Theor Appl Genet (1992) 85:89–100

Interpreting genotype-by-environment interaction using redundancy analysis

F. A. van Eeuwijk

DLO-Center for Plant Breeding and Reproduction Research (CPRO-DLO), P.O. Box 16, 6700 AA Wageningen, The Netherlands

Received January 3, 1992; Accepted March 24, 1992
Communicated by A. R. Hallauer

Summary. Methods for the interpretation of genotype-by-environment interaction in the presense of explicitly measured environmental variables can be divided into two groups. Firstly, methods that extract environmental characterizations from the data itself, which are subsequently related to measured environmental variables, e.g., regression on the mean or singular value decomposition of the matrix of residuals from additivity, followed by correlation, or regression, methods. Secondly, methods that incorporate measured environmental variables directly into the model, e.g., multiple regression of individual genotypical responses on environmental variables, or factorial regression in which a genotype-by-environment matrix is modelled in terms of concomitant variables for the environmental factor. In this paper a redundancy analysis is presented, which can be derived from the singular-value decomposition of the residuals from additivity by imposing the restriction on the environmental scores of having to be linear combinations of environmental variables. At the same time, redundancy analysis is derivable from factorial regression by rotation of the axes in the space spanned by the fitted values of the factorial regression, followed by a reduction of dimensionality through discarding the least explanatory axes. Redundancy analysis is a member of the second group of methods, and can be an important tool in the interpretation of genotype-by-environment interaction, especially with reference to concomitant environmental information. A theoretical treatise of the method is given, followed by a practical example in which the results of the method are compared to the results of the other methods mentioned.

Key words: Genotype-by-environment interaction – Factorial regression – AMMI analysis – Multiple regression – Redundancy analysis – Lettuce

Introduction

In plant breeding, genotype-by-environment interaction typically refers to non-additivity in two-way tables of genotypes by environments. The data consist of evaluations of genotypes collected in a number of environments. The environments usually are made up of combinations of years and locations, but they may also involve different treatments. Environments may be characterized by a number of variables, e.g., soil, climatological, and treatment variables.

The classical approach of Yates and Cochran (1938), revived by Finlay and Wilkinson (1963), uses the following model for an observation on genotype j ($j = 1, 2, \ldots, m$) in environment i ($i = 1, 2, \ldots, n$)

$$Y_{ij} = \mu + G_j + B_j E_i + \varepsilon_{ij}, \quad (1)$$

in which μ stands for the overall mean, G_j for the genotypical main effect, B_j for the slope of the linear regression of the response of genotype j on the environmental main effect E_i, and ε_{ij} is an error term. B_j is often interpreted as some kind of genotypical stability or sensitivity to the complex of environmental variables embodied in the environmental main effect E_i. Hence the environment effectively is reduced to the mean performance of the genotypes in that environment, and the genotype-by-environment interaction is subsequently described as the heterogeneity of the slopes of the regressions of the individual genotypical responses on this mean. For obvious reasons this model is referred to as the regression-on-the mean model. The environment is modelled in terms of the observations of the matrix; no use is made of explicitly measured environmental variables.

Regression-on-the mean provides modelling opportunities for interactions describable in one dimension. For modelling higher dimensional interaction, recourse can

be taken to an extension of the form

$$Y_{ij} = \mu + G_j + E_i + \sum_{l=1}^{L} U_{lj} V_{li} + \varepsilon_{ij}. \quad (2)$$

Model (2) is partially additive, partially multiplicative, and was first introduced by Gollob (1968) and Mandel (1969, 1971). In the multiplicative part, the U_{lj}s denote genotypical scores (sensitivities, stabilities) and the V_{li}s environmental scores (characterizations, indices). L indicates the number of multiplicative terms required for an adequate description of the interaction. Least-squares fitting of this model can be done in two stages. First, the additive terms are fitted in the usual way, then the remaining residual matrix is decomposed using the singular-value decomposition (Gabriel 1978). This model was already used in the context of plant breeding in the early 70s (Perkins 1972; Freeman and Dowker 1973). Recently it received renewed interest through Gauch and Zobel, who also introduced the term AMMI model, a shorthand for Additive Main effects and Multiplicative Interaction effects model (Gauch 1988; Gauch and Zobel 1988, 1989, 1990; Zobel et al. 1988). Model (2) certainly provides more modelling opportunities than model (1), but still defines the environment by quantities derived from the phenotypical observations themselves. These environmental characterizations may afterwards be related to explicitly measured environmental variables, that is *indirectly*, e.g., by regression or correlation.

Easy ways of *directly* relating genotype-by-environment interaction to environmental variables are: (a) regressing residuals from additivity on environmental variables for each genotype separately, or (b) using concomitant information on the environmental factor in the two-way ANOVA of genotypes by environments (Snedecor and Cochran 1980). The second case is a form of simultaneous regression, and will be referred to as factorial regression (Denis 1988). It amounts to regressing ANOVA interaction parameters on environmental variables. For an early example see Abou-El-Fittouh et al. (1969). An elaboration of factorial regression, but originally arrived at via a generalization of the AMMI approach, was obtained by Rao (1964). The method was dubbed principal components of instrumental variables. It can be understood as an AMMI model with a restriction on the environmental scores. These have to be linear combinations of measured environmental variables. Subsequently, the connection with multiple regression was established, e.g., by Hardwick and Wood (1972), Izenman (1975), Lefkovitch (1986), and Denis (1988). Hardwick and Wood probably were the first to note the applicability of the technique in a plant breeding context. So far Wood (1976) seems to be the only accessible application, though in rudimentary form. Finally, Van den Wollenberg (1977) developed the same method starting from canonical correlation analysis under the name of redundancy analysis.

Despite its apparent potential the technique has remained practically unknown in plant breeding. The present paper intends to stimulate interest in the method by describing the key features of the model together with an application to a real data set consisting of nitrate concentrations in lettuce.

Theory

Multivariate multiple regression

In order to describe the relationship between a set of genotypical responses and a number of environmental variables one could carry out multiple regressions for each of the genotypes on the set of explanatory environmental variables. Multiple regression aims at maximizing the multiple correlation coefficient; a measure of the association between a dependent variable and a set of independent variables. It can be shown that the multiple correlation coefficient is the maximum correlation between the dependent and a linear functions of the independents. The multiple regressions for a number of genotypical responses on the same set of environmental variables can be written in the form of a multivariate multiple regression model as follows

$$\mathbf{Y} = \mathbf{1}\mathbf{C}^t + \mathbf{X}\mathbf{M} + \mathbf{E} \quad (3)$$

in which the columns of the matrix $\mathbf{Y}_{n\times m}$ represent the genotypical responses, the columns of the matrix $\mathbf{X}_{n\times q}$ the environmental variables; $\mathbf{1}_{n\times 1}$ stands for a vector of ones, $\mathbf{C}_{m\times 1}$ for the m intercepts, $\mathbf{M}_{q\times m}$ for the matrix of regression coefficients, while $\mathbf{E}_{n\times m}$ stands for a matrix of independently distributed normal errors with zero expectation and variance σ^2. Inclusion of a term for the row main effect changes (3) into a factorial regression model (Denis 1988), which is more appropriate in the context of genotype-by-environment interaction. However, for ease of exposition below, (3) will be used as a reference model, generalizations to factorial regression being obvious. Model (3) will be called the full-rank regression model for reasons to be explained shortly. In the full-rank model each genotype possesses unique sensitivities to every one of the environmental variables, no inter-relatedness between genotypical responses exists.

