The simultaneous analysis of genotype by environment interaction for a number [of] traits using three-way multiplicative modelling

Eeuwijk, F.A. van; Kroonenberg, P.M.

BIULETYN OCENY ODMIAN
Zeszyt 26-27 1995

FRED A. van EEUWIJK*, PIETER M. KROONENBERG**
* DLO-Centre for Plant Breeding and Reproduction Research (CPRO-DLO), The Netherlands
** Department of Education, Leiden University, The Netherlands

THE SIMULTANEOUS ANALYSIS OF GENOTYPE BY ENVIRONMENT INTERACTION FOR A NUMBER TRAITS USING THREE-WAY MULTIPLICATIVE MODELLING

Summary: For the description of genotype by environment interaction in individual traits two-way multiplicative models have become a popular means of analysis. For the simultaneous analysis of genotype by environment interaction in a number of traits three-way multiplicative models are proposed. Theory for such models is reviewed and illustrated by an application to sugar beet resistance data.

Key words: Interaction; Multiplicative model; Plant breeding; Rhizomania; Singular value decomposition; Sugar beet; Three-way model; Two-way table.

1. INTRODUCTION

Over the last few years considerable progress has been accomplished in the development of more sophisticated methods for the analysis of genotype by environment interaction in plant breeding. These efforts were dedicated almost exclusively at the analysis of the genotype by environment interaction in individual traits. The phenotypic measurements made on plants, however, are known to be interdependent as they are the product of shared developmental processes. Separate analyses of individual phenotypic traits often produce sufficient information for adequate choices in breeding programmes, but enough research questions remain for which the present methods are unsatisfactory as they are unable to combine information over traits. For uncomplicated interactions ad hoc solutions may sometimes be acceptable. For example, when the interaction in a number of traits can be modelled by just one bilinear term, correlations between the genotypic (environmental) coefficients for the interaction in the various traits may reveal some indication of a common underlying structure or process (Paul, van Eeuwijk and Heijbroek, 1993). However, as soon as one of the traits has a more complex

interaction the approach by correlations becomes cumbersome. More systematic approaches are then needed. One such approach, based on three-way multiplicative models, will be presented in this paper. Two-way multiplicative models have become a standard tool for the analysis of genotype by environment interaction in individual traits (Gauch, 1992; van Eeuwijk, Denis and Kang, 1995). The generalization of two-way multiplicative models to three-way multiplicative models will be shown to be useful for the simultaneous modelling of genotype by environment interaction in more than one trait. The composition of the rest of the paper is as follows. First, two-way multiplicative modelling will be reviewed (section 2). Then a three-way multiplicative model will be introduced of which we think that is complies well with our demands (section 3). Subsequently, a few related three-way models are briefly described that may represent interesting alternatives (section 4). Practical three-way modelling is illustrated by an application to data from an international sugar beet research programme (section 5). The paper finishes with some conclusions (section 6).

2. TWO-WAY MULTIPLICATIVE MODELLING

For the analysis of genotype by environment interaction in individual phenotypic traits two-way multiplicative models provide a powerful class of models. Typically data are first arranged in the form of two-way genotype by environment tables of means. Subsequently, two-way models are fitted that include the additive main effects for genotypes and environments besides multiplicative terms for the interaction.

Let z_{ij} denote the non-additivity for the i-th genotype ($i=1,...,I$) in the j-th environment ($j=1,...,J$), i.e. the residual from a model that includes only main effects. Then multiplicative models for interaction state that z_{ij} can be written as (approximated by) a sum of multiplicative terms of the form $a_i b_j$ and an error term, e_{ij}, assumed to be independently and identically distributed (i.i.d.) with zero mean and constant variance. For example, $z_{ij}=a_{1i}b_{1j}+a_{2i}b_{2j}+e_{ij}$, where the model for interaction contains two multiplicative terms. The genotypic coefficients, a_i, can often be interpreted as sensitivities, the environmental coefficients, b_j, as some kind of environmental variable. Various types of multiplicative models can be distinguished, depending on whether genotypic and/or environmental coefficients represent values of measured variables or parameters that need to be estimated. In this section we restrict ourselves to multiplicative models for which both genotypic and environmental coefficients are estimated from the data themselves, without reference to explicit measurements. These models are called bilinear models, because upon fixation of the row parameters the models become linear models in the column parameters, whereas upon fixation of the column parameters they become linear in the row parameters.

