THE ANALYSIS OF VARIATION BETWEEN AND WITHIN GENOTYPES AND ENVIRONMENTS

G. H. FREEMAN and B. D. DOWKER

National Vegetable Research Station, Wellesbourne, Warwick

Summary

Methods of analysing variation between and within genotypes and environments are discussed. Principal component analysis, or an equivalent technique proposed by Mandel (1971) for examining interactions in two-way tables, is suggested as an appropriate method in many circumstances, followed by analysis of variance on these principal components for replicated data. Various techniques are applied to yields of carrots from a trial in which eight varieties were grown in 34 environments representing a set of 17 site/year combinations at two densities. The largest source of variation within genotypes is found to be that between environments but not conversely. Two other sources of variation are identified within genotypes and environments, one representing the interaction of varieties with site/year effects and the other their interaction with densities. Analysis of variance indicates the varieties and environments contributing to these interactions. The general implications of the use of principal component analysis are discussed, particularly in situations such as that with the carrot data where the method of joint regression analysis fails because the genotype-environment interaction contains more than one independent component.

1. Introduction

IT is a common practice in genetical studies, and in trials of varieties and breeding lines, to grow a series of genotypes in a range of different environments. If the genotypes all respond similarly to all the environments tested, their relative performance in other environments may be predicted with some confidence. More often, however, there are differences in the relative performance, and an analysis of the pattern of the variation both between and within the genotypes is necessary to determine which particular genotypes respond differently from the others. This should lead to a prediction of the performance of the same genotypes in other environments and, given the performance of two homozygous parent inbred lines and their F_1, prediction of the performance of generations derived from these lines.

Various statistical techniques have been used to assist in the analysis of this variation. The first stage has usually been to regard the data as coming from a two-way table, with genotypes and environments being considered as rows and columns; a straightforward analysis of variance is then performed, taking out main effects of genotypes and environments and the genotype-environment interaction, and assessing the residual variation by means of replicate × treatment interactions. If the genotype-environment interaction is significant, the next step is frequently to try to describe this interaction in terms of biologically meaningful parameters. It has been observed empirically that the performance of genotypes over a range of environments is frequently approximately linearly related to some measure of these environments. If an external measure is available, this can be used,

but this situation does not usually happen; the commoner situation is that in which environmental effects have to be assessed solely by means of the genotypes grown in these environments. The mean of all the genotypes has been used to assess the environment by a series of authors, for example, Yates and Cochran (1938), Finlay and Wilkinson (1963), Eberhart and Russell (1966) and Perkins and Jinks (1968). The technique, known as *joint regression analysis*, consists of regressing genotype effects on environmental means and examining the difference between genotypes both in their linear regression coefficients and deviations from linearity. The parameters have been included in biometrical genetical models of genotype-environmental interactions (Mather and Jinks, 1971). This method has been criticised on physiological grounds by Knight (1970) and on statistical grounds by Freeman and Perkins (1971), but nevertheless has been highly successful in predicting genotypic performance over a range of species and environments. An alternative method, using structural relationships instead of regression, has been given by Tai (1971).

Statistical examination of interactions in two-way tables has not been undertaken specifically in the context of genotype-environment interactions, having been mostly done theoretically and independent of the subject-matter. An early attempt was the single degree of freedom for non-additivity of Tukey (1949); Williams (1952), in a rather neglected paper, postulated that the effects of one factor might be explained by their being proportional at different levels of the other factor. Tukey (1962), in an attempt to obtain true experimental errors, used a " vacuum cleaner " to remove systematic effects from residuals after taking out effects of rows and columns. This procedure formally consists of regressions on row and column means, both separately and together; the original values are split into four parts—(dual regression) + (deviations of row regression from dual regression) + (deviations of column regression from dual regression) + (residuals). An alternative approach to the problem, in the context of samples analysed at various laboratories, has been adopted by Mandel and his co-workers and presented in a series of papers, of which the most accessible is Mandel (1971). The procedure here is to partition one or more terms of the analysis of variance table into multiplicative functions: usually, the partition will be on the interaction term, but one or more sweeps of Tukey's vacuum cleaner may already have been done, or one main effect may be added to the interaction before partitioning, in line with Williams' approach.

