to be made of the results. It is suggested that various multivariate techniques may be used to assist in the elucidation of interactions, especially when these are not easy to explain by simpler methods of analysis.

### 1. INTRODUCTION

VARIOUS methods have been proposed for the statistical analysis of interactions in general, and genotype-environment (GE) interactions in particular, and these will be reviewed in this paper. It is necessary first to find whether interactions are present, and then to consider their importance and effect on subsequent work. If there are interactions it is sometimes enough to find the components of the variation attributable to the main effects and interactions, but it is usually better to look for the structure underlying the observed non-additivity. This is all part of an attempt to discover more about the data than the crude tests reveal; thus, assuming that "data = pattern + noise", one wants to find as much as possible of the pattern while eliminating the maximum noise.

Much work has been done on this subject, both by statisticians interested in non-additivity in general and by those, including agronomists, breeders and geneticists, who are particularly concerned with GE interactions. The simpler methods, including the use of variance components when there are GE interactions, will be briefly discussed, then regression and similar methods for GE will be considered; the concept of stability will also be mentioned. Statistical methods for testing for various forms of non-additivity in general

\* Based on a paper presented at the Symposium on Genotype  $\times$  Environment Interactions, Birmingham University, September, 1972.

will then be examined, and the difficulties of testing and inference in the presence of interactions discussed. Other methods will be considered, including some based on the use of multivariate methods of analysis and others involving external environmental variables. The interpretation of the results will be discussed, in the light of the use to be made of the results in the wider biological context. The observed components of GE interaction will not be considered in terms of the parameters of biometrical genetics, these being discussed fully by Mather and Jinks (1971). A fairly comprehensive list of papers describing statistical methods is given, together with some giving interesting applications of known methods of analysis; some of the references are very recent, as the subject is still under development. Not all the papers listed are referred to in the text, and there are too many papers on applications of the methods for them all even to be listed.

## 2. EARLY APPROACHES TO THE STUDY OF GE INTERACTIONS, INCLUDING THE USE OF VARIANCE COMPONENTS

The existence of interactions between genotypes and environmental factors has long been recognised, the earliest reference, which indeed precedes the analysis of variance, being Fisher and Mackenzie (1923). In considering the manurial responses of different potato varieties they concluded that "the yields of different varieties under different manurial treatments are better fitted by a product formula than by a sum formula". The methods used here foreshadowed those developed many years later, but were apparently completely forgotten.

Sprague and Federer (1951) showed how variance components could be used to separate out the effects of genotypes, environments and their interaction by equating the observed mean squares in the analysis of variance to their expectations on the random model. Thus, in terms of a mathematical model, the yield  $y_{ijk}$  of the  $k$ th replicate of the  $i$ th genotype in the  $j$ th environment is regarded as made up of a general mean  $\mu$ , a genotype effect  $d_i$ , an environmental effect  $\epsilon_j$ , an interaction effect  $g_{ij}$  and a random error  $e_{ijk}$ , i.e.

$$y_{ijk} = \mu + d_i + \epsilon_j + g_{ij} + e_{ijk}, \quad (1)$$

where

$$\sum_i d_i = \sum_j \epsilon_j = \sum_{ij} g_{ij} = 0.$$

The components of variance corresponding to each of these terms are extracted, and can be used to obtain heritability estimates. To quote from Sprague and Federer, "the existence of . . . interaction is the cause of the small increase in average genetic advance". Many others followed this procedure, for example, Miller, Williams and Robinson (1959), who put the problem squarely in the plant breeding context in considering GE interactions for cotton. Miller, Robinson and Pope (1962) found that the three-factor interaction of varieties with sites and years was particularly important. Hanson (1964) considered environmental effects as random combinations of years and sites, and went on to deal with the case where some of these combinations were missing. Allard and Bradshaw (1964) reviewed previous work, emphasising the importance of interactions, particularly varieties  $\times$  years, to plant breeders.

One possible way of overcoming the difficulties caused by the presence of interactions is to seek to transform the data to a scale in which there are no interactions. This is a well-known statistical procedure (see *e.g.* Bartlett (1947), Tukey (1949)), and can sometimes be applied to interactions between genotypic and environmental effects. Mather (1971) considers in some detail the question of the scale of measurement and quotes one example in which interactions are eliminated by a log transformation. However, he concludes that "we must always be prepared to bring interactions explicitly into an analysis".

In genetical studies Mather and Morley Jones (1958) said: "The growing mass of evidence for interaction between genotype and environment lends urgency to the need for its close consideration". The same authors (Morley Jones and Mather, 1958) set out how GE interactions influence variances and covariances used for measuring variation in biometrical genetical models. When interactions exist "the measures of genetic effects apply only to the range of environments studied" and conversely. The approach by geneticists was thus entirely different from that of plant breeders and agronomists: the latter group wished merely to minimise the effects of GE interactions on their field trials, while the former wished to understand the causes of the interactions, in terms of biometrical genetical parameters.

### 3. THE USE OF REGRESSION METHODS IN STUDYING GE INTERACTIONS

The idea of breaking down an interaction into several parts is entirely missing in the variance component approach described above. In the GE context a method of partitioning had in fact been given by Yates and Cochran (1938), though it was largely neglected for 20 years. In their words, "the degree of association between varietal differences and general fertility (as indicated by the mean of all varieties) can be further investigated by calculating the regression of the yields of the separate varieties on the mean yields of all varieties". That is,  $g_{ij}$  in equation (1) is regressed on  $\epsilon_j$ , *i.e.*

$$g_{ij} = \beta_i \epsilon_j + \delta_{ij}, \quad (2)$$

where  $\beta_i$  is a linear regression coefficient for the  $i$ th genotype and  $\delta_{ij}$  a deviation. Using (2), (1) can then be rewritten as:

$$y_{ijk} = \mu + d_i + (1 + \beta_i) \epsilon_j + \delta_{ij} + e_{ijk}. \quad (3)$$

Yates and Cochran showed that this regression accounted for a large part of the interaction in a set of barley trials, but their ideas were not really taken up until Finlay and Wilkinson (1963) rediscovered the same method and used it for an analysis of adaptation in a trial with 277 varieties of barley in seven environments. This method was also used by Perkins and Jinks (1968a), who used it for estimating parameters in a biometrical genetical model.