The environmental information as collected by the researcher will generally not have the form that is most relevant to the plants. Environmental variables of importance to the plants can be approximated by linear combinations of measured variables (possibly transformed, and including squares and cross products). In addition, it seems reasonable to assume that different genotypes react to *similar* environmental factors, though with varying sensitivity. A model that describes genotype-by-environment interaction in terms of heterogeneity in genotypical sensitivity to *common* linear combinations of environmental variables is given by the redundancy analysis model (Rao 1964; Hardwick and Wood 1972; Izenman 1975; Van den Wollenberg 1977; Davies and Tso 1982). The supposition of common linear combinations of environmental variables as the basis of genotype-by-environment interaction marks the distinction between the redundancy analysis model and the multivariate multiple regression model. The common linear combinations are found by rotation of the axes in the space spanned by the fitted values of the full-rank regressions for the genotypes. The rotation step may be followed by a reduction step in which only the most explanatory linear combinations are retained.

Redundancy analysis

Instead of maximizing the correlations between the individual dependent variables and the set of independent variables, as in

multiple regression, in redundancy analysis linear combinations of independent variables are formed that account for successively maximal proportions of the total sum of squares over the set of dependent variables. The quantity of central importance is the index of redundancy, introduced by Stewart and Love (1968).

Let $\mathbf{Y}^t = (Y_1, \ldots, Y_m)$ and $\mathbf{X}^1 = (X_1, \ldots, X_q)$ be two sets of centered variables, and $\mathrm{SSSP}(\mathbf{Y}) = \mathbf{S}_{11}$, $\mathrm{SSSP}(\mathbf{Y}, \mathbf{X}) = \mathbf{S}_{12}$, $\mathrm{SSSP}(\mathbf{X}, \mathbf{Y}) = \mathbf{S}_{21}$, and $\mathrm{SSSP}(\mathbf{X}) = \mathbf{S}_{22}$, with SSSP a sums of squares and sums of products matrix. The index of redundancy is defined as

$$R^2(\mathbf{Y}:\mathbf{X}) = \frac{\mathrm{trace}(\mathbf{S}_{12}\,\mathbf{S}_{22}^{-1}\,\mathbf{S}_{21})}{\mathrm{trace}(\mathbf{S}_{11})}, \tag{4}$$

being the proportion of the total sum of squares in the **Y**-set which is accounted for by the linear prediction of **Y** by **X**. The analogy with the squared multiple correlation coefficient from multiple regression is obvious.

The coefficient vector **b** for the linear combination of independent variables $\mathbf{b}^t\mathbf{X}$ that describes the maximum proportion of the total sum of squares in the set of dependent variables **Y** can be found by maximizing the following function of **b**

$$\xi(\mathbf{b}) = \mathbf{b}^t\,\mathbf{S}_{21}\,\mathbf{S}_{12}\,\mathbf{b} - \lambda(\mathbf{b}^t\,\mathbf{S}_{22}\,\mathbf{b} - 1) \tag{5}$$

(Van den Wollenberg 1977). For understanding (5) one should note that the sums of products between the dependent variables **Y** and the linear combination of independent variables $\mathbf{b}^t\mathbf{X}$, are given by $\mathrm{SSSP}(\mathbf{b}^t\mathbf{X}, \mathbf{Y}) = \mathbf{b}^t\,\mathbf{S}_{21}$, and the sum of the squares of these sums of products is simply $\mathbf{b}^t\,\mathbf{S}_{21}\,\mathbf{S}_{12}\,\mathbf{b}$. For convenience, and without loss of generality, the linear combinations are scaled to unit sum of squares, explaining the second term on the right in (5).

Differentiating (5) with respect to **b** and setting the result equal to zero leads, after some reshuffling, to the generalized eigenvalue problem

$$(\mathbf{S}_{21}\,\mathbf{S}_{12} - \lambda\,\mathbf{S}_{22})\,\mathbf{b} = 0. \tag{6}$$

The first eigenvector, **b**, contains the weights for the X-variables, which are called canonical coefficients. The first eigenvalue, λ, represents the amount of the total sum of squares in **Y** explained by the linear combination $\mathbf{b}^t\mathbf{X}$. This linear combination represents the first redundancy variate. Subsequent redundancy variates, uncorrelated with preceding ones, can be obtained from subsequent eigenvectors.

Inspection of (6) also reveals the inter-connectedness of redundancy analysis and principal components analysis. When the Y- and X-set are the same $\mathbf{S}_{12} = \mathbf{S}_{21} = \mathbf{S}_{22}$ and (6) reduces to the equation for the principal components problem.

In the terminology of the genotype-by-environment problem, theoretical environmental variables are formed that minimize the total residual sum of squares of the regressions of the genotypical responses on these linear combinations of environmental variables. Genotypes, now, can be characterized by their covariances with the newly formed theoretical environmental variables.

Reduced rank regression

An alternative derivation of the method of redundancy analysis, which displays more clearly its least squares properties, is given by Davies and Tso (1982). They subsumed redundancy analysis under the wider class of reduced-rank regression models. The basic assumption underlying these models is that the matrix of regression coefficients is a matrix of low rank, in any case of lower rank than the full-rank multivariate multiple regression coefficients matrix. Reduced rank regression models arise naturally in situations where a number of Y-variables are known to be inter-related, as for genotypical responses.

The reduced-rank equivalent of the full-rank regression model (3), assuming the number of environments n to be greater than the number of measured environmental variables q, is written as

$$\mathbf{Y} = \mathbf{1}\,\mathbf{C}^t + \mathbf{Z}\,\mathbf{A} + \mathbf{E}, \tag{7}$$

in which $\mathbf{Z}_{n\times s}$ contains $s \leq q$ redundancy variates, linear combinations of the original environmental variables, that is, $\mathbf{Z} = \mathbf{X}\,\mathbf{B}$, with $\mathbf{B}_{q\times s}$ a matrix whose columns contain the weights for the environmental variables in **X**, the canonical coefficients. The columns of $\mathbf{A}_{s\times m}$ are made up of the covariances of the m responses in **Y** with the redundancy variates in **Z**, they are comparable with the regression coefficients in the Finlay-Wilkinson model.

Effectively, the reduced-rank argument is carried through by a factorization $\mathbf{M} = \mathbf{B}\,\mathbf{A}$ in (3). When **M** has rank $s = q$, model (7) represents the full-rank model (3), whereas for $s < q$ (7) denotes a reduced-rank model. The factorization can be found following a least squares argument (Davies and Tso 1982).

Methods

Assessing rank; maximum likelihood

A major issue arising in the application of redundancy analysis concerns the determination of the maximum rank s. It is appealing to base this decision on the residual sum of squares from the rank s fit

$$\mathrm{SS}_{\mathrm{res\,(s)}} = \|\,\mathbf{Y} - \hat{\mathbf{Y}}\,\|^2 + \sum_{i=s+1}^{\mathrm{rank}(\hat{Y})} \lambda_i, \tag{8}$$

with $\|\,\mathbf{D}\,\|^2 = \sum_{ij} d_{ij}^2$ for a matrix **D** with elements d_{ij}, $\hat{\mathbf{Y}}$ the matrix of fitted values from the full-rank regression, and λ_i the i-th eigenvalue from (6) (which is equivalent to the i-th eigenvalue of $\hat{\mathbf{Y}}^t\hat{\mathbf{Y}}$ or $\hat{\mathbf{Y}}\hat{\mathbf{Y}}^t$). $\mathrm{SS}_{\mathrm{res\,(s)}}$ consists of the ordinary residual sum of squares from the full-rank fit plus a contribution of the least significant eigenvalues of (6).

