J. agric. Sci., Camb. (1987), **108**, 267–274
With 3 text-figures
Printed in Great Britain

A method of assessing the yield stability of crop genotypes

By B. WESTCOTT

*Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT), Mexico**

(*Revised MS. received* 3 *June* 1986)

SUMMARY

A method of assessing genotypic stability is proposed in the case where genotype yields are measured across environments but no concomitant environment data are available. The method is introduced in a practical context and two examples of its use with actual data are presented.

The method depends ultimately on the choice of a suitable measure of similarity between genotypes. Most of the stability information appears in a sequence of plots, where stable genotypes are immediately highlighted as consistently more remote points.

INTRODUCTION

Methods used for examining genotype–environment interaction could also be used in the examination of genotypic stability. However, difficulties are apparent with some of these methods and with particular stability measures which have been put forward in the past, while other methods still have to be fully explored (Westcott, 1986). In particular, methods based on linear regression analysis (Yates & Cochran, 1938; Finlay & Wilkinson, 1963; Eberhart & Russell, 1966) have been criticized on various grounds: a definite model for the expected response function is unnecessary and may be misleading (Mungomery, Shorter & Byth, 1974), the proportion of the genotype–environment interaction sum of squares due to linear regression may be small (Baker, 1969; Byth, Eisemann & De Lacy, 1976), regression fits may be unduly influenced by performance in relatively few environments (Westcott, 1986) and, in practice, regression parameters may fail to identify stable or unstable genotypes (Easton & Clements, 1973). Moreover, the sum of squares for deviations from regression, which was regarded by Eberhart & Russell (1966) as an important component of stability assessment, is not independent of the slope of the regression line (Hardwick & Wood, 1972).

It might be thought that a genotype could be regarded as stable if it showed no genotype–environment interaction when analysed with other genotypes. However, two genotypes which gave zero means everywhere would exhibit no interaction with environment, but would hardly be worthy of consideration as a result. Clearly, the level of performance is also important in the practical assessment of genotypic stability. Eberhart & Russell (1966, p. 38) stated: 'In the past, the term "stable variety" often has been used to mean a variety that does relatively the same over a wide range of environments. This means that a "stable variety" by this definition performs relatively better under adverse conditions and not so well in favourable environments. Analyses of several sets of data from Iowa State University maize yield trials have indicated that hybrids with a regression coefficient less than 1·0 usually have mean yields below the grand mean. In situations where production does not give a surplus that can be stored, or where long storage is not possible, such a variety may still be the most desirable. However, under conditions such as exist for maize in the United States, the breeder usually wants a variety that does above average in all environments.' Verma, Chahal & Murty (1978) attempted to examine low-yielding and high-yielding environments separately by dividing the environmental range into two major regions and fitting a separate linear regression in each. Although it was an attempt at a more practical analysis, this method still suffers from those defects of the linear regression approach detailed by Westcott (1986) and outlined earlier. Putting aside considerations involving regression coefficients, the difference between the two types of situation mentioned by Eberhart & Russell (1966) may just be the choice of the set of environments to analyse. The variety that does not do so well in favourable environments may still be above average if the set of environments is restricted to adverse ones. From this point of view, stability can simply be regarded

* Present address: Plant Breeding Institute, Cambridge, CB2 2LQ.

as above-average performance in an appropriately chosen set of environments. However, consistency of performance over the set of environments has also to be taken into account.

METHOD

The method of assessment which is presented here is based on a suitable measure of similarity between genotypes. In a particular environment, if L and S denote the largest and smallest genotype yields, then the similarity between genotype yields x_i and x_j is defined by $s(x_i, x_j) = (L-(x_i+x_j)/2)/(L\text{-}S)$ if i and j are unequal, while $s(x_i, x_i) = 1$. The higher yielding the genotypes, as measured by their mean, the more dissimilar they will be according to this measure. The similarity is standardized by dividing by the yield range for the environment. When a set of environments is being considered, the similarity between x and y is just the mean of the similarities between x and y across environments.