An apparently completely different approach to the analysis of variation between and within genotypes is by means of multivariate analysis. Multivariate techniques have usually been applied to groups of individuals on which a whole set of measurements have been taken. All the measurements of a set have usually been made on one individual of the group, but there is no theoretical need for this, and it is perfectly legitimate to regard the set of measurements on different phenotypes of a genotype as all being made on that individual genotype. The data can then be examined by any method of multivariate analysis, one possibility being *principal component analysis*. In this method, components of variation within genotypes are taken out successively so that each extracts the maximum possible variation remaining in the system and all are orthogonal to each other. The discussion so far has been in terms of variation within genotypes, but can equally well refer to variation within environments and, making a few more assumptions, to

residual variation after main effects between genotypes and environments have been eliminated. The dual nature of the use of principal component analysis may be seen particularly clearly when an attempt is made to identify those genotypes with large contributions to the genotype-environment interaction. The individual principal components of the variation within environments may be subjected to analysis of variance, provided that replicated observations are available, and by this means differences between genotypes may be tested for significance. Again, examination of variation within genotypes will pick out important differences between environments. It is the purpose of this paper to compare the method of principal component analysis with methods appropriate to a two-way table for the resolution of variation between and within genotypes and environments.

2. Methods of analysis

The response y_{ijk} of the kth replicate of the ith genotype in the jth environment can be expressed in various mathematical forms. If there are t genotypes and s environments, with r replicates, the basic model for many methods of analysis is:

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

where μ = grand mean over all replicates, genotypes and environments;
d_i = additive genetic contribution of the ith genotype $(i = 1, ..., t)$;
ϵ_j = additive environmental contribution of the jth environment $(j = 1, ..., s)$;
g_{ij} = genotype-environment interaction of the ith genotype in the jth environment;
e_{ijk} = residual variation contributed by the kth replicate $(k = 1, ..., r)$ of the ith genotype in the jth environment.

Equation (1) provides the basis for the straightforward analysis of variance, in which the effects of genotypes and environments, and their interaction, are examined. By expressing the interaction effect g_{ij} in the form $\beta_i z_j + \delta_{ij}$, where β_i is a genotypic parameter, z_j some measure of the environment and δ_{ij} a residual term, this same equation provides the model for regression methods of analysis. In particular, if z_j is taken as ϵ_j, the environmental mean, the basis for the joint regression analysis approach is obtained; the subsequent analysis is described by the proponents of this method quoted in the last section.

If only the components of variation between and within genotypes are considered, it is more convenient to express the model in a different form, namely:

$$y_{ij} = \mu + d_i + w_{ij},\qquad(2)$$

where w_{ij} represents all the variation other than between genotypes, *i.e.* w_{ij} is the component of variation within genotypes. To preserve the idea of a two-way table, the combined set of replicates and environments is regarded as a set of environments. Thus, equation (2) does not explicitly contain a term for residual variation, or variation between replicates.

The value w_{ij} in equation (2) may be considered as the element in the ith row and jth column of a matrix W. Then, the method of principal component analysis may be used as follows. If W' is the transpose of W and

I a unit matrix, the *latent roots* or *eigenvalues* of the matrix WW' may be found as solutions θ_q^2 of the determinantal equation $|WW' - \theta^2 I| = 0$. Then, (2) may be expressed as:

$$y_{ij} = \mu + d_i + \theta_1 u_{1i} v_{1j} + \theta_2 u_{2i} v_{2j} + \theta_3 u_{3i} v_{3j} + ..., \tag{3}$$

where $\sum_i u_{qi}^2 = 1$, $\sum_j v_{qj}^2 = 1$, and:

$$v_{qj} = \frac{1}{\theta_q} \sum_i w_{ij} u_{qi}.$$

The vectors u_{qi} are the *latent vectors* or *eigenvectors* corresponding to the roots θ_q^2. The corresponding values of $\theta_q v_{qj}$ are sometimes known as *scores* for the various environment × replicate combinations.