Assuming the errors making up the matrix **E** in (7) to be distributed independently normal, with zero mean and variance σ^2, the loglikelihood can be written as

$$\mathrm{loglik} = -\tfrac{1}{2}\,\mathrm{n\,m}\,[\log_e(2\,\pi\,\sigma^2) + 1], \tag{9}$$

with $\log_e$ denoting the natural logarithm. From (9) the maximum likelihood estimator for σ^2 is obtained as $\hat{\sigma}^2 = \mathrm{tr}(\hat{\mathbf{E}}^t\,\hat{\mathbf{E}})/\mathrm{nm}$, with $\hat{\mathbf{E}}$ containing the residuals from the rank s fit (Van der Leeden 1990). The loglikelihood ratio test for the hypothesis of rank t against $t-1$ is most conveniently written as

$$\mathrm{lr} = \mathrm{nm}\log_e\left[\frac{\mathrm{SS}_{\mathrm{res\,(t-1)}}}{\mathrm{SS}_{\mathrm{res\,(t)}}}\right], \tag{10}$$

with $\mathrm{SS}_{\mathrm{res\,(t-1)}}$ and $\mathrm{SS}_{\mathrm{res\,(t)}}$ the residual sums of squares fom the rank $s-1$ and rank s fit. Asymptotically lr in (10) has a χ^2 distribution with a number of degrees of freedom equal to the difference between the degrees of freedom for the rank t model and the rank $t-1$ model. Assume that the data, **Y**, are corrected for the genotypical and environmental main effect, then the number of degrees of freedom for redundancy variates is equal to $q + (m-1) - (2\,t-1)$ for the t-th redundancy variate, where q stands for the rank of the **X** matrix ($n - 1 > q$), and $(m-1)$ for the rank of the corrected $\mathbf{Y}_{n\times m}$ matrix ($n > m$).

Assessing rank; randomization test

As an alternative for the loglikelihood ratio test a randomization test can be used. A possible approach is based on permutation of the rows of the **X** matrix (Ter Braak 1988). Calculate the first eigenvalue, then permute the rows of **X** and recalculate the first eigenvalue, repeat this v times. The significance level for the first eigenvalue is $(u+1)/(v+1)$, where u is the number of eigenvalues of the permuted set greater than the eigenvalue for the unpermuted **X**. For testing the second eigenvalue, correct **Y** for the first axis, etc. When the errors are uncorrelated, the columns of **X**, the environmental variables, may be permuted independently.

Variable selection and model building

Selection of variables in redundancy analysis can be performed either by techniques akin to those in discriminant analysis or, alternatively, by techniques used for multiple regression problems. One possibility is a stepwise procedure within a factorial regression set-up, in which the usual ANOVA tests for contrasts can be used (Snedecor and Cochran 1980). Subsequently, the rank of the matrix of regression coefficients, and simultaneously the number of axes to retain, can be assessed by means of the test given in (10). This procedure was recommended for modelling a matrix in terms of concomitant variables for rows (and/or columns) by Gabriel and Odoroff (1985).

An alternative, using backward elimination, is inspired by an idea of Jolliffe (1986, p. 108) in the context of principal components. Discard variables with high absolute coefficients in redundancy variates which express exact or nearly exact linear relationships between the explanatory variables, i.e., with zero or near-zero eigenvalues. This can be done iteratively. Fit the full-rank model, test whether or not the last redundancy variate contains significant information, e.g., by (10), and if yes discard the variable with highest absolute coefficient. Repeat this for the now reduced set of explanatory variables until the last redundancy variate appears no longer non-significant. Note that we will end up with a full-rank model, but a reduced set of explanatory variates. Nevertheless, the rotation in the space spanned by the fitted values of the individual genotypes can add to the interpretation of the interaction.

A word of caution should be expressed with respect to too heavy reliance on statistical variable selection procedures. Especially for genotype-by-environment problems, a reasonable choice of variables expected to be most influential should be possible beforehand, thereby reducing the need for elaborate statistical selection procedures.

Goodness of fit for individual genotypical responses

Evaluation of individual fits to responses can be done by considering the reduced-rank regression as a method to derive best linear predictors, the redundancy variates, for the set of responses. The regressions of the responses on the s redundancy variates then can be treated in a univariate fashion, making use of univariate evaluation procedures. Mean square errors of fit can be compared with known levels of precision for the type of response. In addition, prediction error on an independent set can be a useful evaluation criterion.

Precision of estimates of canonical coefficients

In order to say something about the precision with which canonical coefficients are estimated, a result of Tyler (1982) can be used. This shows that the canonical coefficients corresponding to the i-th redundancy variate, $\mathbf{b}_i$, can be interpreted, if scaled appropriately, as the vector of regression coefficients for the regression of $\mathbf{a}_i'\mathbf{Y}$, a linear combination of the responses Y weighted by their covariances with the i-th redundancy variate, on the X-set. Using the normalization $\mathbf{a}_i'\mathbf{a}_i = 1$ the regression of $\mathbf{a}_i'\mathbf{Y}$ on **X** gives as regression coefficients $\sigma_i \mathbf{b}_i$. Standard errors and t-values from the regression may be used for exploratory purposes.

Visualization of results

An important aid in the interpretation of the results of eigenvalue techniques is the biplot (Gabriel 1971). For an exposition on the use of biplots in genotype-by-environment problems see Kempton (1984).

In case of the AMMI model it is customary to depict scores for genotypes and environments on the first two axes in two-dimensional biplots. A rank-two approximation of the matrix of interaction residuals can be found from the biplot using the inner-product definition. Imagine the scores for the genotypes and the environments to determine vectors in two-dimensional space. Then, the interaction effect of a certain genotype in a certain environment is approximated by the inner-product between their respective vectors. The inner-product between two vectors is simply the length of the orthogonal projection from one vector onto the other, multiplied by the length of the other. A factor -1 or 1 is used as a multiplication factor depending on the angle between the two vectors; -1 for obtuse angles, 1 for acute angles. Ranking of interaction effects for all the genotypes in a particular environment can easily be done by just considering the ordering of the orthogonal projections of the genotypical vectors on that environmental vector.

For redundancy analysis the story is about the same as for the AMMI analysis. The major difference is that for redundancy analysis it is not the matrix of interaction residuals, but the matrix of *fitted* interaction residuals, which forms the raw material. Biplots for redundancy analysis have as an additional feature the possibility of representing measured environmental variables. For details on this and related aspects see Ter Braak (1990).

Computation

The calculations for a redundancy analysis can be done by any package that includes facilities for the singular value decomposition of matrices (in which case the matrix of full-rank fitted values must be the input) or for solving generalized eigenvalue problems such as (6). The calculations for the Application section were programmed in Genstat (1987). The package CANOCO (Ter Braak 1988) includes redundancy analysis among a number of other multivariate techniques, all furnished with facilities for forward selection of variables and permutation tests.