A great advantage of the similarity matrix defined here is that in its principal coordinates analysis (Gower, 1966), no negative eigenvalues are obtained. A proof of this is given in the Appendix. Coordinates of points in a Euclidean space thus result, referred to principal axes, such that the distance between two points represents the dissimilarity between the corresponding genotypes. Each analysis produces a two-dimensional picture, in which the first two principal coordinates are plotted for each genotype. If distances are adequately approximated in this representation for a particular set of environments, genotypes which are above average yielding over these environments will be more dissimilar to the lesser yielding genotypes than the latter will be to each other and so will be represented by points which are more remote.

Such plots show their value when the stability assessment is based on the sequential accumulation of environments. The environments are first ranked in descending order of mean yield and the low- and high-yielding environments are then examined in cycles. Thus, for the low-yielding environments, the first cycle (called L1) involves the analysis of the lowest-yielding environment, the second cycle (L2) involves analysing the two lowest-yielding environments, the third cycle (L3) adds the next lowest-yielding environment and so on, the lowest-yielding environment of those remaining being added at each cycle. Similarly, cycles H1, H2, etc. involve the highest-yielding environment, the two highest-yielding environments, etc. Analysing these cycles produces a succession of pictures, in each of which the first two principal coordinates are plotted for each genotype. Neglecting information in the third and later principal coordinates can somewhat distort the relationship between genotypes revealed by such a picture but the distortion can be reduced by superimposing a minimum spanning tree (Gower & Ross, 1969). Such a superimposition also provides a natural centre for the picture as, with the above similarity matrix, a minimum spanning tree can be found such that all the branches radiate from a single node. A proof of this can also be found in the Appendix. The good genotypes are simply the ones furthest from the centre and their identification is generally immediate. The stable genotypes are then just the ones which are consistently good over cycles.

Table 1. *Mean yield (t/ha) of the 12 winter wheat varieties analysed in Example 1, ranked in descending order*

Rank	Variety number	Variety name	Mean
1	8	Hobbit	5·61
2	9	Sportsman	5·32
3	10	TJB259/95	5·09
4	5	Kinsman	5·06
5	12	Hustler	5·05
6	7	Durin	5·05
7	11	TJB325/464	4·77
8	2	Maris Ranger	4·72
9	4	Maris Templar	4·71
10	6	Maris Fundin	4·61
11	1	Cappelle-Desprez	4·42
12	3	Maris Huntsman	4·42

EXAMPLES

Example 1

The yield values for this example come from trials carried out by Blackman, Bingham & Davidson (1978) to compare the responses of semi-dwarf and tall varieties of winter wheat to nitrogen fertilizer. Twelve varieties were grown at seven sites with contrasting soil types. Two trials were grown at each site with a high and a low top dressing of nitrogen, respectively, making 14 environments in all.

Natural sets of environments to be analysed were the high- and low-nitrogen environments in turn. In the principal coordinates analysis of the high-nitrogen environments (the first two eigenvalues contributing 29 % of the total), the varieties Hobbit and Sportsman appear best, followed by Hustler (referred to as TJB368/268 by Blackman *et al.* (1978)) and TJB259/95. In the low-nitrogen environments (the first two eigenvalues amounting to 31 % of the total), Hobbit and Sportsman are again best, followed this time by Durin. These results agree with those obtained via tables of means.

Hobbit is clearly the top variety overall, Sportsman is clearly second, followed by TJB259/95, Kinsman, Hustler and Durin, which lie close together and form a group (Table 1). The four

Table 2. *Mean yields (t/ha) of the 14 environments used in Example 1, ranked in descending order*

