Methods for analysing multi-site plant variety trials
II. Selection for yield and stability

G. Ye, D.L. McNeil and G.D. Hill

Plant Sciences Group, PO Box 84, Lincoln University, Canterbury, New Zealand

Abstract

This is the second of two manuscripts describing statistical methods for analysing multi-site plant variety trials. In this paper, we summarise methods for estimating and comparing genotypic means across sites. These include analysis of variance, nonparametric methods for testing genotypic effects, and some methods for selecting high-yield and stable genotypes including joint regression, segmented regression, principle co-ordinate analysis, general superiority measure, yield-stability statistic, safety-first, expected utility maximisation and desirability index.

Additional key words: *ANOVA, nonparametric method, mixed model*

Introduction

The existence of Genotype by Environment (GE) interactions is well documented in the plant breeding literature. Although selecting specific genotypes for specific environments is the best way to utilise GE interactions, it may not be practicable. GE interactions usually cannot be related to a single or even a few environmental factors, thus it may be not possible to group cultivation environments into groups which give the same GE response. If the cultivation environments can be grouped, then limited resources for plant breeding may dictate that stable genotypes with wide applicability are the best option. Because good performance is the main objective of plant breeding, selection of stable genotypes means that simultaneous selection for good performance and stability is needed. Therefore, breeders must weigh the importance of a genotype's stability relative to its mean performance across sites. Methods have been suggested to assist in the simultaneous selection of yield and stability and some of these are summarised in this paper.

Estimating and comparing genotypic means across sites

The genotypic mean across sites is an important indicator of the potential of the genotype to be successful in the whole production region. Therefore, an accurate assessment of yield performance of new genotypes across environments is crucial for plant breeding programs (Cullis *et al*., 1996).

Estimating genotypic means across sites

The simplest way to estimate the across-site mean is to take the arithmetic average. However, the arithmetic average is an unbiased estimate only if all genotypes are tested at all the sites and the genotypic mean variance is homogenous.

When not all the sites contain all the genotypes, but the genotypic mean variance is homogenous, the least square means can be used to account for the different numbers of test sites for different genotypes. This is the 'fitting constants' method (Searle, 1971). When genotypes are absent in environments with generally high responses, their means are corrected upwards. Means for genotypes absent in unfavourable environments will be corrected downward (van Eeuwijk, 1995). The general linear model (GLM) for analysis of variance provided by statistical software such as SAS and Genstat includes the fitting constants method as one option. However, the assumption underlying this method is that there is no genotype-by-site interaction; this clearly is unsatisfied by most of the multi-site tests.

Gauch and Zoble (1990) proposed a method based on an expectation maximisation algorithm to input the missing values by use of the additive main effects and multiplicative interaction effect (AMMI) model. The

AMMI model was discussed in the first paper of this review (Ye *et al*., 2001). The Gauch and Zoble (1990) method involves the following steps: 1) Compute cell means for every cell with data, then initialise the additive parameters by computing the unweighted genotype means, environment means and grand mean. 2) Initialise the interaction residuals as usual for cells with data, but input an interaction residual of zero for missing cells. 3) Solve the multiplicative parameters. 4) Re-estimate and revise each missing cell with the current AMMI model. 5) Fit the AMMI model to these revised data. 6) Iterate this process until convergence, i.e., the imputed missing values show acceptably small changes.

A more efficient method to deal with unbalanced GE data is to use a mixed model by assuming either genotype or environment effects to be random. Using this method the predicted means of the unobserved cells (genotype and environment combination) can be obtained by replacing the necessary terms in the model with their expected value. In other words, all the random effects are set to zero and the fixed effects to their generalised least squares estimate. Indeed, the predicted mean for any cell corresponds to the hypothetical means that would have been obtained if the data were orthogonal and equally replicated (van Eeuwijk, 1995). Whether the genotype or environment effect is assumed to be random depends on the situation. In general, an effect can be regarded as a random effect if the levels of the effect may reasonably be assumed to come from a probability distribution (Maclean *et al*., 1991; Stroup and Multize, 1991; Piepho, 1994). In practice the random effect should have sufficient degrees of freedom (say 10) to allow proper checking of the distributional assumptions about this effect (Stroup and Multize, 1991).