The use of latent roots ensures that all the terms in (3) are orthogonal; they may be written so that $\theta_1 u_{1i} v_{1j}$ is the linear term corresponding to the largest proportion of the variation within genotypes that can be accounted for by one term, and it is then called the *first principal component*. Similarly, if $\theta_2 u_{2i} v_{2j}$, $\theta_3 u_{3i} v_{3j}$, etc., are terms corresponding to the largest amounts of the remaining variation, they are the second, third, ..., principal components. The latent roots θ_q^2 are the successive amounts of variance extracted by the principal components. Equation (3) is the same as that obtained by Mandel (1971), except that Mandel usually, though not always, advocates taking out a component of variation between environments and explaining the interaction component in terms of linear functions.

Approximate tests of significance are available for seeing how many principal components are meaningful. The test usually used is an approximate chi-squared test due to Lawley (1956), there being no exact test; however, in order to carry out this test, all the components have to be calculated, and this may be very laborious when there are many environments. An alternative approach to significance testing has been adopted by Mandel, who uses an approximate F-test based on degrees of freedom obtained from simulation studies. This is preferable in practice when only a few components are likely to be significant. Anyway, in the biological context, it is not usually difficult to determine how many terms are meaningful, using an approximate significance test as a guide but not relying too heavily on it.

It is commonly, but not always, found that the first latent vector represents a general effect with the values of u_{1i} for varying i being only slightly different; then, v_{1j} for varying j is similar to the ϵ_j of equation (1). Whether or not this is so, the next stage of the analysis of the data is an analysis of variance on the scores v_{1j}, v_{2j}, ..., continued as far as the components are meaningful. This analysis, to be useful, requires that the observed environmental differences be separated again into those for replicates and those for genuine environmental effects.

The whole of this section, from equation (2) onwards, has been concerned with variation between and within genotypes. However, the methods apply in exactly the same fashion to variation between and within environments: indeed, they are used later in this paper both ways.

3. DESCRIPTION OF EXPERIMENTAL EXAMPLE

The analytical methods just discussed will be illustrated on data for yield from a series of carrot trials whose underlying concepts have been described

by Dowker (1971). The series consisted of eight carrot varieties (four Chantenay and four Autumn King types) each grown at two intended densities (86 and 258 plants/m², abbreviated as D_1 and D_2 respectively) at six sites over the three years 1968 to 1970. The varieties and sites, together with the abbreviations used for them in the text, are listed below:

Varieties

    Chantenay types (Ch):
        New Model Red cored (NMRC)
        Royal Chantenay (RCh)
        Long Chantenay Red cored (LCh)
        Chantenay Red cored (ChRC)
    Autumn King types (AK):
        Early Giant (EG)
        Autumn King (AuK)
        Flakkee French strain (FFS)
        Red Giant (RG)

Sites

    N.V.R.S., Wellesbourne, Warwick (W)
    Stockbridge House E.H.S., Yorks. (S)
    Arthur Rickwood E.H.F., Cambs. (R)
    Cawston, Norfolk (C)
    Burscough, Lancs. (M)
    Retford, Notts. (N)

The last three sites were on growers' holdings, and the soils were fen peat at site R, moss peat at M, and sand or sandy loam elsewhere.

Each individual trial (*i.e.* at one site in one year) had a split plot design with two replicates, having varieties as main plots and densities as sub-plots. The plot size harvested for estimates of yield was 2·09 m².