Rank	Environment number	Site	Nitrogen	Mean
1	13	Edinburgh	Low	7·40
2	14	Edinburgh	High	7·05
3	11	Earith	Low	5·58
4	12	Earith	High	5·14
5	4	Begbroke	High	5·01
6	8	Trumpington	High	4·99
7	6	Fowlmere	High	4·94
8	10	Boxworth	High	4·75
9	9	Boxworth	Low	4·46
10	2	Crafts Hill	High	4·35
11	7	Trumpington	Low	4·29
12	5	Fowlmere	Low	4·19
13	3	Begbroke	Low	3·28
14	1	Crafts Hill	Low	3·19

highest-yielding environments involve only two sites, the Edinburgh trials being clearly the highest-yielding, while the Earith trials outyielded the remainder (Table 2). Surprisingly, the low-nitrogen trial had a higher mean yield in both of these sites, probably due to residual effects of previous treatments. The natural separation into low- and high-nitrogen environments for initial stability assessment may not, then, have been appropriate in this case. In stability assessment over cycles H1 to H5, Hobbit proves to be the most stable, Kinsman coming second while Durin and Sportsman also show signs of stability over high-yielding environments. In analyses over L1 to L5, Sportsman, Hobbit and Maris Templar prove to be stable.

Taking L4 as an example, Table 2 shows that the four lowest-yielding environments all had low nitrogen. The yields of each variety over these environments are ranked in Table 3.

In the principal coordinates analysis, the first two eigenvalues provide 31 % of the total. In the resulting picture (Fig. 1), Sportsman is the best variety, followed closely by Hobbit, the next best being (in order) Maris Templar, Durin, TJB259/95, Kinsman and Hustler. Incidentally, Maris Templar had by far the largest third principal coordinate and so is actually better than appears from the plot: it is stable over low-yielding environments, although ranked only ninth overall (Table 1).

The data values used in this example were also used by Kempton (1984) to produce biplots. The good performance of Sportsman and Hobbit in the four lowest-yielding environments (Fig. 1) shows up in his Finlay–Wilkinson biplot: the good performance of Maris Templar does not, however. His principal components biplot was based on residual yields after variety effects had been removed and so was not intended to reveal the actual performance of varieties. From Table 3 and similar tables, it can be seen that the plots used here are much closer to reality than Finlay–Wilkinson biplots when matters of stability are under consideration: further comparisons are needed with the principal components biplot based on raw yields.

Example 2

The yield values for this example come from a CIMMYT international maize trial (EVT 12) which was carried out in 1979. Twenty-two white-grained experimental varieties (Table 4) selected from CIMMYT's full season tropical maize populations 21, 22, 25, 29 and 43 and intermediate season populations 23 and 32 were grown in 29 sites (Table 5). The two highest-yielding varieties overall, La Maquina 7843 and Poza Rica 7822, performed well in the H cycles, the latter being best in H1 and

Table 3. *Means of all the varieties in Example 1 in each of the four lowest-yielding environments*

Variety	Site				Mean
	Crafts Hill	Begbroke	Fowlmere	Trumpington	
Sportsman	3·69	3·81	5·22	4·26	4·25
Hobbit	3·32	3·74	4·90	4·87	4·21
Maris Templar	3·46	3·60	4·55	4·13	3·94
Durin	3·14	3·18	4·57	4·66	3·89
TJB259/95	3·25	3·36	4·47	4·41	3·87
Kinsman	3·56	3·12	4·42	4·07	3·79
Hustler	3·22	3·39	3·73	4·53	3·72
Maris Ranger	2·85	3·28	4·18	4·09	3·60
TJB325/464	2·93	2·89	3·95	4·57	3·59
Cappelle-Desprez	3·21	3·17	3·64	4·08	3·53
Maris Huntsman	2·87	3·16	3·41	3·82	3·32
Maris Fundin	2·78	2·66	3·29	4·01	3·19
Mean	3·19	3·28	4·19	4·29	3·74