When the genotypic mean variances are not homogeneous, the estimation of genotypic means becomes more complicated. Unfortunately, genotypic mean variances are usually not homozygous across sites due to the heterogeneity of the genotype-by-environment interaction which is an indicator of different stabilities of tested genotypes and/or the heterogeneity of the within site experimental errors. In this case, the across-site genotypic means (arithmetic) are mostly affected by the mean from the sites with larger variances, and the generalised or weighted least squares method is a method to get more precise estimates. Using this method, observations in different sites are weighted by the reciprocal of the error mean squares of the site. Thus, the genotypic means from the sites with higher precision have a greater influence on the estimation. By defining a weight factor in the GLM model, the weighted least squares means can be obtained easily if all effects are regarded as fixed. Another method is to use a mixed model. The heterogeneous variances can be taken into account in mixed model analysis by correctly defining the two variance/covariance matrices (see Ye *et al*., 2001). SAS procedure MIXED is very useful in analysing multi-site data sets. Using the 'REPEATED /SUB = SITE TYPE = UN' statement, a site-specific error variance will be generated, and the estimate for fixed effect and the prediction for random effect will be obtained.

Using within environment error variances estimated directly from the observed data as weights has been criticised by many authors. For instance, more weight is usually given to the less productive sites since the error variance is usually positively correlated to the environmental yield response (Gauch, 1988; Crossa, 1990). Some authors have argued that more weight should be put on the more productive sites because they are of more interest to the experimenters (Gauch, 1988; Crossa, 1990). To overcome this problem Cullis *et al.* (1996) and Frensham *et al.* (1997) first modelled the error variances as a function of the log of site mean yield and other environmental variables. The predicted error variances from the prediction equation were then used as weights for estimating genotypic means across sites. The advantages of this procedure are: 1) It provides insight into those factors affecting error variance. 2) The influence of data recording and transcription errors is reduced. 3) It does not require that the additivity or homogeneity assumption is true for the other random effects.

Using error variances as weights does not take the heterogeneity of the genotype-by-environment interaction mean square into account. This can be overcome by weighting the genotypic means by the reciprocals of the residual mean squares, which includes both the interaction and the error variances (Bernardo, 1992).

The heterogeneous interaction variance can also be dealt with using the MIXED procedure of SAS. Using the 'REPEATED / GROUP = GEN TYPE = UN' statement, a genotype-specific variance can be generated. As for the error variance, Frensham *et al.* (1997)

proposed a method that models the GE interaction variances by a log-linear function of the explanatory variables. This approach produces a GE interaction variance for each genotype. Alternatively, Denis and Dhorne (1989) modelled the GE interaction variance using genotypic and/or environmental variables directly in the mixed model analysis. This is method is known as the 'mixed factorial regression'.

When the weighted least square method is used, the following points apply. Firstly, unless the variances are quite different, the simple arithmetic means are still valid, although the significance tests using a pooled error are no longer valid. Secondly, the weighted means can lose their expected superiority over the arithmetic mean if the estimated variances lack sufficient precision (Yates and Cochran, 1938). This was further confirmed by Bernardo (1992) using a maize yield trial containing 34 varieties and 53 environments. Thirdly, multi-site testing data usually exhibit large ranges in site mean yields so that much of the heterogeneity may be related to scale. Therefore, data transformation may be necessary to remove scale-dependent heterogeneity; otherwise, misleading interpretation of the heterogeneity may be obtained (Frensham *et al*., 1997).

Compare genotypic means across sites

Generally the purpose of multi-site testing is the statistical estimation of genotypic performance. Breeders are also interested in comparing the genotype means across environments. Analysis-of-variance (ANOVA) combined with multiple comparisons is normally used by breeders to achieve this objective. Nonparametric methods have also been suggested by different authors to be used when the assumptions underlying ANOVA cannot be satisfied.