4. ANALYSIS OF VARIANCE AND JOINT REGRESSION ANALYSIS

An overall analysis of variance on yield in tons/acre is given in table 1. Dowker (1971) has described the partitioning of treatments, expectations of the mean squares and resultant variance ratio tests for this analysis. In the analysis main effects and interactions involving varieties are split into three, a comparison of types (Ch *v* AK) and comparisons within each type (Within Ch, Within AK). For this and subsequent tables, significant values at P = 0·05, 0·01 and 0·001 are indicated by one, two and three stars respectively. The analysis shows a number of significant main effects: differences in mean performance of varieties within both the Chantenay and Autumn King groups (although the overall difference between the two groups just failed to reach significance at the 0·05 level) and differences within site/year combinations and densities. Interactions between genotypes and environments are also found, with site/year combinations interacting with the Ch *v* AK effect and that within the Autumn King group; there are density interactions with the effects within both Chantenay and Autumn King groups. Finally, the two environmental components, effects of site/year combinations and densities, interact with each other.

These genotype-environment interactions may be further examined by joint regression analyses on environmental means, first on the overall data

TABLE 1

Analysis of variance

Source of variation	d.f.	Mean square and significance†
Sites × Years (S × Y)	16	3082·9***
Replicates within S × Y	17	12·0
Ch *v* AK	1	406·5
Within Ch	3	56·9*
Within AK	3	230·5***
S × Y × (Ch *v* AK)	16	96·7***
S × Y × (Within Ch)	48	17·9
S × Y × (Within AK)	48	31·2***
Error 1	117·6‡	16·6
Densities (D)	1	5352·5***
D × S × Y	16	102·6***
D × (Ch *v* AK)	1	74·8**
D × (Within Ch)	3	79·4***
D × (Within AK)	3	40·1*
D × S × Y × (Ch *v* AK)	16	9·6
D × S × Y × (Within Ch)	48	7·4
D × S × Y × (Within AK)	48	13·3
Error 2	134·4‡	10·7

† Expectations of mean squares and appropriate F-tests given in Dowker (1971).
‡ 1 degree of freedom lost because of a missing plot.

TABLE 2

(a) *Joint regression analysis (based on means over two replicates)*

Source of variation	d.f.	Mean square	F-test (1)†	(2)†
Genotypes (G)	7	91·0	—	—
Environments (E)	33	853·5	—	—
G × E	231	—	—	—
Heterogeneity of regressions:				
Ch *v* AK	1	3·3	< 1	< 1
Within Ch	3	27·9	4·16**	4·15**
Within AK	3	44·3	4·26**	6·58***
Remainder:				
Ch *v* AK	32	27·6	—	4·11***
Within Ch	96	6·7	—	< 1
Within AK	96	10·4	—	1·54**
Pooled error (from table 1)	252	6·73	—	—

Regression coefficients

NMRC	RCh	LCh	ChRC	EG	AuK	FFS	RG
0·028	0·071	−0·007	−0·136	0·122	−0·022	−0·132	0·075

(b) *Joint regression analyses separately for Chantenay and Autumn King*

Source of variation	d.f.	Chantenay		Autumn King	
		Mean square	F-test (1)	Mean square	F-test (1)
Genotypes (G)	3	28·2	—	115·7	—
Environments (E)	33	431·0	—	449·4	—
G × E	99	—	—	—	—
Heterogeneity of regression	3	29·8	4·49**	63·3	6·80***
Remainder	96	6·6	—	9·3	—

† Test (1) Heterogeneity of regression items against corresponding remainders.
    (2) Heterogeneity of regression and remainder terms against pooled error from analysis of variance converted to same scale by dividing by 2.

and then on the two groups of varieties separately, as shown in table 2. The Ch v AK genotype-environment interaction is not explained by differences between the linear regression slopes for the two groups, most of this inter-action being in the remainder term. However, the genotype-environment interaction within the Chantenay group is largely described by the linear regression item and that within the Autumn King group partly so, though with a significant remainder term as well. The regression coefficients from the separate analyses on the two groups are very similar to those for the combined analysis, allowing for the fact that each set of coefficients separately is constrained to sum to zero. An analysis by the method of Tai (1971) shows general agreement with that in table 2. The absolute values of Tai's coefficients of linear response must always be bigger than those of the regression coefficients, but the difference here is very small because the variation between replicates within environments is much less than between the environments themselves.