Application: nitrate concentration in lettuce

Data

In the period between March 1987 and June 1988 eight lettuce (*Lactuca sativa* L.) genotypes (Table 1) were evaluated at 18 harvesting times (Table 2) with respect to their nitrate concentrations (Reinink 1991). Each evaluation consisted of an experiment in eight blocks. The 18 evaluations in time were treated as environments in which genotypical performances were assessed. The average nitrate concentrations (g/l) of the eight genotypes observed in the 18 environments are given in Table 3. After a pre-

liminary selection eight environmental variables thought to exert influence on nitrate concentration (Tables 4, 5) were chosen for a characterization of the circumstances. Their usefulness to describe the genotype-by-environment interaction was investigated.

Preliminaries

Before searching for an explanation in terms of environmental variables, the existence of interaction has first to be proven. This involves testing for interaction [see Krishnaiah and Yochmowitz (1980) for a review] and, when present, determining whether the interaction is not due to a few outliers or removable by transformation. Then various methods should be tried to relate environmental variables to the interaction. In what follows the results of the following methods will be used: (a) stepwise regression of residuals from additivity on the set of environmental variables for each genotype separately; (b) factorial regression on the environmental variables; (c) AMMI analyis; (d) redundancy analysis. Different methods will elucidate different aspects of the data. At the same time, however, certain main features should become evident, as if looked upon from different angles.

Testing interaction in the two-way analysis of variance set-up (Table 6), using the mean intra-block error as an estimate for the error gave a highly significant result, $P \ll 0.001$. Another estimate for the error can be obtained via principal components analysis of the matrix of interaction residuals, which is part of the AMMI analysis, using the non-significant eigenvalues. The eigenvalues ex-

Table 1. Lettuce genotypes (*Lactuca sativa* L.) and their abbreviations

Pa	Panvit
DM	Deci-Minor
Pi	Pinto
GT	Große Brune Têtue
RW	Reichenauer Winter
Wi	Winterbutterkop
Tr	Trocadero
Ls	*Lactuca sativa capitata*

Table 2. Trial numbers and harvesting times (day-month-year) of the trials corresponding to the environments 1 to 18

Trial	Harv. time	Trial	Harv. time	Trial	Harv. time
1	08-04-1987	7	25-11-1987	13	10-05-1988
2	06-05-1987	8	06-01-1988	14	18-05-1988
3	03-07-1987	9	19-02-1988	15	03-06-1988
4	10-09-1987	10	08-03-1988	16	14-06-1988
5	07-10-1987	11	30-03-1988	17	20-06-1988
6	05-11-1987	12	26-04-1988	18	30-06-1988

Table 4. Measured environmental variables in the environments of Table 2

Number	Variable
1	Electrical conductivity of the medium
2	Summed global radiation in Joule/cm^2/day on eighth last day before harvest
3	As 2 on fourth last day before harvest
4	As 2 on second last day before harvest
5	As 2 on last day before harvest
6	Daylength on sowing day in hours
7	As 6 on introduction NFT system
8	As 6 on harvesting day

Table 3. Mean nitrate concentrations (g/l) over the eight replicates of a randomized blocks design for the genotypes from Table 1 in the environments of Table 2

Environment	Genotype							
	Pa	DM	Pi	GT	RW	Wi	Tr	Ls
1	3.113	2.835	2.629	1.988	2.199	2.414	1.248	2.380
2	3.379	3.222	2.848	2.823	3.002	2.950	2.176	3.196
3	3.067	2.326	2.511	2.120	2.692	2.598	1.032	2.355
4	3.202	2.663	2.230	1.638	2.187	2.171	1.062	1.599
5	3.921	3.365	3.028	2.653	2.935	2.931	2.007	2.942
6	4.153	3.970	3.444	2.813	2.865	3.232	2.341	3.289
7	4.851	4.512	4.010	3.504	3.135	3.624	3.080	3.612
8	4.547	4.203	3.429	2.944	2.616	3.052	2.817	3.070
9	3.721	3.505	3.337	2.425	2.177	2.525	1.917	2.830
10	3.581	3.298	3.287	2.389	2.159	2.681	1.744	2.726
11	3.312	3.130	2.959	2.280	1.797	2.152	1.365	2.178
12	3.439	3.329	3.254	2.561	2.843	3.035	1.927	3.058
13	3.195	3.047	2.948	2.696	2.610	2.902	1.914	3.138
14	2.890	2.297	2.295	2.237	1.930	2.414	1.462	2.274
15	2.700	2.430	2.172	2.004	2.194	2.392	1.374	2.144
16	3.143	2.710	2.429	2.260	2.406	2.438	1.536	2.464
17	2.746	2.470	2.226	2.126	2.332	2.185	1.287	2.621
18	3.273	2.384	2.555	2.167	2.545	2.386	1.616	2.813

Table 5. Values of the environmental variables of Table 4 in the environments of Table 2

Environment	Environmental variable							
	1	2	3	4	5	6	7	8
1	2.1	1,136	993	911	881	14.75	11.23	13.15
2	2.1	1,345	1,277	1,250	1,815	11.78	13.28	14.93
3	2.1	1,700	2,191	2,586	2,556	16.14	16.48	16.37
4	2.1	1,076	1,090	1,323	1,065	14.81	13.61	12.74
5	2.2	960	779	539	457	13.61	12.46	10.89
6	2.0	316	482	421	556	12.81	11.23	9.04
7	2.0	145	117	102	42	11.91	10.29	8.05
8	1.9	109	93	127	42	10.96	8.71	7.73
9	2.2	555	504	415	383	9.64	8.50	9.90
10	2.0	641	663	596	780	8.45	7.98	11.78
11	2.0	676	666	541	546	7.70	9.04	12.60
12	1.6	1,951	2,427	2,413	2,286	9.22	12.05	14.39
13	1.5	1,651	1,789	1,276	1,518	11.78	13.01	15.21
14	1.5	2,281	2,359	2,376	2,514	13.15	14.45	15.62
15	1.5	1,244	1,456	1,604	1,398	14.07	15.37	16.23
16	2.3	1,398	1,852	2,719	2,975	14.51	15.96	16.45
17	1.5	2,041	1,515	1,350	988	14.93	16.23	16.48
18	1.5	1,326	1,416	1,779	1,580	15.26	16.39	16.41

Table 6. Two-way analysis of variance on the genotype-by-environment matrix of Table 3. The error is the mean intra block error over the 18 trials

Source	Df	SS	MS
Genotypes	7	31.333	4.476
Environments	17	29.325	1.725
Interaction	119	6.772	0.057
Error	882		0.009

Table 7. Sum of squares for interaction per genotype, selected explanatory set from a stepwise regression, percentage sum of squares explained, R^2, by the selected set, and by the first and second redundancy variate (linear combinations of variables 7 ad 8), and residual mean square from regression on first and second redundancy variate

Genotype	SS int.	MR-set	R^2			RMS-RA
			MR	RA1	RA2	
Pa	0.970	3 4 5 8	80	44	17	0.023
DM	1.297	6 8	81	84	0	0.013
Pi	0.674	4 7	61	35	30	0.015
GT	0.314	6 7	39	31	10	0.012
RW	1.744	1 3 5 7	85	75	3	0.024
Wi	0.440	2 3 4 6	75	49	0	0.014
Tr	0.495	1 2 3 5 8	67	9	5	0.027
Ls	0.840	2	31	27	4	0.036

plained respectively 61, 16, 11, 5, 4, 2, and 1% of the interaction sum of squares. Eigenvalues below 0.7 times the average percentage (i.e., $0.7 \times 100/7 = 10\%$) can be interpreted as noise (Jolliffe 1986). So the first three eigenvalues represent structure, the rest noise. Approximate degrees of freedom can be attributed using Mandel's (1971) simulation studies. Summing the last four eigenvalues and dividing by the appropriate degrees of freedom, 34.8, led to an error estimate of 0.023, again leading to a highly significant interaction. A reason for the difference between both estimates of error might be the extra contributions of environment-by-block, and genotype-by-environment-by-block interactions, to the estimate derived from the non-significant eigenvalues. As a corollary it can be remarked that the dimensionality of three for the interaction implied inappropriateness of the regression-on-the-mean model.