ANOVA

For convenience, assume there are '*v*' genotypes tested in '*s*' environments with '*b*' replications in a randomised block design and '*n*' individuals planted within each plot. The general model for analysing a multi-site test based on cell (plot) means can be written as:

$$y_{ijk} = \mu + g_i + e_j + b_{k(j)} + (ge)_{ij} + \varepsilon_{ijk} \ ,$$

where y_{ijk} is the mean of the *i*-th genotype in the *j*-th environment in the *k*-th block, μ is the overall mean, g_i is the effect of the *i*-th genotype, e_j is the effect of the *j*-th environment, $b_{k(j)}$ is the effect of *k*-th block in *j*-th environment, $(ge)_{ij}$ is the interaction of the *i*-th genotype with the *j*-th environment, and ε_{ijk} is the error associated with the mean of the *i*-th genotype in the *j*-th environment in the *k*-th block.

When all genotypes are tested at all the sites and the within-site error variances are homogeneous, ANOVA based on the cell means is the simplest method to compare the genotypes. The ANOVA table is given in Table 1. If both genotypic and environmental effects are assumed to be fixed, or genotypic effects are random and the environmental effects are fixed, the significance of genotypic differences can be tested by F = MSG/MSE with (*v*-1) and (*v*-1)*s*(*b*-1) degrees of freedom. Assuming that genotypic effect is fixed and environmental effect is a random effect or if both effects are random, the significance of genotypic difference can be tested by F = MSG/MSGE with (*v*-1) and (*v*-1)(*s*-1) degrees of freedom.

Table 1. Combined analysis of variance for multi-site data.

source	df	MS	F
Genotype	*v*-1	MSG	MSG/MSGE
Site	*s*-1	MSE	
Block(Site)	(*b*-1)*s*	MSB	
Genotype*Site	(*v*-1)(*s*-1)	MSGE	
Error	(*v*-1)*s*(*b*-1)	MSE	

When the within-site error variances are not homogeneous, most practitioners use some form of transformation to remove the heterogeneity before ANOVA is done on the transformed data. However, the interpretation of the results from analysing transformed data may become difficult or biologically meaningless. For most multi-site tests, if ANOVA is used just to test the difference between genotypes, using original data directly does not create much bias (Crossa, 1990). However, the test for an interaction can be misleading because too many significant results are produced (Crossa, 1990). An alternative is to combine sites into groups that have homogenous within group variance, and the combined analysis is done for each group separately. However, the results from different groups cannot be combined to give a recommendation over all the sites.

Sometimes, instead of the plot means the genotypic means at each site have to be used to perform an ANOVA. In this case, the interaction and the experimental error cannot be separated. The ANOVA table is given in Table 2. The usual assumptions for an ANOVA are more difficult to satisfy because the interaction mean squares for genotypes are unlikely to be the same and the covariances between a pair of genotypes are also unlikely to be the same.

Table 2. ANOVA of multi-site testing data based on genotypic means at each site.

Source	df	MS	F
Genotype	*v-1*	MSG	MSG/MSGE
Site	*s-1*	MSE	
Residual	*(v-1)(s-1)*	MSGE	

However, an ANOVA can be valid under a less restrictive assumption about the variance-covariance structure. The sufficient and necessary condition for a valid ANOVA is the circularity structure of the variance-covariance matrix (Winer *et al.*, 1992). In the context of a GE two-way table, this condition requires that for each pair of genotypes '*i*' and '*k*', the quantity $\sigma_{ii} + \sigma_{kk} - 2\sigma_{ik}$ is a constant, where σ_{ii} and σ_{kk} are the variances of the *i*-th and *k*-th genotypes respectively, and σ_{ik} is the covariance between the *i*-th and *k*-th genotypes. In other words, the variance of the difference between the observations of genotype '*i*' and '*k*' in the same environment is the same.