The formal analysis of variance for GE interactions is given in an identical manner by Yates and Cochran, Finlay and Wilkinson and Perkins and Jinks. In the notation of the last-named, there are  $t$  genotypes and  $s$  environments and the partition of the interaction is into a term with  $(t-1)$  degrees of freedom (d.f.) for heterogeneity of regressions and one with  $(t-1)(s-2)$  d.f. for deviations from regression. Each of these terms can be

compared with the residual experimental error, and the heterogeneity of regression further compared with the deviations in order to see if it accounts for a significantly large part of the observed interaction. This particular approach, now commonly known as *joint regression analysis*, has been widely adopted in practice, some applications being given in the list of references.

Regression methods were also considered by Rowe and Andrew (1964) and Eberhart and Russell (1966), who added together the sums of squares for environments and GE interactions and repartitioned this. Their partitioning is into a linear component between environments with 1 d.f., a linear component of the GE interaction with  $(t-1)$  d.f., and deviations from regression, the deviations being found separately for each of the  $t$  genotypes with  $(s-2)$  d.f. each. The trouble with this approach, as pointed out by Freeman and Perkins (1971), is that the sum of squares for the linear component between environments, which is allocated one degree of freedom, is the same as the total sum of squares for environments with  $(s-1)$  d.f. (Incidentally, it is now apparent to me that some of the criticisms in that paper of the method of regressing on the environmental mean were misdirected, since marginal means may be regarded as fixed when inferences are made about values in the body of a table. Nevertheless, the comment about the confusion over degrees of freedom for the environmental sum of squares remains valid.) Perkins and Jinks recognised that the environmental sum of squares is the same as that for the combined regression over all genotypes but did not use it so avoided this difficulty.

Significance testing in either of these methods is dependent upon even more assumptions than is usual in the analysis of variance. Eberhart and Russell point out that in their approach the comparison of the linear component of the interaction against deviations from regression assumes that the deviations within the various genotypes are homogeneous. The same is true in the Yates and Cochran approach. For the same reason it is better to test the significance of the  $\beta_i$  for a particular genotype by comparing the appropriate sum of squares against the deviations for regression for that genotype rather than against the pooled deviations term.

#### 4. ALTERNATIVES TO REGRESSION ON THE ENVIRONMENTAL MEAN

Although the approach described in the last section has been very widely used, other methods have been suggested. Those concerned with assessing the environment by physical measures will be considered in section 9, while others will be discussed here. For two inbred lines Bucio Alanis (1966) regressed the difference between the values for a range of environments on their mean; this is still regression on the environmental mean. Bucio Alanis and Hill (1966) went on to regress the values for the  $F_1$  hybrid between these lines on the midparental mean, thus using what amounted to an entirely independent assessment of the environment. Bucio Alanis, Perkins and Jinks (1969) later extended this to prediction across several generations from the values of the parents and  $F_1$  hybrid. In all these papers biometrical genetical parameters were obtained.

An entirely different regression method for relating an  $F_1$  hybrid to its parents was considered by Knight (1971), who used separate regression coefficients for the two parents for the purpose of assessing hybrid vigour. This is equivalent to taking the midparental value as one independent

variable, as do Bucio Alanis and Hill, but taking the difference between the parents as another independent variable rather than regarding all other variation as being part of deviation from regression. The dominance in a cross is then predictable from the parental yields, and it may well vary between environments, reflecting variation in parental values.

An alternative to regression methods is the use of structural relationships, which gets round the logical difficulties of regressing one set of variables on another which is not independent of them. In the context of GE interactions this method was introduced by Tai (1971), who used a parameter  $\alpha_i$  for the linear response of the  $i$ th genotype to environments and showed that  $(\beta_i - 1)$  was a biased estimate of  $\alpha_i$ , being always smaller in absolute value. The bias inherent in the regression estimate of  $\beta_i$  was also pointed out by Shukla (1972a); a slightly different point was made by Utz (1972), namely that the estimate of  $\beta_i$  contains a contribution from the regression of the interaction on the two main effects.

Hardwick and Wood (1972) use an entirely different approach by seeking for an underlying physiological model and regressing on the parameters in this model. They show that the empirical regression on the environmental mean is an approximation to this that holds only if genotypes differ in just one dimension. They also show that the estimate of  $\beta_i$  is biased, because of the assumption of lack of error in the independent variate.

## 5. STABILITY PARAMETERS

One of the main reasons for growing genotypes in a wide range of environments is to estimate their stability, a concept defined in various ways: for example, Wricke (1962) considered *ecovalence*, which is the contribution of a genotype to the GE interaction sum of squares. Shukla (1972a) also partitioned the GE interaction sum of squares into components for each genotype separately by considering the *stability variance*  $\sigma_i^2$  of the  $i$ th genotype, defined as the variance over environments of  $(g_{ij} + \bar{e}_{ij})$  in equation (1), where  $\bar{e}_{ij}$  is the mean of  $e_{ijk}$  over replicates. He found an unbiased estimate of  $\sigma_i^2$  and an approximate criterion for testing whether  $\sigma_i^2$  was so large that the genotype should be considered unstable.

Other stability parameters are obtained from the regression approach. Thus, Finlay and Wilkinson suggest that genotypes with a low value of  $\beta_i$  (negative in the formulation in (3) above) are to be regarded as stable while those with a high value are unstable. Eberhart and Russell also use  $\beta_i$  as a first measure of stability but go further and regard the sum of squared deviations  $s_{di}^2$  as a second measure. The use of two stability parameters may well be valuable for some purposes, and indeed Jowett (1972), using sorghum data, found Wricke's *ecovalence* to be less informative than Eberhart and Russell's two parameters.