A check on outliers revealed no severe anomalies in the data. The maximum normed residual, the maximum absolute interaction residual divided by the square root of the interaction sum of squares (Stefansky 1972), amounted to only 0.21, which was far from significant. The estimate for the Box-Cox parameter for a power transformation (see Atkinson 1982) included the value 1 in its 95% confidence interval, so that there was no reason for a transformation either.

Multiple regression

The environmental variables from Table 5 were used as the explanatory set in stepwise regressions for the interaction residuals of the individual genotypes. The cut-off values were chosen as $F_{in} = F_{out} = 4$ (Montgomery and Peck 1982). The results are given in Table 7. Substantial parts of the interaction sums of squares can be described by the environmental variables. The problem, however, is that no pair of genotypes has the same set of explanatory variables. In fact all environmental variables end up three times in the eventual explanatory set, except variable 1

Table 8. Selected sets of variables for factorial regression, with order of variables within sets reflecting stepwise selection. Further columns; distribution of explained sums of squares over redundancy variates, and total sums of squares explained. All subset regressions and redundancy axes are significant at at least 5%, unless non-significancy (ns) is indicated

Variable (s)	RA-1	RA-2	RA-3	Total
7, 8	3.645	0.514		4.158
8, 6	3.593	0.522		4.117
7, 2	3.647	0.393		4.040
7, 3	3.586	0.378		3.965
2, 6	3.444	0.499		3.943
5, 6, 1	3.290	0.537	0.112 ns	3.938
3, 6	3.404	0.441		3.845
7, 6	3.430	0.413		3.844
4, 6, 1	3.139	0.495	0.159 ns	3.794
5, 6	3.062	0.348		3.441
7	3.361			3.361
8	3.360			3.360
4, 6	3.012	0.326 ns		3.338
2	3.019			3.019
3	2.846			2.846
6, 1	2.287	0.469		2.756
4	2.535			2.535
5	2.403			2.403
6	1.952			1.952
1	0.903			0.903

Table 9. Correlations between environmental variables and environmental scores from AMMI- and redundancy analysis (axes are linear combinations of variables 7 and 8)

Variable	AMMI-1	AMMI-2	AMMI-3	RA-1	RA-2
1	−0.40	−0.45	−0.11	−0.42	0.13
2	0.85	0.10	−0.12	0.84	−0.24
3	0.82	0.02	−0.17	0.83	−0.23
4	0.78	−0.13	−0.07	0.85	−0.12
5	0.76	−0.03	−0.07	0.80	−0.19
6	0.63	−0.45	0.36	0.66	0.59
7	0.89	−0.19	0.26	0.95	0.30
8	0.90	0.05	−0.12	0.95	−0.30
RA-1	0.94	−0.08	0.07	1.00	0.00
RA-2	−0.01	−0.40	0.63	0.00	1.00

(only two times) and variable 3 (four times). The multiple regression approach thus leads to a highly idiosyncratic description of the interaction.

Factorial regression

More parsimonious descriptions of the interaction residuals are possible with factorial regression. Just as in the case of separate multiple regressions, the interaction is related directly to environmental variables, but this is done simultaneously for all genotypes. Testing of the contributions of one or several variables can be done by means of usual F-tests. With the inclusion or exclusion of a variable, 7 degrees of freedom from the interaction are involved. Contributions were tested against the remainder of the interaction at 5%. The remainder might be tested against an independent estimate of the error, e.g., 0.023.

Table 8 gives the results of an all-subsets procedure. For pairs and trios, variables are given in order of inclusion following a stepwise procedure: for pairs starting from every one of the individual variables, which were all found significant at 5%, for trios starting from each of the pairs remaining after the elimination part of the preceding step. The pair consisting of variables 7 and 8 (daylengths at the introduction of the NFT system and at harvesting time) performs best with respect to the amount of the interaction sum of squares explained. However, the pair 8 and 6 (daylength at sowing date) does only slightly worse.

AMMI analysis

Part of the AMMI analysis was already presented above under Preliminaries (testing for interaction). To gain some insight into the meaning of the axes, the correlation of the environmental scores for the axes 1 to 3 with the environmental variables was calculated (Table 9). Only axis 1 shows a relationship with the environmental variables, especially with variables 7 and 8. For an easier understanding of the meaning of this result one can look at the biplots of axis 2 against 1, 3 against 1, and 3 against 2 (Fig. 1 a, b, c). The scaling is such that the score vectors for the environments have squared lengths equal to the eigenvalues, whereas the genotypes have squared lengths of 1. With this scaling in the biplot of axis 2 against 1 the squared distance between the environmental approximates to twice the amount of interaction between them (Kempton 1984).

AMMI axis 1, AMMI-1, can be seen in Fig. 1 to represent roughly a contrast between summer (environments 2, 3, 13, 15, 17, and 18, having high positive scores) and and winter (environments 7, 8, 9, and 11, having high negative scores). This conclusion is in accordance with the high positive correlations of AMMI-1 with daylength at introduction NFT, variable 7, and harvesting date, variable 8. Daylength is greater in summer than in winter. AMMI-2 is dominated by the environments 4 (highly positive) and 13 (highly negative). To say AMMI-2 represents a contrast between spring and autumn would be overinterpreting. Just as AMMI-2 is not very easily relatable to environmental circumstances, neither is AMMI-3.

Redundancy analysis

One way of starting the redundancy analysis is by investigating the possibilities for rank reduction of the matrices of regression coefficients of the factorial regressions. In Table 8 the distribution of the interaction sum of squares over the redundancy variates is given for each