When circularity is violated, the F-test for the significance of the differences between genotypes F = MSG/MSGE can be approximated by the F-distribution with $\varepsilon(v\text{-}1)$ and $\varepsilon(v\text{-}1)(s\text{-}1)$ degrees of freedom, where ε is a measure of departure from circularity. If the ε value is less than 0.8, the departure from the circularity is serious. The ε can be estimated by using the sample variances and covariances as suggested by Geisser and Greenhouse (1958):

$$\varepsilon = \frac{v^2(\bar{s}_{ii} - \bar{s}_{..})^2}{(v-1)\sum\sum s_{ik}^2 - 2v\sum s_{i.}^2 + v^2 s_{..}^2} \ ,$$

where s_{ik} is sample covariance of the *i*-th and *k*-th genotypes, $\bar{s}_{..} = \dfrac{\sum_i s_{ii}}{v}$ is the average of all genotype variances, $\bar{s}_{ik} = \dfrac{\sum_i \sum_k s_{ik}}{v^2}$ is the pooled covariance between all pairs of genotypes, and $\bar{s}_{i.} = \dfrac{\sum_k s_{ik}}{v}$ is the average covariance between the *i*-th genotype and all other genotypes.

Huynh and Feldt (1976) gave a modified estimate of ε,

$$\varepsilon_{adj} = \frac{s(v-1)\varepsilon - 2}{(v-1)[(s-1)-(v-1)\varepsilon]} \ ,$$

which is preferable when ε is not much smaller than unity.

When the number of sites is larger than the number of genotypes, the circularity of the covariance structure can be tested by Mauchley's procedure and the Hotelling multivariate T^2 test. Because this is rarely the case in plant variety tests, the detail of these methods are not given, but interested readers can refer to Winer *et al.* (1992) and Piepho (1996).

When multiple comparisons between means are made under the condition that circularity is violated, one may use paired t-tests. That is, to carry out two-way ANOVA for each pair of genotypes separately (the 'lsmeans' statement associated with MIXED of SAS provides this pair-wise comparison) and the experiment-wise error rate can be controlled using the Bonferroni procedure: the paired comparisons are performed at the $2\alpha/(v\text{-}1)v$ significant level, where α is the predetermined significance level.

Nonparametric methods

In the previous section a method was introduced to deal with the situation when the underlying assumptions of the ANOVA are violated. Some nonparametric methods, which do not rely on any restrictive assumption, may be used as alternatives. All nonparametric methods transform the original observations into ranks, which are then used for subsequent analyses. Because we are only interested in testing the genotypic effect, in the following sections, only methods for testing the genotypic effect are introduced.

1. Hildebrand (1980) method: The original observations are expressed as the differences from the replication mean, then the differences are transformed into a singe rank order. The test statistic,

   $$\frac{12}{v(vsb+1)}\sum_{i=1}^{v}(\bar{R}_{i..}-\bar{R}_{...})^2 \ ,$$

   is approximately χ^2-distributed with (v –1) degrees of freedom.

2. Kubinger (1986) method: The original observations are ranked into a single rank order (R_{ijk}), then the ranks are transformed by subtracting the average rank over replicates ($\bar{R}_{ij.}$) and adding the overall rank of the genotype ($\bar{R}_{i..}$), that is, $R^*_{ijk} = R_{ijk} - \bar{R}_{ij.} - \bar{R}_{i..}$. The test statistic,

   $$\frac{12}{v(vsb+1)}\sum_{i=1}^{v}(\bar{R}^*_{i..}-\bar{R}_{...})^2 \ ,$$

   is approximately χ^2-distributed with (v –1) degrees of freedom.

3. Van der Lann-de Kroon (1981) method: The original observations are ranked for each site separately into the ranks (R_{ijk}). The test statistic,

   $$\frac{12}{vsb^2(sb+1)}\sum_{i=1}^{v}\bar{R}^2_{i..}-3v(sb+1) \ ,$$

   is approximately χ^2-distributed with (v –*1*) degrees of freedom.

When comparisons among genotypes are made the sign test or the signed Wilcoxon test can be used (they are available in most commercial statistical software). Again, the significance level needs to be modified using the Bonferroni procedure. In addition, the Spearman and the Kendall rank coefficient can also be used for each pair of genotypes.