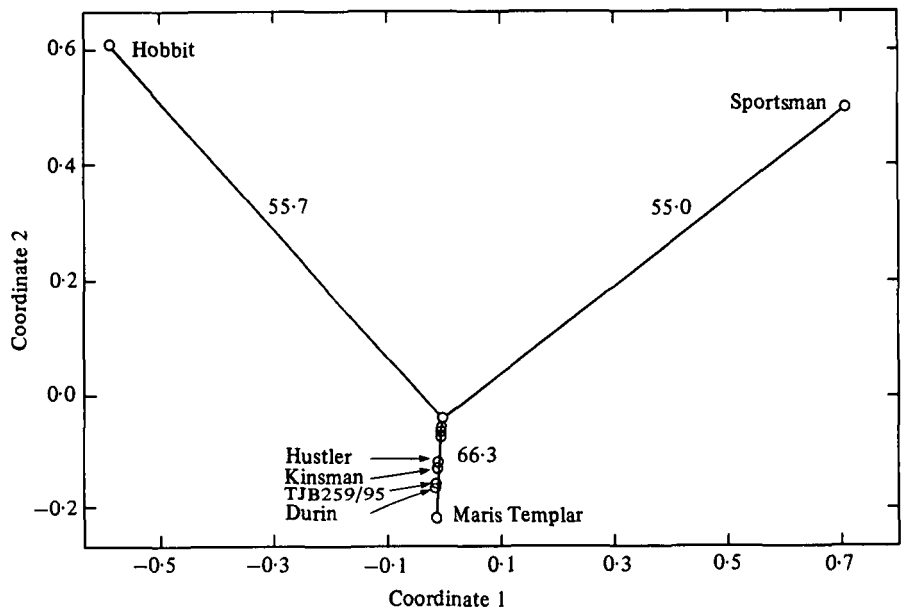

Fig. 1. Plot of the first two principal coordinates of a set of winter wheat varieties in cycle L4 of the assessment of Example 1 (see Table 3). Part of the minimum spanning tree is superimposed on the plot, the distances shown between varieties being the similarities expressed as percentages.

Table 4. *Varieties in CIMMYT trial EVT 12 in 1979 ranked in descending order by their mean yields (t/ha) over all sites*

Rank	No.	Name	Country of origin	Mean	*b**
1	12	La Maquina 7843	Guatemala	4·96	1·148
2	7	Poza Rica 7822	Mexico	4·87	1·093
3	11	Poza Rica 7843	Mexico	4·86	1·112
4	6	Across 7622	—	4·83	1·058
5	5	Dholi 7622	India	4·74	1·058
6	8	Guanacaste 7729	Costa Rica	4·70	1·043
7	10	Across 7729	—	4·68	1·074
8	9	Kisanga 7729	Zaire	4·61	1·094
9	14	San Andres (1) 7823	El Salvador	4·61	0·990
10	3	Across 7721	—	4·59	0·996
11	2	Gandajika 7721	Zaire	4·58	0·968
12	4	Ferkessedougou (2) 7622	Ivory Coast	4·54	1·031
13	1	San Andres 7721	El Salvador	4·40	0·988
14	15	Poza Rica 7823	Mexico	4·39	0·886
15	16	Santa Rosa 7823	Nicaragua	4·37	0·855
16	17	Alajuela 7725	Costa Rica	4·33	0·961
17	22	Across 7632	—	4·28	0·944
18	20	Poza Rica 7832	Mexico	4·27	0·885
19	13	Cotaxtla (INIA) 7623	Mexico	4·23	0·968
20	21	Across 7523	—	4·17	0·944
21	18	Kaniama 7725	Zaire	4·12	0·918
22	19	Across 7725	—	3·99	0·987

* If the variety yield at a site is plotted against the site effect (site mean yield–overall mean yield) for all sites, the *b* value for the variety is the slope of the regression line fitted to the plotted points by least squares (Finlay & Wilkinson, 1963).

Table 5. *Sites used in EVT 12 in 1979 ranked in descending order by site mean yield (t/ha)**