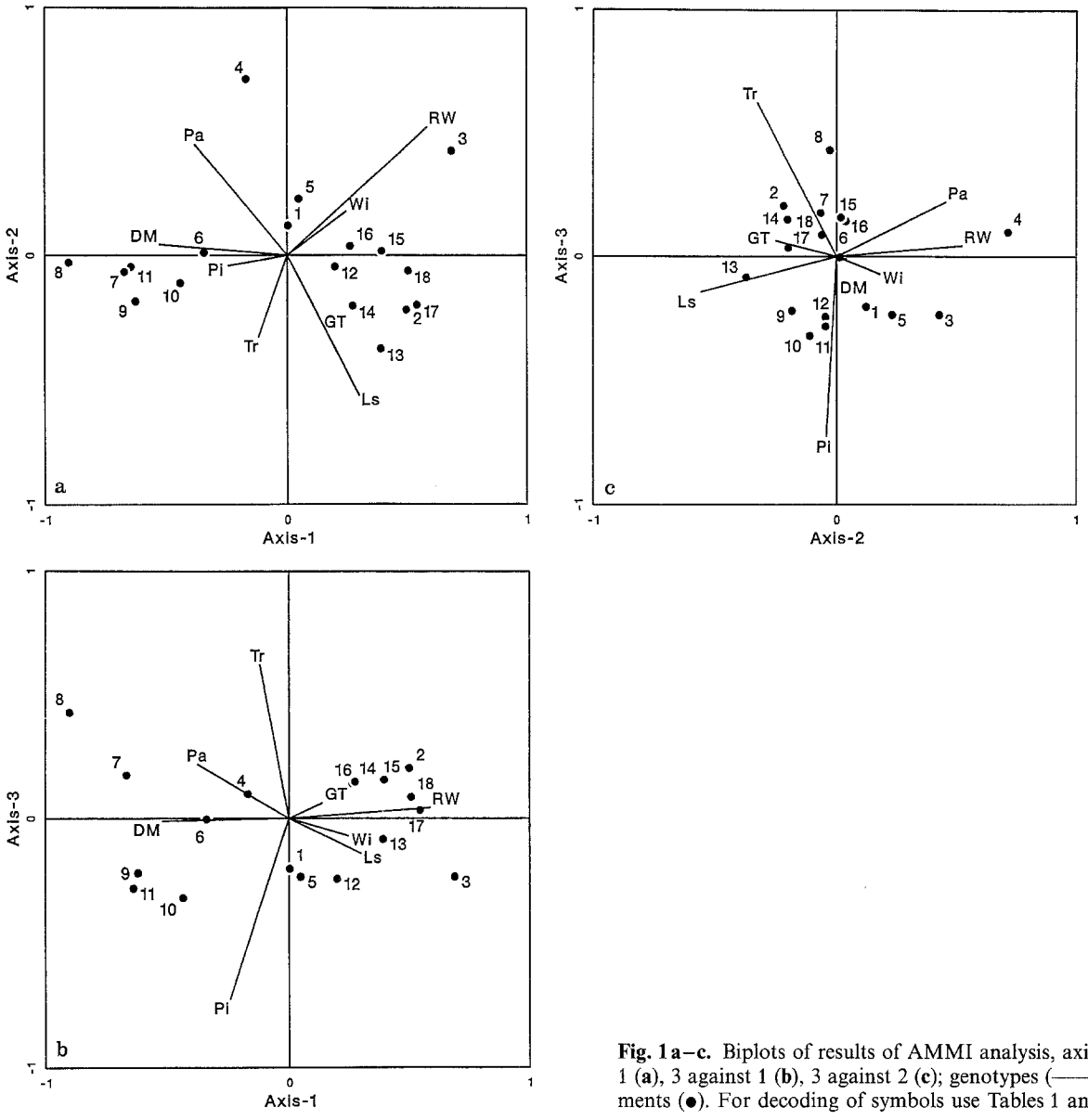

Fig. 1a–c. Biplots of results of AMMI analysis, axis 2 against 1 (**a**), 3 against 1 (**b**), 3 against 2 (**c**); genotypes (——), environments (●). For decoding of symbols use Tables 1 and 2

factorial regression. It seems natural to take the best set, pair 7 and 8. The test for a rank reduction using (10) reads: $\text{lr} = 18 \times 8 \times \log_e \left[\frac{6.772 - 3.645}{6.772 - (3.645 + 0.514)} \right] = 25.859.$ The quantity lr is, under the null-hypothesis, of no second dimension, asymptotically distributed as a χ^2 with $q + (m-1) - (2t-1) = 6$ degrees of freedom, $q = 2$, $m = 8$, $t = 2$ (see Methods section). This means that no rank reduction is possible, as the 5% point for $\chi^2_6 = 12.592$. The necessity for the full-rank model was confirmed by a permutation test for the second dimension, conditional on the first dimension, $P \leq 0.05$ (see Methods section). The coefficients for the (standardized) variables in the first redundancy variate were 0.13 for 7, and 0.13 also for 8, approximate t-values (see Methods section) were 3.36 and 3.37. Corresponding values of the second redundancy variate were 0.40 and −0.40, with t-values of 4.85 and −4.85. The coefficients were scaled in such a way that the sum of squares for the environmental scores was 1.

The first redundancy variate is the sum of the daylenghts at harvesting time and a month earlier, so high values will be found in summer and low values in winter (recall that X-variables were centered to mean zero), while intermediate values will be found in spring and autumn. The second redundancy variate is the difference between both daylength variables. During summer and winter daylength will not change very much, resulting in almost zero values for this redundancy variate. However, in spring and autumn daylength changes, and the second redundancy variate will become positive in autumn and

negative in spring. The two redundancy variates together thus describe a reaction of nitrate concentration to day-length throughout the year.

The biplot for the nitrate data (Fig. 2) immediately reveals that the genotype-by-environment interaction is a season-dependent phenomenon; the environments are arranged in a closed curve running counter-clockwise from summer at the right via autumn at the top, winter at the left, and spring at the bottom, to summer again at the right. Scaling is just as for the AMMI biplots; that is, environmental scores have sum of squares equal to the eigenvalue of the corresponding axis. The distance between environments is proportional to the amount of interaction between them. Most interaction can be identified between the extreme winter environments 7, 8 and 9 on the left, and the extreme summer environments 3, 15, 16 and 17 on the right.

The data set offers the opportunity for an internal check of the adequacy of the model because, for some dates, data are available from 1987 as well as 1988. To be more specific; environment 1 (8-4-87) may be expected to be located between 11 (30-3-88) and 12 (26-4-88), 2 (6-5-87) has to be in the neighbourhood of 13 (10-5-88) and 14 (8-5-88), and 3 (3-7-87) has to be near 18 (30-6-88). Inspection of Fig. 2 corroborates these expectations, thereby vindicating the chosen model.

Further evidence for the correctness of the redundancy solution is given by the position of the genotype RW in the biplot. This genotype was selected for its extremely low nitrate concentrations under low light conditions (Reinink et al. 1987). The genotype RW has above average nitrate concentrations in summer, so that highly positive inner-products results from the projection of summer points (3, 15, 16, 17, 18) on the RW vector, whereas RW has below average nitrate concentration in winter, and highly negative inner-products result from the projection of winter points (8, 9, 10) on the RW vector.

The cosine of the angle between the genotypical vectors may be interpreted as an estimate of the correlation between genotypical responses over environments. Genotypes RW and DM seem to behave as antipodes.

Information about the fits for the individual genotypical responses (in fact individual genotypical deviations from additivity) to the redundancy variates is given in the last three columns of Table 7. There it can be seen that the genotypes with the greater amounts of non-additivity, DM and RW, seem especially to determine the first redundancy component; that ist, their explained sums of squares are the highest. For the second component, genotypes Pa and Pi seem to be the most important. The proportion of variance explained by the regressions on both redundancy variates is a measure for the quality of the representation of the individual genotypes in the biplot. Genotypes DM and RW are well represented, genotypes Tr and Ls are poorly represented.