5. PRINCIPAL COMPONENT ANALYSES

Principal component analyses are presented for various partitions of the data. The number of significant components in each arrangement was assessed by Mandel's test and the results are shown in table 3. The sums of

TABLE 3

Mandel's test for significance of principal components

	Within genotypes		Within environments		Residual after elimination of main effects	
	d.f.	Mean square	d.f.	Mean square	d.f.	Mean square
Component:						
1	110·2	517·4***	82·2	49·9***	60·8	20·4***
2	93·4	32·0***	67·9	22·9**	47·6	12·4**
3	81·2	22·1**	57·9	17·3*	38·4	8·9
Residual	251·2	14·2	302·0	12·2	84·2	6·8

squares from which the mean squares shown were derived are the first, second and third latent roots. The third component within environments appears from this table to be just significant at the 0·05 level, but, as no biological meaning has been found for it, it has been assumed to be a chance effect and ignored henceforth. It will be seen from table 3 that the degrees of freedom in Mandel's test, which are obtained from simulation studies, are not necessarily integers.

Principal component analysis within genotypes

A principal component analysis within genotypes is presented in table 4. The first of the three significant components is seen from table 3 to be much the largest. The ranking of the environmental scores is different for each component, and the analyses of variance on these scores show significant effects of site/year combinations and densities, and an interaction between them, for the first component, a site/year effect for the second component and a density effect for the third component.

For the first component, the ranking of the environmental scores corres-ponds closely to the mean yields for each environment, and so this component

TABLE 4

Principal component analysis within genotypes

Genotypic latent vectors:

	Component		
Variety	1	2	3
NMRC	0·362	0·250	−0·010
RCh	0·378	0·404	0·000
LCh	0·351	0·305	0·372
ChRC	0·302	0·267	−0·164
EG	0·397	−0·535	0·676
AuK	0·342	−0·078	−0·332
FFS	0·305	0·070	−0·143
RG	0·379	−0·560	−0·497

Environmental scores:

	Component								
	1			2			3		
Site/ year	D1	D2	Mean	D1	D2	Mean	D1	D2	Mean
W68	8·4	36·9	22·7	−2·4	−2·8	−2·6	2·1	9·7	5·9
S68	−19·6	−8·3	−14·0	−1·3	2·9	0·8	−0·8	3·2	1·2
R68	1·6	3·4	2·5	−9·4	0·5	−4·4	−1·5	3·6	1·0
C68	−35·7	−11·2	−23·4	−0·5	2·5	1·0	−0·5	−0·1	−0·3
M68	−5·9	22·6	8·4	−16·4	−16·5	−16·5	−1·3	−0·0	−0·7
W69	2·2	16·4	9·3	−1·2	9·5	4·1	2·7	2·8	2·7
S69	−23·1	−13·7	−18·4	2·3	2·7	2·5	0·2	3·8	2·0
R69	−23·9	−12·7	−18·3	−0·9	1·6	0·4	1·2	0·4	0·8
C69	−6·6	22·8	8·1	−4·7	0·1	−2·3	4·0	2·5	3·2
M69	−15·3	14·0	−0·7	−0·2	1·0	0·4	−4·4	4·3	−0·1
N69	−13·6	−8·9	−11·2	6·8	8·7	7·8	−0·9	2·4	0·8
W70	44·5	59·9	52·2	9·5	9·6	9·5	−0·9	−1·4	−1·1
S70	14·4	48·9	31·6	2·0	−2·2	−0·1	−6·7	0·7	−3·0
R70	−2·0	21·3	9·7	3·1	1·7	2·4	−7·0	−0·1	−3·5
C70	−61·0	−57·4	−59·2	1·1	0·0	0·6	0·7	0·5	0·6
M70	28·2	47·5	37·9	−2·2	−6·5	−4·3	−13·5	−1·9	−7·7
N70	−44·4	−29·7	−37·1	−0·6	2·2	0·8	−2·8	−0·7	−1·8
Mean	−8·9	8·9	—	−0·9	0·9	—	−1·7	1·7	—