Rank	Site	Country	Mean	Rank	Site	Country	Mean
1	Sids	Egypt	7·08	16	Santa Cruz	El Salvador	4·12
2	Malkerns	Swaziland	7·04	17	Njala	Sierra Leone	3·98
3	Las Acacias	Honduras	6·78	18	Ekona	Cameroon	3·71
4	Ferkessedougou	Ivory Coast	6·26	19	San Andres	El Salvador	3·65
5	Potchefstroom	Rep. S. Africa	6·21	20	La Huerta Jal.	Mexico	3·62
6	Poza Rica	Mexico	6·07	21	Taiz	Yemen A.R.	3·56
7	Cotaxtla (INIA)	Mexico	5·86	22	San Cristobal	Dom. Rep.	3·53
8	Chirinas	Honduras	5·82	23	Ngabu	Malawi	3·26
9	Mount Makulu	Zambia	5·66	24	Nayarit (INIA)	Mexico	2·87
10	Hofuf	Saudi Arabia	5·43	25	Ilonga	Tanzania	2·86
11	Jamalpur	Bangladesh	5·38	26	Sefa	Senegal	2·74
12	La Maquina	Guatemala	5·08	27	Turipana	Colombia	2·68
13	Pergamino	Argentina	4·74	28	Ibadan (IITA)	Nigeria	2·32
14	Cuyuta	Guatemala	4·38	29	Sebele	Botswana	1·81
15	Afgoi	Somalia	4·20				

* Sites ranked 1–7 and 23–29 were used in the stability assessment cycles.

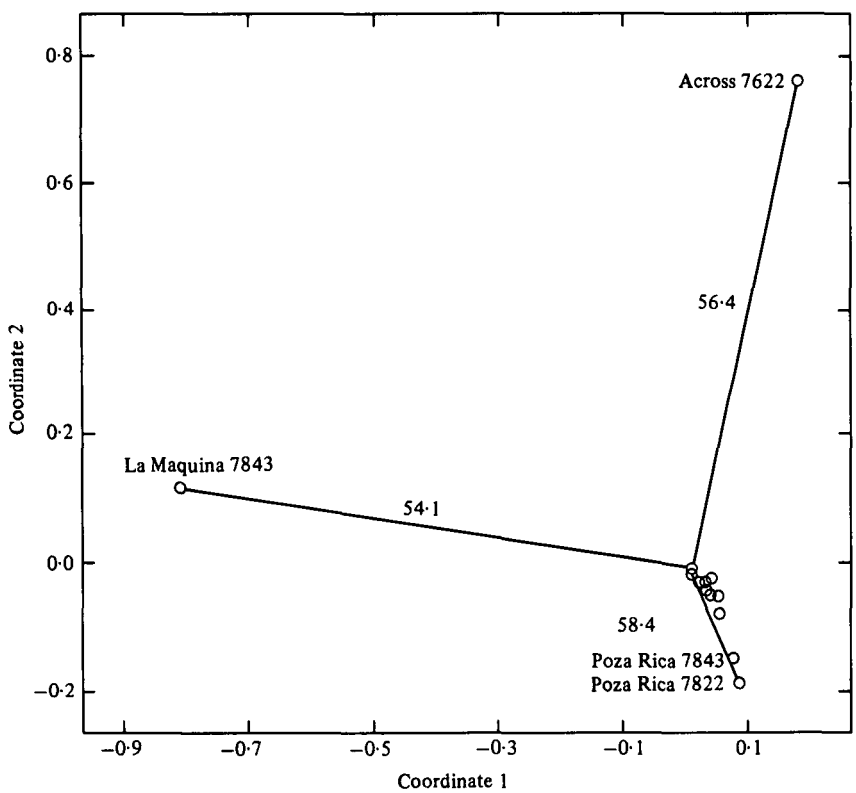

Fig. 2. Plot of the first two principal coordinates of a set of experimental maize varieties in cycle H5 of the assessment of the 1979 CIMMYT trial EVT 12 (Example 2). As before, part of the minimum spanning tree is superimposed on the plot, the distances shown between varieties being the similarities expressed as percentages.