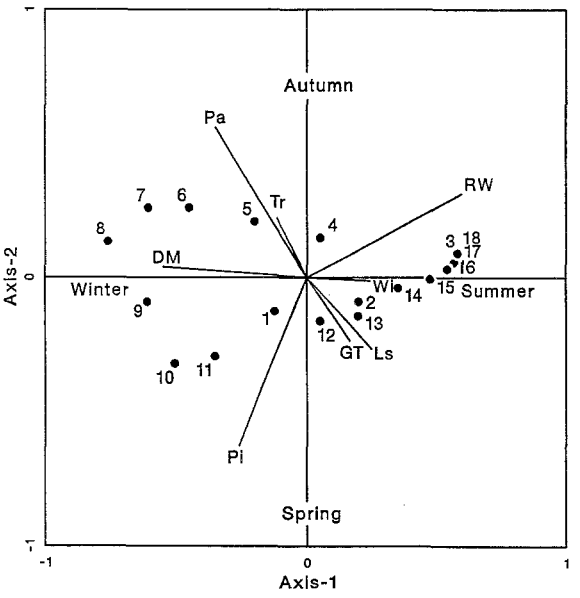

Fig. 2. Biplot of results of redundancy analysis; axis 2 against 1; genotypes (——), environments (•). For decoding of symbols use Tables 1 and 2

Table 10. Backward elimination of variables by discarding variable with highest coefficient on non-significant last redundancy variate

Variable	Coefficients						
1	0.02	0.02	−0.13	0.03	0.07	0.26	–
2	−0.44	−0.70	–	–	–	–	–
3	0.58	0.40	−0.02	−0.73	−0.45	–	–
4	−0.32	0.38	−0.19	0.77	–	–	–
5	0.60	–	–	–	–	–	–
6	−0.16	−0.30	0.46	0.18	0.08	−0.13	0.26
7	0.14	0.44	−0.60	–	–	–	–
8	−0.43	−0.33	0.44	−0.18	0.36	0.21	−0.19
Last eigenvalue	0.000	0.003	0.012	0.041	0.061	0.194	0.522
Sum eigenvalues	5.512	5.370	4.998	4.775	4.449	4.365	4.117

The residual mean squares for all genotypes except Ls are quite comparable, supporting the view that the redundancy analysis has taken up almost all structure from the data. The exception, Ls, has a higher residual mean square, probably due to interaction caused by factors other than the amount of light. The mean of the residual mean squares over the genotypes is 0.021, which is close to the 0.023 that was derived from the AMMI analysis.

An alternative to the above procedure is to start off from a full-rank model incorporating all environmental variables, and then test the significance of the last redundancy variate. Upon non-significance the variable with the highest coefficient is discarded (see Methods section). This process is repeated until the last redundancy variate

turns out to be significant. The results are given in Table 10. The test for the fourth redundancy variate for the model with the variables 1, 3, 6, and 8 reads $\mathrm{lr} = 18 \times 8 \times \log_e \left[\frac{6.772 - (4.459 - 0.061)}{6.772 - 4.459} \right] = 3.748$. Compared to the 5% value of χ^2_4, 9.488, this means non-significance. As variable 3 had the highest coefficient it was discarded. The test for the third redundancy variate for the model with the variables, 1, 6, and 8 reads $\mathrm{lr} = 18 \times 8 \times \log_e \left[\frac{6.772 - (4.365 - 0.194)}{6.772 - 4.365} \right] = 11.162$. The 5% value for χ^2_5 is 11.070. On this criterion the final set would be 1, 6, and 8. However the loglikelihood ratio test is slightly over-sensitive (see Discussion) and, therefore, it is better to continue until clearer significance for the last redundancy variate is found. After removing variable 1, a final set, 6 and 8 (daylength at sowing and at harvest time) is found for which both redundancy variates are clearly significant; lr for the second redundancy variate is 25.847 ($P < 0.001$). The interpretation is equivalent to the one arrived at earlier. The first redundancy variate is again a sum of both environmental variables, with most extreme values in summer and winter, and the second their difference, being extreme in spring and autumn. This is not surprising; in Table 8 it could already be seen that the pairs 7 and 8, and 8 and 6, explain the interaction almost equally well. Variables 6 and 7 have a correlation of 0.82, and should be exchangeable in combination with 8. In fact all pairs of variables selected in Table 8, except those including variable 1, would have led to the interpretation given above.

Discussion

Comparison of analyses

Various methods can lead to a very similar interpretation of the interaction. This important conclusion follows from the analyses in the Application section. In analysing genotype-by-environment tables one should use different approaches and, upon agreement, interpretation is straightforward, whereas upon disagreement closer inspection is necessary thereby acknowledging the differences between the method and the kind of structure they are supposed to detect.

For the nitrate data, AMMI and redundancy analysis gave comparable results, though the first extracts environmental scores as linear combinations of residuals from additivity, whereas the second forms environmental scores from linear combinations of measured environmental variables. The first AMMI axis paralleled the first redundancy axis, while the second and third AMMI axis more or less collapsed into the second redundancy axis (Table 9). The resemblance of AMMI and redundancy solutions means that for the redundancy analysis all relevant variables were selected (Ter Braak 1987). This is a useful diagnostic for the interpretation of interaction.

The individual regressions per genotype were mainly given as a reference point for the other analyses. Individual regressions have the advantage of high specificity, but the disadvantage of low parsimony. Moreover, it seems more likely that genotypes react to *common* environmental factors as can be uncovered by redundancy analysis. The dimension reduction property of the redundancy analysis was eventually not used for the redundancy analysis departing from factorial regression. Nevertheless, the interpretation of the interaction was certainly facilitated by the rotation, and the axes do bear on the physiology of the plants, as witnessed by the repeatability of the environmental scores in time. Besides, the fact that the positions of the genotypes in the redundancy biplot (Fig. 2) were scattered over all four quadrants means that the axes transcend a purely statistical interpretation, because in the latter case genotypes would be more likely to be situated near the lines $y = x$ and $y = -x$, since genotypes would bear no particular relationship to the extracted axes.

The dimension-reducing faculty of redundancy analysis proved very beneficial in the backward elimination procedure in the search for a good subset. However, strictly speaking, after final selection of variables 6 and 8, further rank reduction was not allowed. Real rank reduction can be seen to occur in Table 8 for the sets 5, 6, 1; 4, 6, 1; and 4, 6. For these sets the last redundancy variate turned out to be non-significant. Though one would not base an interpretation on these sets, since better ones are available, the estimation of the regression coefficients for these sets should be more accurate using the lower rank approximation of the matrix of regression coefficients due to the separation of structure in the retained dimension(s), and noise in the discarded dimension(s) (Gauch 1982).

In the Application section a slight over-sensitiveness of the loglikelihood ratio test was mentioned. This phenomenon is best illustrated by situations for which F-tests, as well as loglikelihood ratio tests, can be calculated. Consider the inclusion of variable 1 in the model after having fitted main effects. The loglikelihood ratio test is $\mathrm{lr} = 144 \times \log_e [6.772/(6.772 - 0.903)] = 20.608$ (see Table 8), to be compared with a χ^2_7 distribution, so $P = 0.004$. The F-test is $f = [0.903/7]/[6.772 - 0.903)/112] = 2.46$, to be compared with an $F_{[7;112]}$ distribution, giving $P = 0.022$. Somewhat less obvious is the following example. Take the pair 8, 6, and the trio 8, 6, 1. From Table 8 we know that 6 and 8 together explained a sums of squares of 4.117. Adding 1 raises this amount to 4.365 (see Table 10). An F-test for inclusion of 1 has the form $f_{[7;98]} = [(4.365 - 4.117)/7]/[(6.772 - 4.365)/98] = 1.44$, $P = 0.198$, so inclusion of variable 1 seems not to be supported by this F-test. On the other hand having found that both

redundancy variates are significant for the pair 8 and 6 (Table 8), a possible test for the need of the inclusion of 1 is to test the third redundancy variate for the trio 8, 6, 1. The loglikelihood ratio test here gives $P = 0.048$ (see Application section). A reason for the liberality of the loglikelihood ratio test could be that, though it is based on the comparison of two estimates for the residual variance, it does not take into account the different degrees of freedom on which the estimates are based. However, in general, F-test and loglikelihood ratio test do not deviate much, and it seems recommendable anyway, not to adhere too strictly to the results of significance testing. They are best used as rough guides.