Tai uses  $\alpha_i$  as one measure of stability and also defines a second measure  $\lambda_i$ . If the variance component for deviations from linearity for the  $i$ th genotype is  $\sigma_{di}^2$  and the error variance is  $\sigma_e^2$ , then:

$$\lambda_i = \frac{\sigma_{di}^2 + \sigma_e^2}{\sigma_e^2} \quad (4)$$

Tai shows that the term (deviation mean square  $\times$  number of replicates/error mean square) is greater than the estimate of  $\lambda_i$  when there are many more



genotypes than environments, *i.e.* in the sort of situation of Finlay and Wilkinson. Shukla (1972*a*) extended his definition of stability variance to allow for regression on a covariate, usually but not necessarily the environmental mean. Using the original data of Yates and Cochran he found that, without allowance for regression, two varieties were unstable; the instability of one was entirely removed and of the other greatly reduced by considering deviations from regression on the environmental mean.

Some writers prefer to take just one measure of stability. Thus, Hardwick and Wood state that in terms of their underlying model the deviations  $\delta_{ij}$  are not independent of the regression on the environmental mean, so that the second stability parameter of Eberhart and Russell is not meaningful. Hanson (1970) considers a stability measure that is similar to Wricke's but takes account of regression. In terms of the parameters of the model, the ecovalence of the  $i$ th genotype is  $\sum_j g_{ij}^2$ , while Hanson's parameter is  $D_i$ , where:

$$D_i^2 = \sum_j \{g_{ij} + (1 - \alpha)\epsilon_j\}^2. \quad (5)$$

$\alpha$  here is not the same as Tai's  $\alpha$ , being defined as the minimum observed value of  $(1 + \beta_i)$ . In a soybean trial Hanson found that similar values of  $D_i^2$  were obtained by taking  $\alpha = 0$  and its empirical value of 0.75, but that taking  $\alpha = 1$ , thus making the parameter the same as Wricke's, gave very different values. When there are only a few genotypes, or the absolute value of  $\beta_i$  is not more than about 0.2, Utz suggests that similar results are obtained by taking any value of  $\alpha$  between 0 and 1, and shows that this holds, at least approximately, for Yates and Cochran's data.

## 6. STATISTICAL METHODS FOR TESTING FOR NON-ADDITIVITY

So far, all the work on GE interactions has been described as if it was going on in a world of its own, completely divorced from the main stream of statistical thought. For many years this was to a large extent true, although statistical procedures for a two-way table had already been worked out. It had long been appreciated by statisticians that although the overall test for interactions by means of analysis of variance is usually simple it is not very sensitive or efficient. Tukey (1949) developed a specific procedure for testing for the presence of non-additivity in a two-way table by extracting one degree of freedom and showing that a valid  $F$ -test was possible. In this, although Tukey did not say so, the underlying model is that, in equation (1) above, the interaction parameter  $g_{ij}$  is a constant multiple of  $d_i\epsilon_j$ , as was first shown by Ward and Dick (1952); further, Rojas (1973) showed that Tukey's test is equivalent to testing the significance of the regression coefficient in a model with a single covariate taking the value  $d_i\epsilon_j$ . Ward and Dick went on to extend the model slightly, but Elston (1961) showed that, although this extended model gives a sum of squares that is more effective than Tukey's in accounting for non-additivity, Tukey's method is computationally simpler as well as having a test statistic known to follow the  $F$ -distribution, unlike Ward and Dick's. Little is known about the power of Tukey's test; Ghosh and Sharma (1963) found that, in one situation where Tukey's test could be directly compared with another  $F$ -test, the former was the more powerful

when main effects were large and very little inferior in any circumstances. Scheffé (1959, sec. 4.8) showed that, under the null hypothesis, the conditional and unconditional distributions were the same in Tukey's test. Mandel (1961) used a regression approach which amounted to the same as that in equation (2) above and validated it; as he said, this model is more flexible than that of Ward and Dick.

The whole subject was tied up by Milliken and Graybill (1970), who worked with a very general model and showed that all the methods previously described were special cases of it. They used matrix notation, and wrote the general model as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\alpha} + \mathbf{e}, \quad (6)$$

where  $\mathbf{y}$  is an  $(n \times 1)$  random observation vector,  $\mathbf{X}$  is an  $(n \times p)$  matrix of known numbers of rank  $q$  ( $\leq n-1$ ),  $\boldsymbol{\beta}$  is a  $(p \times 1)$  vector of unknown parameters,  $\mathbf{F} = f_{ij}(\cdot)$  is an  $(n \times k)$  matrix where the functional form of  $f_{ij}(\cdot)$  is known but it is a function of the unknown estimable functions of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$  is an unknown vector and  $\mathbf{e}$  is a random normal error vector. They showed that a valid test of the null hypothesis  $\boldsymbol{\alpha} = \mathbf{0}$  can always be obtained, though the computations for the test may be rather tedious unless, as is usual in practice,  $\mathbf{F}$  is a column vector. They stressed that the functions in  $\mathbf{F}$  do not have to be multiplicative functions of  $\mathbf{X}\boldsymbol{\beta}$  so a very general form of non-additivity can be tested for.

## 7. DIFFICULTIES IN TESTING AND INFERENCE CAUSED BY INTERACTIONS

The usual form of the null hypothesis in an analysis of variance includes the assumption of no interactions between main effects. When the data do not invalidate this hypothesis the main effects may then be validly tested against the residual variation. Further, the factors exert their effects unconditionally: that is, in the present context, any differences found between genotypes will be the same for any environment, whether or not it has been included in the experiment. If components of variance are to be estimated, those given by equating mean squares to their expected values in the random model have minimum variance in the class of all unbiased estimates (Graybill and Wortham, 1956) whether or not there is interaction.