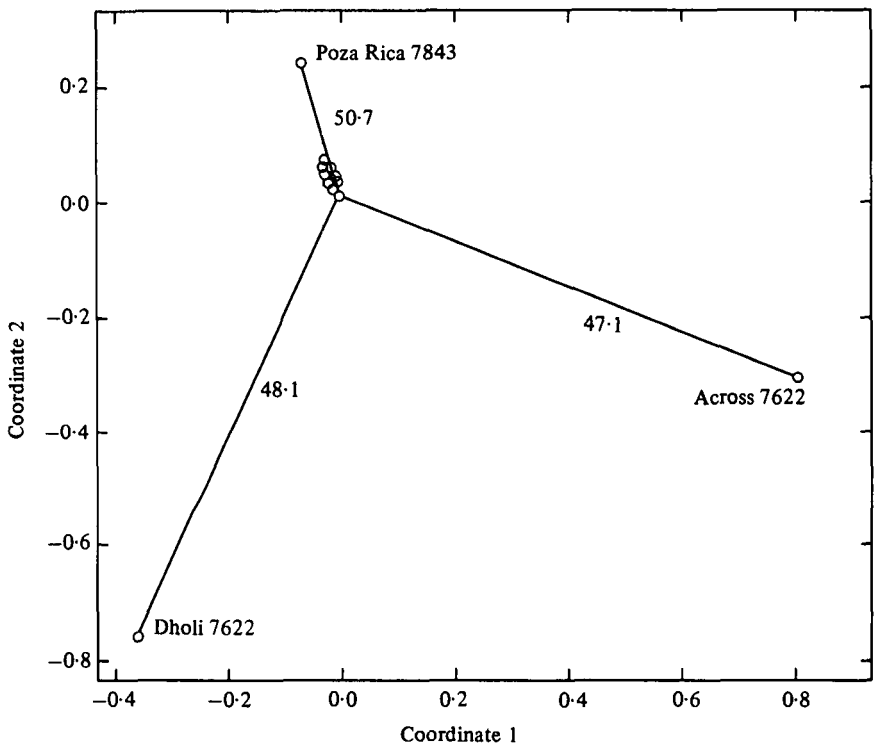

Fig. 3. Plot of the first two principal coordinates of a set of experimental maize varieties in cycle L3 of the assessment of the 1979 CIMMYT trial EVT 12 (Example 2). As before, part of the minimum spanning tree is superimposed on the plot, the distances shown between varieties being the similarities expressed as percentages.

H2 while the former was best in H3 to H7. However, while not doing badly, they did not do particularly well in the L cycles. Poza Rica 7843 (3rd overall) showed reasonable stability in both high- and low-yielding sites. In cycles L1 to L7, Across 7622 (ranked 4th overall) was clearly the most stable, being the most remote point in each cycle apart from L4 where it was second behind Dholi 7622. Across 7622 also performed well in the H cycles, particularly in H3 to H6 (Fig. 2 shows the picture for H5). Dholi 7622 (5th overall) showed good stability in low-yielding sites, being an outlying point in all cycles but hardly featuring in the H cycles. A typical picture for the L cycles (L3) is shown in Fig. 3.

It is instructive to compare the results in this example with those using the regression approach. According to Finlay & Wilkinson (1963), information on the adaptation of a variety could be obtained from the joint consideration of its regression coefficient and its mean yield. From Table 4, the varieties Dholi 7622, Across 7622 and Guanacaste 7729 all have about the same values for these two parameters and so, on this criterion, should show similar adaptation. However, according to the present assessment, Across 7622 is adapted to both low- and high-yielding sites, Dholi 7622 performs less well overall but is better adapted to low-yielding sites than to high, while Guanacaste 7729 is relatively adapted to neither.

DISCUSSION AND CONCLUSIONS

In these examples, the two-dimensional plot associated with an analysis over a particular set of environments immediately shows the best genotypes. The picture can be a good guide even when only a small proportion of the trace is due to the first two eigenvalues. When there are few genotypes, the interpretation is straightforward as the points are mostly distinct. When there are more genotypes, points representing the worst genotypes tend to cluster near the centre. If the order of these genotypes is of interest, care needs to be exercised in interpreting the points: some of them may have large differences in the third or later coordinates and so will be further away in reality than appears from the plot. Usually, however, attention is focused on the stable genotypes which are easily identified and there is no interest in the order of the least stable ones. If the plots are produced by computer then, as a

precaution, software is needed that displays all coincident points. The analysis of the above examples was done using the GENSTAT statistical package in which this facility is provided.