Selection for yield and stability

Joint regression analysis

Finlay and Wilkinson (1963) developed a method to study genotypic stabilities using multi-site testing data. This method is now known as joint regression analysis. It consists of regressing genotypic means at each site onto the environmental indexes defined as the environmental means. The stable genotypes are those with regression coefficients less than one.

Westcott (1987) method

Westcott (1987) proposed a method based on principle coordinate analysis. His definition of the dissimilarity between two genotypes in a given environment is

$$S_j(i,k)=\frac{y_{mj}-(y_{ij}-y_{kj})/2}{y_{mj}-y_{lj}} \ ,$$

where y_{mj} and y_{lj} represent the genotypes with the highest and lowest mean performance in the j-th environment, respectively; y_{ij} and y_{kj} are the mean performance of genotypes i and k in the j-th environment, respectively. When more than one environment is considered the similarity between the i-th and k-th genotypes is the mean of $S_j(i,k)$ across environments. The measure of similarity between any pair of genotypes compares their average performance with the best genotype (y_{mj}) in a given environment. Genotypes with smaller $S_j(i,k)$ values are closer to y_{mj}.

This method can be used for selecting stable genotypes. The testing environments are ranked in descending order according to their means (i.e., environmental index); the sites outside the lower and upper quartile are the poor and good sites. Genotype performance is first analysed for the poorest site, next the two poorest sites, and so on. The same procedure is applied to the good sites. For each cycle of analysis, a two-dimensional diagram is developed that represents the first two principle coordinates. Genotypes that have consistently shown an above average performance throughout the cycles are the most stable genotypes.

General superiority measures

Lin and Binns (1988) defined the measure of general superiority P_i as the mean square of the distance between a genotype's response and the maximum response at each site averaged over all sites. They demonstrated that P$_i$ may be regarded as the mean square (MS) of the joint effect of the genotypic (G) and GE interaction, thus

$$P_i = \sum_{j=1}^{s} \left(y_{ij} - y_{mj}\right)^2 / 2v \ ,$$

where y_{mj} is the maximum response among all genotypes in the *j*-th site: the smaller the value the better the genotype.

Yield-stability statistic

Kang (1991) proposed the 'rank sum method'. Ranks are assigned to mean yield with the highest yield receiving the rank of one and another rank is assigned to the stability variances with the lowest value having a value of one. Then the yield rank and stability rank are summed for each genotype. Genotypes with smaller rank sums are preferred. Kang (1993) modified this method and gave it another name, the yield-stability statistic (YS$_i$). The necessary calculations are as follows:

1. Rank genotypes according to yield, the genotype with lowest yield receives a rank of one.
2. Adjustment of yield rank: +1 if the genotype mean yield is higher than the overall mean yield for a test (OMY); +2 and +3 if the genotype mean yield is higher than OMY by one least significance difference (LSD) or two LSDs or more respectively; -1 if the genotype mean yield is lower than OMY; -2 and -3 if the genotype mean yield is lower than OMY by one LSD or more and lower than OMY by two LSDs or more. The adjusted rank was labelled (Y$_i$).
3. Assignment of stability rating (S$_i$): S$_i$ = 0 if stability variance is not significant; and –2, -4 and -8 if it is significant at 10%, 5% and 1% probability levels, respectively.
4. Compute and select genotypes: YS$_i$ = Y$_i$ + S$_i$.

The genotypes that have YS$_i$ values larger than the average are selected.

Segmented regression analysis

When selecting for wide adaptation for variable environmental conditions, the selected genotypes should ideally possess relatively high yield and stable performance in high stress environments. At the same time the genotypes should possess the capability to respond positively to favourable environments. Therefore, the environments are grouped into high-yielding and low-yielding groups first, then the response pattern for each group is fitted to a linear model by joint regression (Finlay and Wilkinson, 1963). The ideal genotypes are the genotypes with regression coefficients in low-yielding environments less than one and regression coefficients in high-yielding environments larger than one.