When there are interactions it is well known that there are difficulties in testing and inference (see, for example, Scheffé). To test if there are differences between genotypes, it is usually regarded as important to decide whether the environments should be considered as fixed or random. There is general agreement that in the fixed effects model the main effects should always be tested against the residual variation. However, in most experiments in which genotypes are tested in a range of environments it is implicitly assumed that the environments are a random sample of all possible environments, so that there is either a random or mixed effects model. The usual test here is to compare the effect of genotypes against the interaction, but, according to Plackett (1960), this test is extremely sensitive to departures from normality. In the discussion on Plackett's paper Sampford (1960) pointed out that the test is not justified, and proposed treating the analysis of variance here "as an Art Form rather than a sausage machine". In this situation Yates (1967) goes so far as to regard this or any other test as pointless, since genotype effects are bound to exist if there are interactions. In

practice, it is likely that most experimenters would not pay too much attention to a test of significance of main effects of genotypes when GE interactions have been shown to exist.

The problem of inference may be summarised as follows: what can one say about the main effects of genotypes and environments when there is GE interaction? When interactions are present, estimates of main effects are conditional: that is, one can only validly assert that genotypic effects are as observed *in this particular set of environments*, not over all possible environments. This was pointed out by Morley Jones and Mather, as quoted above, but is still not always realised, a point emphasised by Fripp and Caten (1973). Those who wish to infer the behaviour of genotypes in other environments should at least ensure that the new environments come, so far as is possible, from the same population from which the experimental environments are a sample. When an interaction is largely expressible in terms of a linear effect, then it may be reasonable to assume that for a new environment with a mean in the range of those tested the genotypes will behave in the same relative fashion as before. Even this may be dangerous if some of the experimental environments are sub-optimal and other are super-optimal, as pointed out by Knight (1970).

#### 8. DATA FROM A TWO-WAY TABLE REGARDED AS A MULTIVARIATE SET

In 1923 Fisher and Mackenzie had used least squares to get an approximation to what would now be regarded as the first principal component of the interaction between two main effects. For 30 years this remained an isolated result, until Williams (1952) extended it and showed that the sum of squares for interactions could be represented by the sum of the eigenvalues of a matrix and that if the sums of squares for either or both of the main effects were added to that for interaction the same result held with a different matrix. In the GE case, combining replicates and environments, the model for variation between and within genotypes becomes:

$$y_{ij} = \mu + d_i + w_{ij}, \quad (7)$$

where  $w_{ij}$  represents all the variation within genotypes, including that for error. The value  $w_{ij}$  may be considered as the element in the  $i$ th row and  $j$ th column of a matrix  $\mathbf{W}$ . Then, the eigenvalues of the matrix  $\mathbf{W}\mathbf{W}'$ , where  $\mathbf{W}'$  is the transpose of  $\mathbf{W}$ , may be found as solutions  $\theta_q^2$  of the determinantal equation  $|\mathbf{W}\mathbf{W}' - \theta^2\mathbf{I}| = 0$ , where  $\mathbf{I}$  is the unit matrix. Equation (7) may then be expressed in the following form:

$$y_{ij} = \mu + d_i + \theta_1 u_{1i} v_{1j} + \theta_2 u_{2i} v_{2j} + \theta_3 u_{3i} v_{3j} + \dots, \quad (8)$$

where

$$\sum_t u_{qt}^2 = \sum_j v_{qj}^2 = 1, \quad \text{and} \quad v_{qj} = \frac{1}{\theta_q} \sum_t w_{tj} u_{qt}.$$

In this formulation, due to Mandel (1969*b*), the vectors  $u_{qi}$  are the eigenvectors corresponding to the roots  $\theta_q^2$  and the corresponding values of  $\theta_q v_{qj}$  are the scores for the replicate-environment combinations. With  $r$  replicates,  $s$  environments and  $t$  genotypes, the number of non-zero values



of  $\theta$  in equation (8) is the smaller of  $(t-1)$  and  $rs$ . This result, which links analysis of variance and principal component analysis, is of considerable importance but was neglected for many years after Williams first derived it, possibly because it is difficult without electronic computers to calculate eigenvalues and always difficult to assess their significance.

Tukey (1962), though not mentioning Williams' work, suggested the use of a "vacuum cleaner" to break down values in a two-way table into four parts, (i) dual regression on row and column means simultaneously, (ii) deviations of row regression from dual regression, (iii) deviations of column regression from dual regression, (iv) residuals. In the course of this very wide-ranging paper Tukey also said: "The problem of determining the eigenvectors and eigenvalues associated with a two-way array is beginning to appear in diverse places in data analysis". Williams' method was rediscovered by Gollob (1968) and Mandel (1969*b*), who applied it to psychological testing and samples from chemical analytical data respectively. Both gave methods for judging the significance of interactions, while Mandel related all the various methods, including the vacuum cleaner, to each other. He further showed that most of the significance tests, including Gollob's, suffered from misconceptions regarding the number of degrees of freedom to allocate to the eigenvalues. Mandel calculated these degrees of freedom from Monte Carlo studies and refined the method to make it easier to apply (Mandel, 1971), noting again that it is not necessary to eliminate main effects before extracting principal components. Finally (Mandel, 1972), he made the point that the assessment of significance in principal component analysis by means of the criterion "per cent of the trace" is a very poor guide; with random numbers the first three components account for about 70 per cent. of the trace in a  $9 \times 11$  table and over 80 per cent. in a  $5 \times 19$  table.

Corsten and van Eijnsbergen (1972) and Johnson and Graybill (1972*b*) both gave the full statistical analysis underlying the principal component approach, and derived the appropriate distribution for a test of significance of the first component. This distribution cannot be written in an explicit form, but Mandel's approximation was shown by Johnson and Graybill to be close to the exact form obtained by numerical integration in a few cases where a comparison was possible. The estimate of the residual error variance  $\sigma^2$  is obtained after subtracting the sums of squares and degrees of freedom appropriate to any effects found to be significant. If no specific form is assumed for the interaction in equation (1) an estimate of  $\sigma^2$  was given by Johnson and Graybill (1972*a*), but this may be rather cumbersome.