Analysis of variance for environmental scores

		Component		
		1	2	3
Source of variation	d.f.	Mean square	Mean square	Mean square
Sites × Years (S × Y)	16	3103·7***	126·2***	36·4
Error 1	17	12·8	21·4	31·8
Densities (D)	1	5418·4***	53·1	205·4**
D × S × Y	16	104·2***	16·8	15·0
Error 2	17	3·8	16·3	13·5

represents the main environmental effect. The ranking of the corresponding genotypic latent vectors, the u_i values, is that of the mean yields of the varieties.

The ranking of the environmental scores for the second component is not related to the mean yields, nor to any obvious pattern of sites and years. Instead, certain site/year combinations differ considerably in their environmental scores, with Burscough 1968 (M68) being at the opposite extreme from Retford 1969 (N69) and Wellesbourne 1970 (W70). The corresponding

genotypic vectors show a sharp contrast between the Chantenays and Autumn Kings, and also, within the Autumn King group, a difference between Early Giant and Red Giant on the one hand and Autumn King and Flakkee French strain on the other.

The third component is much the least important, but the scores for the two densities are significantly different. The rankings of the genotypic vectors are quite different from those for both the first and second components, with Early Giant and Red Giant at opposite ends of the scale.

With data of this kind, it is possible that some of the effects may be due to small differences between the densities actually achieved for each variety. To investigate this, an analysis of covariance adjusting yields for densities achieved was carried out, and a principal component analysis then done on the adjusted yields. It was found that this principal component analysis gave results very similar to those from the unadjusted data.

Principal component analysis within environments

A principal component analysis within environments is shown in table 5. The first two components only are presented, since, as mentioned above, the

TABLE 5

Principal component analysis within environments

Environmental latent vectors:

	Component			
	1		2	
Site/year	D1	D2	D1	D2
W68	−0·157	−0·192	0·108	0·347
S68	−0·082	0·043	−0·128	0·267
R68	−0·279	−0·111	−0·088	0·095
C68	−0·035	−0·031	−0·099	−0·009
M68	−0·461	−0·572	−0·116	0·065
W69	−0·127	0·057	0·140	0·182
S69	−0·024	−0·013	0·069	0·149
R69	−0·056	−0·032	0·002	0·010
C69	−0·200	−0·122	0·152	0·191
M69	−0·035	−0·102	−0·065	0·194
N69	0·059	0·113	0·013	0·259
W70	0·045	0·051	0·110	0·063
S70	−0·042	−0·247	−0·123	0·151
R70	−0·008	−0·077	−0·155	0·054
C70	0·006	−0·020	0·021	0·013
M70	−0·092	−0·322	−0·611	−0·132
N70	−0·071	−0·017	−0·125	0·001

Analysis of variance on genotypic scores:

Genotypic scores:

	Components			Source of		Component	
						1	2
Variety	1	2		variation	d.f.	Mean square	Mean square
NMRC	10·2	−2·7		Ch v AK	1	1855·2**	137·4
RCh	7·4	3·6		Within Ch	3	20·6	144·3
LCh	10·4	14·7		Within AK	3	517·1*	203·8*
ChRC	15·1	−3·9		Error	8	79·0	46·7
EG	−27·1	11·3					
AuK	−0·8	−10·7					
FFS	6·4	−3·0					
RG	−21·5	−9·3					

third, though just significant at the 0·05 level, does not appear biologically meaningful. The analysis of variance of the genotypic scores for the first component shows differences both between types and within the Autumn King group, where Early Giant and Red Giant differ from Flakkee French strain and Autumn King. The corresponding environmental vectors show differences in ranking between the M68 environment at one extreme and N69 and W70 at the other. The second component shows differences within Autumn King only, with EG significantly different from RG and AuK. The corresponding environmental vectors show a difference between the two densities. Analyses on yields adjusted for differences in densities achieved again give broadly similar results.