Safety-First

Eskridge (1990) introduced a decision-making concept known as safety-first to the selection of stable genotypes. The model he used was the Kataoka (1963) model. Based on this model, the general safety-first index (SFI) is

$$\text{SFI}= \bar{y}_{i.} - Z_{(1-\alpha)} S_i^2 \ ,$$

where α is an acceptable probability of having a disastrously low performance, $\bar{y}_{i.}$ is the sample mean yield across sites for the *i*-th genotype, S_i^2 is a measure of stability for the *i*-th genotype, and $Z_{(1-\alpha)}$ is the (1-α) percentile from a standard normal distribution.

The genotypes with larger SFI values are the desirable ones. Eskridge (1990) developed the safety-first index for several commonly used stability parameters.

Expected utility maximisation

Eskridge and Johnson (1991) introduced the expected utility maximisation (EUM) to select stable plant cultivars. It can be separated into four major steps:

1. Enumeration of all possible choices: if the single 'best' genotype from a set of genotypes is selected, then the list of all possible choices is simply the set of genotypes being evaluated.

2. Define utility function: to evaluate the genotypes by the utility function the utility function should have the following characteristics. First, if a breeder prefers A to B, then the utility of A is larger than B. Second, the scale on which the utility is defined is arbitrary, which means that the ordering of geno-

types must not change under a positive linear transformation. Finally, it is likely to be a concave function of performance, where the curvature of the utility function defines the breeder's attitude toward stability. The more curved the utility function, the greater the importance placed on stability. Eskridge and Johnson (1991) used the negative exponential utility function as the functional form. Therefore, $U(Y)=1- e^{-aY}$, where 'a' is defined as the stability preference coefficient, $a \geq 0$.

3. Specify a probability distribution of genotype response, $f(y_{ij})$: the performances of each genotype in all environments are rarely fully tested and need to be predicted. Application of EUM to selection requires these 'predictions' be made in terms of a probability distribution for each genotype. Eskridge and Johnson (1991) assumed that genotypic performance is normally distributed, and that sample estimates of mean and variance from trials were used to replace the unknown true parameters.

4. Calculate the indices for selection of stable genotypes based on EUM: the 'value' the breeder may expect to obtain from the *i*-th genotype is simply the expected value of the utility of genotype performance, i.e.,

$$E[U(y_{ij})] = \int U(y_{ij}) f(y_{ij}) dy_{ij} \quad ,$$

   where integration is over all possible yields.

If the performance of the *i*-th genotype in the *j*-th environment y_{ij} is normally distributed with the mean $E(y_i)$ and variance $V(y_i)$, then the general form of an expected utility index is

$$E(y_i)-(a/2)\ V(y_i)\ .$$

The genotype with largest index value is considered to be the 'best'. In practice, a stability model needs to be chosen so as to estimate $E(y_i)$ and $V(y_i)$.

Hernandez *et al.* (1993) desirability index

Hernandez *et al.* (1993) proposed a desirability index that is expressed as the area under the regression function. It can be written as

$$D_i = \bar{y}_{i.} + b_i C$$

where $\bar{y}_{i.}$ is the mean yield of the *i*-th genotype, b_i is the linear regression coefficient of the *i*-th genotype on the environmental index (*I*) which is defined as the mean of an environment minus the grand mean, and $C= \frac{I_a + I_b}{2}$ is the mean of the environmental indices at two extreme environments.

Conclusion

The heterogeneous variances of the genotypic means, within-site error variances and/or the heterogeneous GE interaction variances complicate the estimation of genotypic means across sites. The across-site genotypic mean of a genotype is a weighted mean of its means at every site. Therefore, how to determine the weights is very important. The method introduced in textbooks and currently applied by plant breeders is to use the within-site error variances as weights (Cochran and Cox, 1957). Modelling error variance using a function of other variables has been used to replace the sample error variances as weights. A mean that is more meaningful in the context of breeding may be more appreciated. With a multi-trait selection index appropriate weights need to be determined by the relative importance of each trait in determining the economic value of the genotypes. Similarly the genotypic means at different sites may also be weighted by the relative importance of the production conditions represented by each site.