The principal component approach has been used to investigate interactions in various contexts, such as Mandel's use of them in interlaboratory studies, as described in various of his papers cited, and the analysis of response curve data (Snee, 1972*a*); however, it has not yet been much used in GE studies, though it may be useful when regression on the environmental mean shows wide deviations from linearity. A version of it has been used by Perkins (1972) without applying Mandel's test for significance. Freeman and Dowker (1973) have taken it further by considering the components of the variation within genotypes, combining replicates and environments, and then doing analysis of variance on the resulting scores; the variation within environments is examined similarly. In both these studies the first principal component of the variation within genotypes is closely related to the main

effect of environments, but Freeman and Dowker find that the converse does not necessarily hold, the first component within environments not always being the main effect of genotypes. It would seem unwise to generalise about what results may be expected from the limited number of analyses so far published. Analysis of variance on principal component scores has also been done by Snee (1972*b*) when studying shapes of different carrot varieties. Here, positions of measurements on carrot roots formed one main effect while genotype and environment were the two factors in the other main effect.

The techniques of factor analysis and cluster analysis have also been used. Grafius and Kiesling (1960) used factor analysis methods to construct orthogonal vectors representing environmental effects, and thus predict genotypic responses in terms of these vectors. This work has also been extended to cover components of yield, but this, though interesting, is outside the scope of the present review. In cluster analysis an attempt is made to find similarities between clusters (environments, here) on the basis of measurements taken on the individuals of a cluster, the measurements being the genotypes grown there. The first attempt to do this was by Abou-El-Fittouh, Rawlings and Miller (1969*a*), using cotton. Because they had only a few genotypes common to all their environments, the usefulness of the method is hard to establish from this paper; furthermore, there are various methods of calculating similarities, and these may affect the clusters obtained.

A fairly new multivariate technique is *principal co-ordinate analysis* (Gower, 1966), which is, in a sense, the dual of principal component analysis. That is, if there is a matrix  $\mathbf{W} = (w_{ij})$ , as in equation (7), principal component analysis uses the matrix  $\mathbf{W}\mathbf{W}'$  whose size is the number of variates, there environments  $\times$  replicates: if the dual matrix  $\mathbf{W}'\mathbf{W}$  is considered, whose size is the number of individuals or genotypes, then the technique is principal co-ordinate analysis. The number of non-zero eigenvalues of the two matrices is the same. Gower gives a method for examining distances, suitably defined, between individual genotypes, and finding co-ordinates for each individual which are referred to principal axes which preserve these distances. Shukla (1972*b*) analysed GE interactions in a series of winter wheat variety trials over three years by various techniques: he regressed genotypic values on the environmental means, calculated principal components of the variation within genotypes and also used principal co-ordinate analysis in an attempt to find clusters of genotypes. The conclusions from this set of analyses were as follows: (i) the heterogeneity of regression on environmental means did not account for a large part of the interaction; (ii) the interaction contained three or four significant principal components; (iii) a few varieties clustered together, consistently over the years, but for most varieties no clustering was found.

## 9. USE OF EXTERNAL VARIABLES TO ASSESS THE ENVIRONMENT

The environment may be assessed by physical factors such as climatic measures, and regressions of yield on environmental variables have been calculated and those for individual genotypes compared by various workers. For example, Dowker (1963) related yield to rainfall for two varieties of maize at four plant population levels, finding that the slopes of the regression lines were very different. This type of study was formalised by Abou-El-

Fittouh, Rawlings and Miller (1969*b*), who partitioned the environment and genotype  $\times$  environment sums of squares for a series of cotton trials into items for regression and deviations from regression.

Freeman and Perkins suggested that the whole regression approach should be based on the use of an independent measure of the environment, either biological or physical. They made proposals as to how this should be done, and carried out an analysis in which a genotype similar to those under test was used to assess the environment. This was followed up by Fripp and Caten (1971), who found that, with a large number of genotypes, the results for regression on a control genotype differed little from those for regression on the mean of the test genotypes. Further, Fripp (1972) compared both biological and physical measures of the environment and found that the analyses, again for a large number of genotypes, gave very similar results for all reasonable external measures and the environmental mean. Finally, Perkins and Jinks (1973) regressed values for a large number of genotypes on their environmental means and on values derived from other, closely related, sets of genotypes; they found that all analyses generally gave similar values for significance, but that regressions on means derived from only a few independent genotypes were sometimes so insensitive as to give rise to problems of interpretation.

Hardwick and Wood recommended multiple regression on external variables as a means of finding the underlying physiological causes of observed regression on the environmental mean, and gave an example where predictions from this approach were successful. Perkins (1972) found that differences between two groups of genotypes could be explained in terms of multiple regression on climatic factors. A method of canonical analysis was briefly considered by Hardwick and Wood, and taken further by Wood (1972) and Shukla (1972*b*). In this method, a linear combination of genotypic values is regressed on linear combination of environmental variables. This approach differs from that of canonical correlation, in which the two linear combinations are chosen so as to maximise the correlation between two sets of variables: here, the aim is to find that combination of environmental variables accounting for the greatest amount of the variation between genotypes. Shukla found that three of his varieties were particularly affected by variation in environmental variables, but that, for his example, there was no clear advantage of canonical analysis over regression analysis on environmental variables. Wood, however, obtained by this technique a more logical explanation of variation in the growth of grasses from different sites than was obtained by any other technique.

Another approach is to regress simultaneously on the genotypic mean over all environments as well as the environmental mean, as suggested by Wright (1971) and, in a rather different context, Mandel (1969*a*). Utz also examines this possibility, but points out that the *F*-tests obtained are only approximate. The condition for this approach to be successful, namely that the regression lines for different genotypes are concurrent, is rather restrictive, and is not likely to be met very often in practice, though an example of reasonably good concurrence for over a hundred clones of grass is given by Troughton (1970). If the regression lines are concurrent, Wright suggests that it may then be possible to predict the performance of a hypothetical genotype in a hypothetical environment, and, in particular, to predict the performance of single cross hybrids from a diallel cross experiment. Even

here, however, prediction may well not be very accurate unless genes are highly correlated in the parental material.