The residuals from additivity, z_{ij}, can be collected in a two-way array, Z, that for convenience will be referred to as *matrix* Z. The matrix formulation of a multi-

plicative model for the non-additivity, $\mathbf{Z}$, reads $\mathbf{Z}=\mathbf{X}+\mathbf{E}$, where $\mathbf{X}$ denotes the multiplicative interaction and $\mathbf{E}$ the error. Multiplicative modelling of the interaction without the use of explicitly measured genotypic or environmental covariables is equivalent to finding a rank P approximation to the matrix $\mathbf{Z}$, where P is always smaller than the minimum of $I-1$ and $J-1$. This implies that $\mathbf{X}$ can be written as a product of two rank P matrices, $\mathbf{X}=\mathbf{AB}^T$.

The problem of finding a rank P approximation to $\mathbf{Z}$ can be translated to the search for a pair of rank P matrices $\mathbf{A}$ and $\mathbf{B}$ that minimize, in least square sense, the function $\phi_1(\mathbf{A},\mathbf{B})=\|\mathbf{Z}-\mathbf{AB}^T\|^2=\text{trace}\,((\mathbf{Z}-\mathbf{AB}^T)(\mathbf{Z}-\mathbf{AB}^T)^T)$. In general, no unique solution will be found unless additional identification constraints are imposed. Usually these constraints embody orthogonality and length constraints on the columns of $\mathbf{A}$ and $\mathbf{B}$. The singular value decomposition (SVD) of $\mathbf{Z}$ can be used to arrive at a unique solution for $\mathbf{A}$ and $\mathbf{B}$ (except for columnwise changes of sign) (Eckart and Young, 1936). Let the SVD of $\mathbf{Z}$ be $\mathbf{Z}=\mathbf{U}\Lambda\mathbf{V}^T$, with the columns of $\mathbf{U}$ and $\mathbf{V}$ consisting of the orthonormal eigenvectors of $\mathbf{ZZ}^T$ and $\mathbf{Z}^T\mathbf{Z}$ respectively, while Λ is a diagonal matrix with as diagonal elements the square roots of the eigenvalues of $\mathbf{Z}^T\mathbf{Z}$ (assumed to be all different). The best rank P approximation to $\mathbf{Z}$ is given by the product $\mathbf{AB}^T$ with $\mathbf{A}=\mathbf{U}_{(P)}\Lambda_{(P)}^c$ and $\mathbf{B}=\mathbf{V}_{(P)}\Lambda_{(P)}^{1-c}$ here the subscripts indicate that only first P columns of the pertinent matrices are retained, i.e. those corresponding to the P largest eigenvalues of $\mathbf{ZZ}^T$ and $\mathbf{Z}^T\mathbf{Z}$. The choice of the scaling constant c is not critical as long as $0\leqslant c\leqslant 1$.

The freedom that exists with respect to the scaling of the component matrices $\mathbf{A}$ and $\mathbf{B}$ can be used to include explicitly a matrix of scaling constants in the discrepancy function ϕ_1, resulting in $\phi_2(\mathbf{A},\mathbf{B},\mathbf{G})=\|\mathbf{Z}-\mathbf{AGB}^T\|^2$. Both the matrices $\mathbf{A}(I\times P)$ and $\mathbf{B}(J\times P)$ are assumed to be orthonormal and of rank P. $\mathbf{G}$ is a $P\times P$ matrix of scaling constants. The SVD of $\mathbf{Z}(=\mathbf{U}\Lambda\mathbf{V}^T)$ provides a direct solution. $\mathbf{A}$ and $\mathbf{B}$ can be obtained from the first P columns of $\mathbf{U}$ and $\mathbf{V}$, while $\mathbf{G}$ can be taken equal to the first P columns of Λ.