For comparing the genotypic means across sites, the ANOVA plus multiple comparison method is often difficult to justify. Since the stability of a genotype should also be taken into consideration when selection is made, the possibility of two genotypes with very similar performance and stability may be rare. Though breeders may be more interested in ranking genotypes rather than detecting statistical significance, there is still a requirement to test for differences among genotypes by statistical methods. The method of Geisser and Greenhouse (1958) for the F test, and nonparametric methods may be more appropriate.

The presence of significant genotype-by-environment interactions in multi-site testing is the rule rather than the exception. Selection for genotypes with good stability has always been the objective of breeders. Many methods for simultaneous selection of performance and stability have been developed recently. The relative performance of these methods under different situations is unknown. The general superiority measure and Westcott method use relative performances at each site, do not require an explicit stability parameter in developing the selection criteria, and are simpler. The main disadvantage of these two methods is that breeders cannot make a subjective evaluation of the importance of the stability. The desirability index requires that the genotypic response can be explained by a linear model. Because a linear model rarely models the genotypic response satisfactorily, this index may be of limited use in practice. Similarly, the segmented regression method requires that linear models in each environmental group can model the genotypic responses. The rank sum method, yield-stability statistic, the safety-first method and the expected utility maximisation procedures all use stability parameters explicitly. Therefore, they all face the problem of selecting a form of stability parameter to use since different types of stability parameters may end up with the selection of different genotypes. The safety-first and the expected utility maximisation procedures need to define a probability distribution of the genotypic performance across sites and the parameters of the distribution need to be estimated accurately. Thus they can only be used when the number of sites is large. Another obvious difficulty of the expected utility maximisation procedure is the specification of a utility function.

References

Bernardo, R. 1992. Weighted vs. unweighted mean performance of varieties across environments. *Crop Science* ***32***, 490-492.

Cochran, W.G. 1954. The combination of estimates from different experiments. *Biometrics* ***10***, 101-119.

Cochran, W.G. and Cox, G.M. 1957. Experimental Design. 2nd Ed, John Wiley & son Inc. New York.

Crossa, J. 1990. Statistical analyses of multilocation trials. *Advances in Agronomy* ***44***, 55-85.

Cullis, B.R., Gleeson, A.C. and Thomson, F.M. 1982. The response to selection of different procedures for the analysis of early generation variety trials. *Journal of Agricultural Science* ***118***, 141-148.

Cullis B.R., Thomson, F.M., Fisher, J.A., Gilmour, A.R. and Thompson, R. 1996. The analysis of the NSW wheat variety database. I Modelling trial error variance. *Theoretical and Applied Genetics* ***92***, 21-27.

Denis J.B. and Dhorne, T. 1989. Modelling interaction by regression with random coefficients. *Biuletyn Oceny Odmian* ***21-22***, 65-73.

Eskridge, K. M. 1990. Selection of stable cultivars using a safety-first rule. *Crop Science* ***30***, 369-374.

Eskridge, K.M. and Johnson B.E. 1991. Expected utility maximisation and selection of stable plant cultivars. *Theoretical and Applied Genetics* ***81***, 825-832.

Frensham, A., Cullis, B.R. and Verbyla, A.P. 1997. Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* ***53***, 1373-1383.

Gauch HG 1988. Model selection and validation for yield trials with interaction. *Biometrics* ***44***, 705-715.

Gauch, H.G. and Zobel, R.W. 1990. Imputing missing yield trial data. *Theoretical and Applied Genetics* ***79***, 753-761.

Geisser, S. and Greenhouse, S.W. 1958. An extension of Box's results on the use of F-distribution in multivariate analysis. *Annals of Mathematical Statistics* ***29***, 885-891.

Gilmour, A.R., Cullis, B.R. and Verbyla, A.P. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics* ***2***, 269-293.

Hernandez, C.M., Crossa, J. and Castillo, A. 1993. The area under the function: an index for selecting desirable genotypes. *Theoretical and Applied Genetics* ***87***, 409-415.