#### 10. THE INTERPRETATION OF GE INTERACTIONS AND THE USE MADE OF THE RESULTS

When any statistical analysis is carried out the basic question always is: what use is to be made of the results? It is only a secondary question to enquire: how are the data to be analysed? Too much concentration on the second point leads to what has been described as the third kind of statistical error, the right answer to the wrong question. Thus, it is not surprising that different methods of analysis have been evolved to answer what is apparently the same question; the context in which the question is asked is all-important. There may well have been the tendency frequently observed in scientific methodology to use only home-grown methods: however, the worst examples of misuse of the techniques mentioned here have usually been by people who took them over from their originators but used them in very different situations. One particular danger is that of failing to recognise the conditional nature of much of the inference, and thus to assume, for example, that genotypes are stable in some absolute sense rather than merely more stable than the rest of those under test, the tests having been conducted in a given set of environments. It is usually assumed that environments are random, and this may be so, but the genotypes tested rarely are.

Of the techniques discussed, there can be no doubt that, for geneticists, the most fruitful has been the regression approach. For its success, a very high proportion of the interaction sum of squares should be explained by linear regression. When, as in the work of Bucio Alanis, Perkins and Jinks, very good linearity is found by regressing results from different generations of inbred lines on midparental means, the method is unchallengeable, predictions across generations being remarkably good. The conditions making for success, *i.e.* linearity of regression, are very difficult to determine, and one set of characters has frequently been found to give linear regressions, while other characters measured on the same set of genotypes have not.

In practice there are often wide deviations from linearity: thus, Witcombe and Whittington (1971), working with rape, found that "regression techniques to characterise genotype responses to the environment are an over-simplification". In such cases some plant breeders and agronomists have then found the concept of two stability parameters useful. However, Easton and Clements (1973) compared various methods for assessing stability and found them to be only moderately good at identifying genotypes with atypical behaviour; by choosing a subset of genotypes it was possible to make a previously stable genotype appear unstable. Similarly, Fripp and Caten (1973) found that the selection of a subset of environments changed the relation between mean performance and both linear and non-linear components of stability. Nevertheless, the use of two parameters sometimes provides a means of examining yield stability or instability: thus, Joppa, Lebsock and Busch (1971) found with wheat that many large values of Eberhart and Russell's  $s^2_{dii}$  could be explained by interactions between genotypes and specific pathogens, though others were still not explained. All these results point to the conclusion that there is often more than one way in which responses differ. That is, the interaction contains more than one

significant principal component or, in the words of Hardwick and Wood, "there is variation between genotypes in more than one dimension".

A measure of the suitability of the environment for discriminating between genotypes is also needed, as pointed out by Utz. Indeed, Dowker (1971) suggests that in a practical breeding programme it is desirable to capitalise on the GE interactions present in order to find those environments in which the genotypic effects of interest may be maximised. It may sometimes be sufficient just to know the structure in a two-way table, but it is often more important to know it in terms of some other variable. Plant physiologists especially are concerned to find the underlying causes of differences between genotypes.

For all these purposes univariate and regression methods will undoubtedly continue to be used, but it is not likely that any essentially different ones will be tried. With the range of modern computing devices available, new techniques are much more likely to be multivariate. Principal component analysis, cluster analysis, factor analysis, principal co-ordinate analysis and various types of canonical analysis may all be useful. Often, multivariate methods yield answers giving insight into particularly complex situations, and this may well happen in the study of GE interactions.

*Acknowledgments.*—I should like to thank those taking part in the Symposium on Genotype  $\times$  Environment Interactions for their helpful comments on the version of this paper presented there, and for the additional references to which they drew my attention. Particular thanks are due to my colleagues, Mr B. D. Dowker and Dr J. T. Wood, for their comments on a draft of the paper.

## 11. REFERENCES

- ABOU-EL-FITTOUH, H. A., RAWLINGS, J. O., AND MILLER, P. A. 1969a. Classification of environments to control genotype by environment interactions with an application to cotton. *Crop Sci.*, 9, 135-140.
- ABOU-EL-FITTOUH, H. A., RAWLINGS, J. O., AND MILLER, P. A. 1969b. Genotype by environment interactions in cotton—their nature and related environmental variables. *Crop Sci.*, 9, 377-381.
- ALLARD, R. W., AND BRADSHAW, A. D. 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.*, 4, 503-508.
- BAKER, R. J. 1969. Genotype-environment interactions in yield of wheat. *Can. J. Plant Sci.*, 49, 743-751.
- BARTLETT, M. S. 1947. The use of transformations. *Biometrics*, 2, 39-52.
- BOGYO, T. P., AND RUSSELL, T. S. 1972. Some properties of the statistics used in the analysis of genotype-environment interactions. *Biometrics*, 28, 1161-1162.
- BOLTON, A., AND SCAIFE, M. A. 1969. Maize variety trials in Western Tanzania, 1963-1966. *E. Afr. agric. for. J.*, 35, 11-20.
- BRESE, E. L. 1969. The measurement and significance of genotype-environment interactions in grasses. *Heredity*, 24, 27-44.
- 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, 399-405.
- 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.
- CORSTEN, L. C. A., AND VAN EIJSBERGEN, A. C. 1972. Multiplicative effects in two-way analysis of variance. *Statist. Neerlandica*, 26, 61-68.
- DOWKER, B. D. 1963. Rainfall reliability and maize yields in Machakos district. *E. Afr. agric. for. J.*, 28, 134-138.