Extensions and other applications of the redundancy analysis model

An appealing extension of redundancy analysis is the so called partial redundancy analysis, in which not only environmental variables, but also one or more covariables, are present (Davies and Tso 1982). To obtain the partial redundancy analysis solution the environmental variables are first regressed on the covariables, after which the residuals of these regressions replace the environmental variables in the subsequent redundancy analysis. In this way the contribution of particular environmental variables conditional on the contribution of other environmental variables is testable.

In the same vein, AMMI analysis and redundancy analysis can be combined. First, extract the significant redundancy variates; next, search for structure in the residuals by performing a singular value decomposition on them to see whether there is any structure left. Of course, covariables or conditioning can again be incorporated in this analysis.

Instead of interpreting the genotypical responses as variables and the environments as sample points, one could analyse the reversed situation of the genotypes within environments constituting variables and the genotypes over environments being sample points. Explanatory variables can then express either group structure in the genotypes or contrasts between them. A straightforward generalization of the redundancy analysis model even makes it possible to investigate both types of dependence simultaneously (Denis 1988; Velu 1991).

Another interesting application of redundancy analysis lies in the search for informative genotypes with respect to environmental circumstances, say indicator genotypes. Consider the model consisting of the genotypical main effect and the first dimension of the singular-value decomposition of the data corrected for the genotypical main effect. This model is almost equivalent to a regression on the mean model; the genotypical scores, in a reparameterized form, are estimates for the regression coefficients, and the environmental scores are estimates for the environmental main effects. Rewrite this model as a redundancy model by choosing as explanatory variables the (centered) genotypical responses themselves. When subsequently a subset selection procedure is applied to the explanatory genotypical responses a maximally adequate subset of informative genotypes will be retained. A similar approach is possible with respect to the environments.

Acknowledgements. I am indebted to Kees Reinink for allowing me to use the data set and for helpful discussions. I also thank Jean-Baptiste Denis, Aad van Eijnsbergen, John Gower, Hans Jansen, Piet Stam, Cajo ter Braak, Eeke van der Burg, and two anonymous referees for their comments on drafts of the manuscript. Paul Keizer was responsible for the graphics.

References

Abou-El-Fittouh HA, Rawlings JO, Miller PA (1969) Genotype by environment interactions in cotton – Their nature and related environmental variables. Crop Sci 9:377–381

Atkinson AC (1982) Regression diagnostics, transformations and constructed variables. J R Stat Soc Ser B 44:1–36

Davies PT, Tso M K-S (1982) Procedures for reduced-rank regression. Appl Stat 31:244–255

Denis JB (1988) Two way analysis using covariates. Statistics 19:123–132

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742–754

Freeman GH, Dowker BD (1973) The analysis of variation between and within genotypes and environments. Heredity, 30:97–109

Gabriel KR (1971) Biplot display of multivariate matrices with application to principal component analysis. Biometrika 58:453–467

Gabriel KR (1978) Least squares approximation of matrices by additive and multiplicative models. J R Stat Soc Ser B 40:186–196

Gabriel KR, Odoroff CL (1985) Some reflections on strategies of modelling: How, when and whether to use principal components. In: Sen PK (ed) Statistics in biomedical, public health and environmental sciences, Elsevier science publisher B.V., North Holland, pp 315–331

Gauch HG Jr (1982) Noise reduction by eigenvector ordinations. Ecology 63:1643–1649

Gauch HG Jr (1988) Model selection and validation for yield trials with interaction. Biometrics 44:705–715

Gauch HG Jr, Zobel RW (1988) Predictive and postdictive success of statistical analyses of yield trials. Theor Appl Genet 76:1–10

Gauch HG Jr, Zobel RW (1989) Accuracy and selection success in yield trial and analyses. Theor Appl Genet 77:473–481

Gauch HG Jr, Zobel RW (1990) Imputing missing yield trial data. Theor Appl Genet 79:753–761

Genstat 5 Committee (1987) Genstat 5 reference manual. Clarendon Press, Oxford

Gollob HF (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika 33:73–115

Hardwick RC, Wood JT (1972) Regression methods for studying genotype-environment interactions. Heredity 28:209–222

Izenman AJ (1975) Reduced-rank regression models for the multivariate linear model. J Mult Anal 5:248–264

Jolliffe IT (1986) Principal components analysis. Springer-Verlag, New York

Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. J Agric Sci 103:123–135

Krishnaiah PR, Yochmowitz MG (1980) Inference on the structure of interaction in two-way classification model. In: Krishnaiah PR (ed) Handbook of statistics, vol 1. North-Holland Publishing Company, Amsterdam, North Holland, pp 973–994

Lefkovitch LP (1986) Linear predictivity: An alternative for MANOVA and Multivariate multiple regression. Biom J 28:771–781

Mandel J (1969) The partitioning of interaction in analysis of variance. J Res Nat Bureau of Standards B Math Sci 73B:309–328

Mandel J (1971) A new analysis of variance model fo non-additive data. Technometrics 13:1–18

Montgomery DC, Peck EA (1982) Introduction to linear regression analysis. Wiley, New York

Perkins JM (1972) The principal components analysis of genotype-environmental interactions and physical measures of the environment. Heredity 29:51–70

Rao CR (1964) The use and interpretation of principal component analysis in applied research. Sankhya A26:329–358

Reinink K (1991) Genotype x environment interaction for nitrate concentration in lettuce. Plant Breed 107:39–49

Reinink K, Groenwold R, Bootsma A (1987) Genotypical differences in nitrate content in Lactuca sativa L. and related species and correlation with dry matter content. Euphytica 36:11–18

Snedecor GW, Cochran WG (1980) Statistical methods, 7th edn. Iowa State University Press, Ames.

Stefansky W (1972) Rejecting outliers in factorial designs. Technometrics 14:469–479

Stewart D, Love WA (1968) A general canonical correlation index. Psych Bull 70:160–163

Ter Braak CJF (1987) Ordination. In: Jongman RHG, Ter Braak CJF, Van Tongeren OFR (eds) Data analysis in community and landscape ecology. Pudoc, Wageningen, The Netherlands, pp 91–169

Ter Braak CJF (1988) CANOCO – a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1). Agricultural Mathematics Group, Wageningen, The Netherlands

Ter Braak CJF (1990) Interpreting canonical correlation analysis through biplots of structure correlations and weights. Psychometrika 55:519–531

Tyler DE (1982) On the optimality of the simultaneous redundancy transformations. Psychometrika 47:77–86

Van den Wollenberg AL (1977) Redundancy analysis an alternative canonical correlation analysis. Psychometrika 42:207–219

Van der Leeden R (1990) Reduced rank regression with structured residuals. DSWO Press, Leiden.

Velu RP (1991) Reduced rank models with two sets of regressors. Appl Statist 40:159–170

Wood JT (1976) The use of environmental variables in the interpretation of genotype-environment interaction. Heredity 37:1–7

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:556–580

Zobel RW, Wright MJ, Gauch HG (1988) Statistical analysis of a yield trial. Agron J 80:388–393