Hildebran, H. 1980. Asymptoisch verteilungsfreie Rangtests in linearen Modellen. *Medicinal Information Statistics* ***17***, 344-349.

Huynh, H. and Feldt, L.S. 1976. Estimation of the Box correction for degree of freedom from sample data in randomised block and split plot designs. *Journal of Educational Statistics* ***1***, 69-82.

Kang, M.S.1991. Modified rank-sum method for selecting high yielding and stable crop genotypes. *Cereal Research Communication* ***19***, 361-364.

Kang, M.S. 1993. Simultaneous selection for yield and stability in crop performance trials: Consequences for the grower. *Agronomy Journal* ***85***, 754-757.

Kataoka, S. 1963. A stochastic programming model. *Econometrica* ***31***, 181-196.

Kempton, R.A. and Fox, P.N. (eds) 1997. Statistical Methods for Plant Variety Evaluation. Chapman & Hall.

Kubinger, K.D. 1986. A note on nonparametric tests for the interaction in two-way layouts. *Biometrics Journal* ***28***, 67-72.

Lin, C.S. and Binns, C.S. 1988. A superiority measure of cultivar performance for cultivar×location data. *Canadian Journal of Plant Science* ***68***, 193-198.

Lin, C.S., Binns, C.S. and Lefkovitch, L.P. 1986. Stability analysis: Where do we stand? *Crop Science* ***26***, 894-900.

Little ,T.M. and Hills, F.S. 1978. Agricultural Experimentation: Designs and Enalysis. John Wiley & Sons Inc. New York.

Lowe, W.J. and Hatcher, A.V. 1983. Progeny test data handling and analysis. *In* Progeny Testing of Forest Trees. South Co-op Series Bull No 275. Texas A & M Univ. College Station, TX pp51-68.

Maclean, R.A., Sanders, W.L. and Stroup, W.W. 1991. A unified approach to mixed model theory. *American Statistician* ***45***, 54-64.

Pearce, S.G., Clarke, R.F. and Kempton, R.E. 1988. Manual of Crop Experimentation. London: Charles Griffin and Company Ltd, New York: Oxford University Press.

Piepho, H.P. 1994. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics* ***89***, 647-654.

Piepho HP. 1996. Comparing cultivar means in multilocation trials when the covariance structure is not circular. *Heredity* ***76***, 198-203.

SAS Institute 1997. SAS/STAT software: changes and enhancements through release 6.12. SAS Institute Inc., Cary, North Carolina, USA.

Searle, S. R. 1971. Linear Models. John Wiley & Sons. New York.

Snedecor, G.W. and Cochran, W.G. 1980. Statistical methods. 7th Ed. Iowa State University Press, Ames/IA.

Stroup W.W. and Mulitze DK. 1991. Nearest neighbour adjusted best linear unbiased prediction. *American Statistician* ***45***, 194-200.

Van Eeuwijk F. A., Keizer, L. C. P. and Baker, J. J. 1995. Linear and bilinear models for the analysis of multi-environment trials: II. An application to data from the Dutch maize variety trials. *Euphytica* ***84***, 9-22.

Van der Lann-de Kroon 1981. Distribution-free test procedures in two-way layouts: A concept for rank-interaction. *Statistics* ***35***, 189-213.

Westcott, B. 1986. Some methods of analysing genotype-environment interaction. *Heredity* ***56***, 243-253.

Westcott, B. 1987. A method of assessing the yield stability of crop genotypes. *Journal of Agricultural Science, Cambridge* ***108***, 267-274.

Winer, B.J., Brown, D.R. and Michelis, K.M. 1992. Statistical Principles in Experimental Design. 3rd Ed. McGraw-Hill, New York.

Yates, F. and Cochran, W.G. 1938. The analysis of groups of experiments. *Journal of Agricultural Science, Cambridge* ***28***, 556-580.

Ye, G., McNeil, D.L. and Hill, G.D. 2001. Some methods for analysing multisite plant variety trials. I. Estimating genotypic means at each site. *Agronomy New Zealand* ***31***, 13-24