Principal component analysis on residuals

A principal component analysis is shown in table 6 on residuals after the main effects of genotype and environment have been removed. As seen from table 3, only two principal components need to be considered. It makes no difference whether environmental scores and genotypic vectors are considered or the other way round; the u_{qi} and v_{qj} values are the same either

TABLE 6

Principal component analysis on residuals

Genotypic latent vectors:

	Component	
	1	2
NMRC	0·243	−0·131
RCh	0·284	0·038
LCh	0·331	0·580
ChRC	0·281	−0·201
EG	−0·563	0·562
AuK	−0·062	−0·423
FFS	0·076	−0·111
RG	−0·589	−0·313

Environmental scores:

	Component			
	1		2	
Site/year	D1	D2	D1	D2
W68	−2·8	−4·3	2·2	10·3
S68	−0·2	3·5	−2·9	3·7
R68	−9·5	0·5	−0·9	4·5
C68	1·5	3·2	−1·6	−1·2
M68	−16·4	−17·5	−1·4	1·9
W69	−1·5	8·2	2·2	3·5
S69	3·5	3·8	−0·3	2·6
R69	0·5	2·3	0·2	−0·3
C69	−4·8	−1·2	2·9	3·5
M69	0·8	0·4	−5·4	3·9
N69	7·4	9·4	−1·9	3·7
W70	7·0	6·7	1·2	0·7
S70	0·6	−4·6	−6·4	3·9
R70	3·1	1·0	−7·0	1·0
C70	3·9	2·6	−1·0	−1·2
M70	−3·2	−8·9	−13·8	−1·2
N70	1·5	3·4	−4·6	−0·8

way.   No analyses of variance are possible on the $\theta_q u_{qi}$ or $\theta_q v_{qj}$ scores, but they can be ranked for each principal component.   For the first component all the Chantenay varieties are at one end of the scale and the Autumn Kings at the other, the ranking being LCh, RCh, ChRC, NMRC, FFS, AuK, EG, RG; the corresponding ranking of environments gives, as before, M68 at one extreme and N69 and W70 at the other.   For the second component the ranking of the varieties is LCh, EG, RCh, FFS, NMRC, ChRC, RG, AuK, and the ranking on the environments suggests a consistent difference between the two densities.

Principal component analyses of the Chantenay and Autumn King groups separately

It is reasonable to argue that where differing groups of genotypes respond differently to environments it is appropriate to carry out analyses within groups, keeping the data for each group entirely separate.   These analyses may give information additional to that already obtained from the analyses on all groups together, and so assist in its interpretation.   Here, principal component analyses were done within each of the Chantenay and Autumn King groups separately, both within genotypes and within environments. The results are not shown in detail, for reasons of space, but show the following effects, the first of which is that, for all analyses, two components only are significant.   For the within environments analysis, no significant variety effects are found within either group, though there is a suggestion in the Autumn King group that, for the first component, Early Giant and Red Giant differ from the other two varieties.   Within genotypes, the first components for both Chantenay and Autumn King groups show highly significant effects of site/year combinations and densities, and an interaction between them; they thus represent the main effects of environments, in a similar manner to the first component within all genotypes together.   The second components show density effects, highly significant within Chantenay and at the 0·10 level of significance within Autumn King, thus resembling the third component for the combined analysis.   As stated above, the third component is not significant for either separate analysis; however, it is possibly meaningful for the Autumn King group, where analysis of variance shows a suggestion of an environmental effect similar to the second component in the combined analysis.