The calculations involved in this method need to be done by computer as the extraction of eigenvalues is very time-consuming by hand. This can be done by using either a library subroutine or an appropriate statistical package, preferably one that includes principal coordinates analysis, such as GENSTAT. Sets of the order of 50 genotypes can be dealt with comfortably in this way. If the number of genotypes is very large, however, the calculation of the similarity matrix and its eigenvalues may exceed the limits of some packages. In such cases, the set of genotypes can be partitioned into mutually exclusive subsets, the analysis being done in turn to identify the most stable genotypes in each subset. A final analysis is then done on the amalgamated set of these genotypes. This two-stage procedure works very well, provided, of course, that the number of stable genotypes identified in each subset is at least as big as the number wanted in the final outcome.

The method of stability assessment proposed here can highlight features of performance which might otherwise be overlooked. It is free from the shortcomings of regression methods, cluster analysis and principal components which were detailed by Westcott (1986). Further research is needed to compare it with relatively unexplored techniques like stochastic dominance procedures, correspondence analysis, biplots and other forms of multidimensional scaling (Westcott, 1986).

In conclusion, the method accurately reveals the highest-yielding crop genotypes in given sets of environments. Its main strength lies in its use over cycles of sequentially accumulated environmental sets, where successive pictures directly and immediately highlight the good genotypes while playing down the rest. The performance of the method in the examples shows promise: however, any recommendation for its use must, of course, depend on wider experience of its performance in practice.

This work was carried out at CIMMYT while the author was on sabbatical leave from the Plant Breeding Institute, Cambridge, England. Thanks are due to Gwynneth Fellowes for preparing the diagrams.

REFERENCES

BAKER, R. J. (1969). Genotype–environment interactions in yields of wheat. *Canadian Journal of Plant Science* **49**, 743–751.

BLACKMAN, J. A., BINGHAM, J. & DAVIDSON, J. L. (1978). Response of semi-dwarf and conventional winter wheat varieties to the application of nitrogen fertilizer. *Journal of Agricultural Science, Cambridge* **90**, 543–550.

BYTH, D. E., EISEMANN, R. L. & DE LACY, I. H. (1976). Two-way pattern analysis of a large data set of evaluate genotype adaptation. *Heredity* **37**, 215–230.

EASTON, H. S. & CLEMENTS, R. J. (1973). The interaction of wheat genotypes with a specific factor of the environment. *Journal of Agricultural Science, Cambridge* **80**, 43–52.

EBERHART, S. A. & RUSSELL, W. A. (1966). Stability parameters for comparing varieties. *Crop Science* **6**, 36–40.

FINLAY, K. W. & WILKINSON, G. N. (1963). The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* **14**, 742–754.

GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

GOWER, J. C. & ROSS, G. J. S. (1969). Minimum spanning trees and single linkage cluster analysis. *Applied Statistics* **18**, 54–64.

HARDWICK, R. C. & WOOD, J. T. (1972). Regression method for studying genotype–environment interactions. *Heredity* **28**, 209–222.

KEMPTON, R. A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science, Cambridge* **103**, 123–135.

MUNGOMERY, V. E., SHORTER, R. & BYTH, D. E. (1974). Genotype × environment interactions and environmental adaptation. I. Pattern analysis – application to soya bean populations. *Australian Journal of Agricultural Research* **25**, 59–72.

VERMA, M. M., CHAHAL, G. S. & MURTY, B. R. (1978). Limitations of conventional regression analysis – a proposed modification. *Theoretical and Applied Genetics* **53**, 89–91.

WESTCOTT, B. (1986). Some methods of analysing genotype–environment interaction. *Heredity* **56**, 243–253.

YATES, F. & COCHRAN, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science, Cambridge* **28**, 556–580.

APPENDIX

Proof that the principal coordinates analysis produces no negative eigenvalues

I am grateful to John Gower and Peter Digby of Rothamsted Statistics Department for providing this proof, which is considerably shorter than my original.