- DOWKER, B. D. 1971. Variation studies in carrots as an aid to breeding. I. Concepts. *J. hort. Sci.*, 46, 485-497.
- EASTON, H. S., AND CLEMENTS, R. J. 1973. The interaction of wheat genotypes with a specific factor of the environment. *J. agric. Sci., Camb.*, 80, 43-52.
- EBERHART, S. A., AND RUSSELL, W. A. 1966. Stability parameters for comparing varieties. *Crop Sci.*, 6, 36-40.
- ELSTON, R. C. 1961. On additivity in the analysis of variance. *Biometrics*, 17, 209-219.
- FINLAY, K. W., AND WILKINSON, G. N. 1963. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.*, 14, 742-754.
- FISHER, R. A., AND MACKENZIE, W. A. 1923. Studies in crop variation. II. The manurial response of different potato varieties. *J. agric. Sci., Camb.*, 13, 311-320.
- FREEMAN, G. H., AND DOWKER, B. D. 1973. The analysis of variation between and within genotypes and environments. *Heredity*, 30, 97-109.
- FREEMAN, G. H., AND PERKINS, JEAN 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.
- FRIPP, YVONNE J. 1972. Genotype-environmental interactions in *Schizophyllum commune*. II. Assessing the environment. *Heredity*, 28, 223-238.
- FRIPP, YVONNE J., AND CATEN, C. E. 1971. Genotype-environmental interactions in *Schizophyllum commune*. I. Analysis and character. *Heredity*, 27, 393-407.
- FRIPP, YVONNE J., AND CATEN, C. E. 1973. Genotype-environmental interactions in *Schizophyllum commune*. III. The relationship between mean expression and sensitivity to change in environment. *Heredity*, 30, 341-349.
- GHOSH, M. N., AND SHARMA, D. 1963. Power of Tukey's test for non-additivity. *Jl. R. statist. Soc. B*, 25, 213-219.
- GOLLOB, H. F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*, 33, 73-116.
- GOWER, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53, 325-338.
- GRAFIUS, J. E., AND KIESLING, R. L. 1960. The prediction of the relative yields of different oat varieties based on known environmental variables. *Agron. J.*, 52, 396-399.
- GRAYBILL, F. A., AND WORTHAM, A. W. 1956. A note on uniformly best unbiased estimators for variance components. *J. Am. statist. Ass.*, 51, 266-268.
- HAMAKER, H. C. 1955. Experimental design in industry. *Biometrics*, 11, 257-286.
- HANSON, W. D. 1964. Genotype-environment interaction concepts for field experimentation. *Biometrics*, 20, 540-552.
- HANSON, W. D. 1970. Genotypic stability. *Theoret. Appl. Genetics*, 40, 226-231.
- HARDWICK, R. C., AND WOOD, J. T. 1972. Regression methods for studying genotype-environment interactions. *Heredity*, 28, 209-222.
- HILL, J., AND PERKINS, JEAN M. 1969. The environmental induction of heritable changes in *Nicotiana rustica*. Effects of genotype-environmental interactions. *Genetics*, 61, 661-675.
- HILL, J., AND SAMUEL, C. J. A. 1971. Measurement and inheritance of environmental response amongst selected material of *Lolium perenne*. *Heredity*, 27, 265-276.
- JINKS, J. L., AND PERKINS, JEAN M. 1970. Environmental and genotype-environmental components of variability. VII. Simultaneous prediction across environments and generations. *Heredity*, 25, 475-480.
- JOHNSON, D. E., AND GRAYBILL, F. A. 1972a. Estimation of  $\sigma^2$  in a two-way classification model with interaction. *J. Am. statist. Ass.*, 67, 388-394.
- JOHNSON, D. E., AND GRAYBILL, F. A. 1972b. An analysis of a two-way model with interaction and no replication. *J. Am. statist. Ass.*, 67, 862-868.
- JOPPA, L. R., LEBROCK, K. L., AND BUSCH, R. H. 1971. Yield stability of selected spring wheat cultivars (*Triticum aestivum* L. em Thell) in the uniform regional nurseries, 1959 to 1968. *Crop Sci.*, 11, 238-241.
- JOWETT, D. 1972. Yield stability parameters for sorghum in East Africa. *Crop Sci.*, 12, 314-317.
- KNIGHT, R. 1970. The measurement and interpretation of genotype-environment interactions. *Euphytica*, 19, 225-235.
- KNIGHT, R. 1971. A multiple regression analysis of hybrid vigour in single crosses of *Dactylis glomerata* L. *Theoret. Appl. Genetics*, 41, 306-311.
- MANDEL, J. 1959. The measuring process. *Technometrics*, 1, 251-267.
- MANDEL, J. 1961. Non-additivity in two-way analysis of variance. *J. Am. statist. Ass.*, 56, 878-888.