6. DISCUSSION

Principal component analysis is a descriptive not an explanatory technique and it could be said that its application to the partitioning of treatment effects only reveals what can be shown by an analysis of variance.   This is so, if one knows what one is looking for.   Thus it is now " obvious " that the most important component of the interaction between genotypes and environments in this trial is reflected in the ranking of the site/year combinations for the differences between the Chantenay and Autumn King groups. This was not evident previously, even though the corresponding term in the analysis of variance in table 1 suggested that the component would be large. In general, principal component analysis will give information not readily available from analysis of variance, though neither analysis can replace the other.   The special benefit of the rankings obtained from principal component analysis is to show which particular genotypes and environments contribute to the important components of the interaction.

Using principal component analysis it is possible to partition the variation within genotypes or within environments, or the residual variation when both main effects have been taken out. These three possible partitions need not all give the same results; in the present example the most important component of variation within genotypes is that between environmental means, but the most important component within environments is not that between genotypic means but a component representing genotype-environment interaction. Nevertheless, apart from the component within genotypes representing variation between environments, all analyses, on all the varieties together or on the carrot types separately, pick out the same components as contributing to the overall variation within genotypes and environments.

In an experiment of the present kind, with classified sets of genotypes and environments with replication, the important principal components may be further examined by an analysis of variance of the scores, to see which particular genotypes and environments are responsible for the interactions, as already noted. When, as here, the environments used are a factorial set the analysis must be in terms of the particular factors imposed in the experiment rather than underlying causes. There are independent interactions of varieties with both density and site/year combinations, and the principal component analysis gives rankings for both genotypes and environments for each interaction separately.

When the first component of variation within genotypes is that between environments, the second and subsequent components represent the genotype-environment interaction. If this interaction has only one component the technique of joint regression analysis, though suffering from formal statistical difficulties, can give a useful guide to the biological interpretation of the results. However, as the carrot example shows, more than one statistically independent component may be needed to describe the interaction, and the value of joint regression analysis is then limited.

The ranking of environments may lead the breeder to an effective choice of environment to use for discriminating between genotypes in respect of particular sensitivities. In the example here, both densities would be used in commercial practice and the interaction of genotypes with density shows that it is not sufficient to grow breeding lines at one density only. Although a ranking of the site/year combinations is available, this does not lead directly to a satisfactory choice of environment for testing breeding lines, since it would not be efficient to test many breeding lines at various sites over a number of years. From a consideration of the contrasting site/year combinations the crucial factor is thought to be length of growing season, though this cannot be confirmed from available meteorological data. Thus, a further series of trials with different imposed lengths of growing season at one site has now begun (Dowker, 1971). This is particularly important since in the breeding programme it is intended to make crosses between the Chantenay and Autumn King types and breed for improved yields by selection from later generations of these crosses.

7. REFERENCES

DOWKER, B. D. 1971. Variation studies in carrots as an aid to breeding. I. Concepts. *J. hort. Sci., 46*, 485-497.

EBERHART, S. A., AND RUSSELL, W. A.   1966.   Stability parameters for comparing varieties. *Crop Sci.*, *6*, 36-40.

FINLAY, K. W., AND WILKINSON, G. N.   1963.   The analysis of adaptation in a plant breeding programme.   *Aust. J. agric. Res.*, *14*, 742-754.

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.

KNIGHT, R.   1970.   The measurement and interpretation of genotype-environment interactions.   *Euphytica*, *19*, 225-235.

LAWLEY, D. N.   1956.   Tests of significance for the latent roots of covariance and correlation matrices.   *Biometrika*, *43*, 128-136.

MANDEL, J.   1971.   A new analysis of variance model for non-additive data.   *Technometrics*, *13*, 1-18.

MATHER, K., AND JINKS, J. L.   1971.   *Biometrical Genetics*, 2nd edition.   Chapman and Hall, London.

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

TAI, G. C. C.   1971.   Genotypic stability analysis and its application to potato regional trials.   *Crop Sci.*, *11*, 184-190.

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.

WILLIAMS, E. J.   1952.   The interpretation of interaction in factorial experiments.   *Biometrika*, *39*, 65-81.

YATES, F., AND COCHRAN, W. G.   1938.   The analysis of groups of experiments.   *J. agric. Sci.*, *28*, 556-580.