Given an $n \times n$ association matrix (a_{ij}), Gower (1966) described a procedure (known as principal coordinates analysis) for finding the coordinates of n points such that the distance d_{ij} between the ith and jth points is given by

$$d_{ij}^2 = a_{ii} + a_{jj} - 2a_{ij}. \qquad (1)$$

Suppose there are n genotypes and m environments. Let x_{ik} be the yield of the ith genotype in the kth environment and let L_k and S_k be the largest and smallest genotype yields in the kth environment. In any environment, we assume that not all the genotype yields are equal, so that $L_k > S_k$ for all k.

The similarity between genotypes i and j described in the text is given by

$$a_{ij} = \sum_{k=1}^{m} (L_k - (x_{ik} + x_{jk})/2)/m(L_k - S_k) \qquad (2)$$

if $i \neq j$, while $a_{ii} = 1$.

The $n \times n$ matrix (a_{ij}) is the association matrix for our principal coordinates analysis. It is well known that if each of several association matrices does not generate negative eigenvalues, then neither does their mean. It is sufficient to prove, therefore, that the association matrix for a particular environment, k, does not produce negative eigenvalues.

Rearranging (2), the association matrix for environment k is given by

$$A_k = \text{diag}\,[(x_{ik} - S_k)/(L_k - S_k)] + B, \qquad (3)$$

where B is the $n \times n$ matrix with elements

$$b_{ij} = (2L_k - x_{ik} - x_{jk})/2(L_k - S_k).$$

The principal coordinates analysis of any such matrix A_k finds the eigenvalues of the matrix given by

$$C_k = (I-N)\,A_k(I-N),$$

where I is the identity matrix and N is the square matrix of order n, all of whose values are $1/n$.

Substituting (3) in the expression for C_k, we find, on multiplying out, that $(I-N)\,B(I-N)$ is the zero matrix and so

$$C_k = (I-N)\,\text{diag}\,[(x_{ik} - S_k)/(L_k - S_k)](I-N).$$

Now $\text{diag}\,[(x_{ik} - S_k)/(L_k - S_k)]$ has no negative values and hence C_k is positive semi-definite and thus has no negative eigenvalues, as required.

Proof that a minimum spanning tree can be found with a single node

Given n points, then a tree spanning these points is any set of straight line segments joining pairs of points such that:

(1) no closed loops occur;
(2) each point is visited by at least one line;
(3) the tree is connected.

When the lengths of all $\binom{n}{2}$ segments are given, a minimum spanning tree (MST) is just a spanning tree of minimum length.

The first algorithm described by Gower & Ross (1969) to compute the MST assigns iteratively to the MST the shortest segment not yet assigned which does not form a closed loop with any of the segments assigned already. When a choice of several equal segments of minimum length occurs, any one may be selected, in which case there is not a unique MST. Initially, no segments have been assigned and iteration stops when the MST contains $(n-1)$ segments.

Using equations (1) and (2), the length d_{ij} of the segment joining the ith and jth points is given in our case by

$$d_{ij}^2 = \sum_{k=1}^{m} (x_{ik} + x_{jk} - 2S_k)/m(L_k - S_k). \qquad (4)$$

Let p be a value of i for which $\sum x_{ik}/(L_k - S_k)$ attains its minimum value. This value is usually (but not necessarily) unique. Thus

$$\sum_{k=1}^{m} x_{pk}/(L_k - S_k) \leqslant \sum_{k=1}^{m} x_{qk}(L_k - S_k),$$

for all $q \neq p$, with strict inequality if p is unique. Therefore, from equation (4), $d_{ip}^2 \leqslant d_{iq}^2$ for all i different from p and q. Taking square roots, the inequality is preserved, since distances must be non-negative, i.e.

$$d_{ip} \leqslant d_{iq} \text{ for all } q \neq p \text{ and all } i \neq p \text{ and } \neq q.$$

Thus, at each stage of the above algorithm, the shortest segment not yet assigned can always be taken to join the point p to some other. Thus, the point p is always the single node in an MST, even if it is not unique.