- MANDEL, J. 1969a. A method for fitting empirical surfaces to physical or chemical data. *Technometrics*, 11, 411-429.
- MANDEL, J. 1969b. The partitioning of interaction in analysis of variance. *J. Res. natn. Bur. Stand.*, 73B, 309-328.
- MANDEL, J. 1971. A new analysis of variance model for non-additive data. *Technometrics*, 13, 1-18.
- MANDEL, J. 1972. Principal components, analysis of variance and data structure. *Statist. Neerlandica*, 26, 119-129.
- MANDEL, J., AND LASHOF, T. W. 1959. The interlaboratory evaluation of testing methods. *ASTM Bulletin*, 239, 53-61.
- MANDEL, J., AND MCCrackin, F. L. 1963. Analysis of families of curves. *J. Res. natn. Bur. Stand.*, 67A, 259-267.
- MATHER, K. 1971. On Biometrical Genetics. *Heredity*, 26, 349-364.
- MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics*. 2nd edition. Chapman and Hall, London.
- MATHER, K., AND MORLEY JONES, R. 1958. Interaction of genotype and environment in continuous variation. I. Description. *Biometrics*, 14, 343-359.
- MILLER, P. A., ROBINSON, H. F., AND POPE, O. A. 1962. Cotton variety testing: additional information on variety  $\times$  environment interactions. *Crop Sci.*, 2, 349-352.
- MILLER, P. A., WILLIAMS, J. C., AND ROBINSON, H. F. 1959. Variety  $\times$  environment interactions in cotton variety tests and their implications on testing methods. *Agron. J.*, 51, 132-134.
- MILLIKEN, G. A., AND GRAYBILL, F. A. 1970. Extensions of the general linear hypothesis model. *J. Am. statist. Ass.*, 65, 797-807.
- MORLEY JONES, R., AND MATHER, K. 1958. Interaction of genotype and environment in continuous variation. II. Analysis. *Biometrics*, 14, 489-498.
- PARODA, R. S., AND HAYES, J. D. 1971. An investigation of genotype-environment interactions for rate of ear emergence in spring barley. *Heredity*, 26, 157-175.
- PERKINS, JEAN M. 1970. Environmental and genotype-environmental components of variability. VI. Diallel sets of crosses. *Heredity*, 25, 29-40.
- PERKINS, JEAN M. 1972. The principal component analysis of genotype-environmental interactions and physical measures of the environment. *Heredity*, 29, 51-70.
- PERKINS, JEAN M., AND JINKS, J. L. 1968a. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity*, 23, 339-356.
- PERKINS, JEAN M., AND JINKS, J. L. 1968b. Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. *Heredity*, 23, 525-535.
- PERKINS, JEAN M., AND JINKS, J. L. 1971. Specificity of the interaction of genotypes with contrasting environments. *Heredity*, 26, 463-474.
- PERKINS, JEAN M., AND JINKS, J. L. 1973. The assessment and specificity of environmental and genotype-environmental components of variability. *Heredity*, 30, 111-126.
- PLACKETT, R. L. 1960. Models in the analysis of variance. *Jl. R. statist. Soc. B*, 22, 195-217.
- PORCEDDU, E. 1970. The environmental component and the genotype  $\times$  environment interaction in plant selection work. *Genet. agr.*, 24, 129-144.
- QUALSET, C. O., AND GRANGER, R. M. 1970. Frequency dependent stability of performance in oats. *Crop Sci.*, 10, 386-389.
- REICH, V. H., AND ATKINS, R. E. 1970. Yield stability of four population types of grain sorghum, *Sorghum bicolor* (L.) Moench, in different environments. *Crop Sci.*, 10, 511-517.
- ROJAS, B. A. 1973. On Tukey's test of additivity. *Biometrics*, 29, 45-52.
- ROWE, P. R., AND ANDREW, R. A. 1964. Phenotypic stability for a systematic series of corn genotypes. *Crop Sci.*, 4, 563-567.
- RUSSELL, W. A., AND EBERHART, S. A. 1968. Testcrosses of one- and two-ear types of corn belt maize inbreds. II. Stability of performance in different environments. *Crop Sci.*, 8, 248-251.
- ST-PIERRE, C. A., KLINCK, H. R., AND GAUTHIER, F. M. 1967. Early generation selection under different environments as it influences adaptation of barley. *Can. J. Plant. Sci.*, 47, 507-517.
- SALMON, S. C. 1951. Analysis of variance and long-time variety tests of wheat. *Agron. J.*, 43, 562-570.
- SAMPFORD, M. R. 1960. Discussion on "Models in the analysis of variance" by R. L. Plackett. *Jl. R. statist. Soc. B*, 22, 211-212.
- SAMUEL, C. J. A., HILL, J., BREESE, E. L., AND DAVIES, ALISON. 1970. Assessing and predicting environmental response in *Lolium perenne*. *J. agric. Sci., Camb.*, 75, 1-9.

- SCHEFFE, H. 1959. *The Analysis of Variance*. John Wiley, New York.
- SHUKLA, G. K. 1972a. Some statistical aspects of partitioning genotype-environment components of variability. *Heredity*, 29, 237-245.
- SHUKLA, G. K. 1972b. Application of some multivariate techniques in the analysis and interpretation of the genotype-environment interaction. Paper presented at the Symposium on Genotype  $\times$  Environment Interactions, Birmingham University, September, 1972.
- SNEE, R. D. 1972a. On the analysis of response curve data. *Technometrics*, 14, 47-62.
- SNEE, R. D. 1972b. A useful method for conducting carrot shape studies. *J. hort. Sci.*, 47, 267-277.
- SPRAGUE, G. F., AND FEDERER, W. T. 1951. A comparison of variance components in corn yield trials. II. Error, year  $\times$  variety, location  $\times$  variety, and variety components. *Agron. J.*, 43, 535-541.
- TAL, G. C. C. 1971. Genotypic stability analysis and its application to potato regional trials. *Crop Sci.*, 11, 184-190.
- TROUGHTON, A. 1970. Intra-varietal variation of yield in two varieties of *Lolium perenne* L. *Euphytica*, 19, 382-389.
- TUKEY, J. W. 1949. One degree of freedom for non-additivity. *Biometrics*, 5, 232-242.
- TUKEY, J. W. 1962. The future of data analysis. *Ann. math. Statist.*, 33, 1-67.
- UTZ, H. F. 1972. Die Zerlegung der Genotyp  $\times$  Umwelt-Interaktionen. *EDV Med. & Biol.*, 3, 52-59.
- WARD, G. C., AND DICK, I. D. 1952. Non-additivity in randomized block designs and balanced incomplete block designs. *N.Z. J. Sci. Technol.*, 33, 430-435.
- WESTERMAN, JANE M. 1971. Genotype-environment interaction and developmental regulation in *Arabidopsis thaliana*. II. Inbred lines; analysis. *Heredity*, 26, 93-106.
- WESTERMAN, JANE M., AND LAWRENCE, M. J. 1970. Genotype-environment interaction and developmental regulation in *Arabidopsis thaliana*. I. Inbred lines; description. *Heredity*, 25, 609-627.
- WILLIAMS, E. J. 1952. The interpretation of interactions in factorial experiments. *Biometrika*, 39, 65-81.
- WITCOMBE, J. R., AND WHITTINGTON, W. J. 1971. A study of the genotype by environment interaction shown by germinating seeds of *Brassica napus*. *Heredity*, 26, 397-411.
- WOOD, J. T. 1972. The multivariate approach to genotype-environment interactions. Paper presented to the Symposium on Genotype  $\times$  Environment Interactions, Birmingham University, September 1972.
- WRICKE, G. 1962. Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z. Pflzücht*, 47, 92-96.
- WRIGHT, A. J. 1971. The analysis and prediction of some two factor interactions in grass breeding. *J. agric. Sci., Camb.*, 76, 301-306.
- YATES, F. 1967. A fresh look at the basic principles of the design and analysis of experiments. *Proc. 5th Berkeley Symp. math. Statist. Probab.*, 4, 777-790.
- YATES, F., AND COCHRAN, W. G. 1938. The analysis of groups of experiments. *J. agric. Sci., Camb.*, 28, 556-580.