The SVD guarantees squareness and diagonality of $\mathbf{G}$. There is no room for generalizations of the SVD by allowing $\mathbf{G}$ to be non-diagonal and/or non-square. For two-way matrices the number of columns, or components, in $\mathbf{A}$ and $\mathbf{B}$ will necessarily be equal, while the elements of $\mathbf{G}$ represent the importance of a particular component.

Due to the close links that exist between the SVD of $\mathbf{Z}$ and the spectral decompositions of $\mathbf{ZZ}^T=\mathbf{U}\Lambda^2\mathbf{U}^T$ and $\mathbf{Z}^T\mathbf{Z}=\mathbf{V}\Lambda^2\mathbf{V}^T$, solutions for our minimization problems could alternatively have been obtained from these spectral decompositions, without having recourse to the SVD, if this had been desired.

3. THREE-WAY MULTIPLICATIVE MODELLING

For the simultaneous analysis of the genotype by environment interaction in a number of traits, the two-way genotype by environment tables of residuals from additivity data are first arranged in a three-way array (matrix). The classifying factors

for this three-way array are those of 'genotype' for the rows, of 'environment' for the columns, and of 'trait' for the layers. Let z_{ijk} represent the non-additivity for the i-th genotype in the j-th environment for the k-th trait ($k=1,...,K$). A multiplicative model for z_{ijk} consists in a structural part, x_{ijk}, containing a sum of multiplicative terms, and a random part, e_{ijk}, containing an i.i.d. error term with zero mean and constant variance. The terms of which x_{ijk} is comprised are each the product of a particular component coefficient for the genotype, the environment and the trait (ignoring for the moment a scaling constant). We will only be concerned with three-way multiplicative models for which neither genotypic nor environmental nor trait covariables were measured, i.e. a lower 'rank' approximation to the three-way array $\mathbf{Z}$ is wanted.

Two-way multiplicative models can be generalized in various ways to the three-way situation. We will focus on the proposal by Tucker (1966), the Tucker3 model. In the Tucker3 model each way is allowed to have a different number of components. This is possible because of the absence of a three-way equivalent of the two-way result 'row-rank=column rank'. Furthermore, every component in one way can combine with every one of the components in the other ways. In the elementwise formulation the Tucker3 model is written as

$$x_{ijk}=\sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}\,b_{jq}\,c_{kr}\,g_{pqr}+e_{ijk}.$$

Here a_{ip} represents the coefficient for genotype i in the genotypic component p ($p=1,...,P$), b_{jq} the coefficient for environment j in the environmental component q ($q=1,...,Q$), and c_{kr} the coefficient for trait k in the trait component r ($r=1,...,R$). The importance of the combination of the p-th genotypic component with the q-th environmental component and the r-th trait component for the description of z_{ijk} is expressed by g_{pqr}, a *core* element.

In matrix formulation the model for $\mathbf{Z}$ is $\mathbf{Z}=\mathbf{X}+\mathbf{E}$, with $\mathbf{Z}$, $\mathbf{X}$ and $\mathbf{E}$ three-way $I\times J\times K$ arrays. To represent the Tucker3 model in matrix form we need to introduce some notation. Denote by $\mathbf{Y}_{\langle 1\rangle}$ the $R\times ST$ two-way rearrangement of the $R\times S\times T$ three-way array $\mathbf{Y}$, i.e. the rows of the three-way array correspond to the rows of the two-way array, while the columns of the two-way array represent combinations of the columns and layers of the three-way array (Cartesian product). In the same way $\mathbf{Y}_{\langle 2\rangle}$ and $\mathbf{Y}_{\langle 3\rangle}$ represent $S\times TR$ and $T\times RS$ two-way rearrangements.

In the Tucker3 model the structural part of the model for $\mathbf{Z}_{\langle 1\rangle}$ is written as $\mathbf{X}_{\langle 1\rangle}=\mathbf{A}\mathbf{G}_{\langle 1\rangle}(\mathbf{C}^T\otimes\mathbf{B}^T)$, with $\mathbf{A}$ an $I\times P$ rank P genotypic component matrix, $\mathbf{B}$ a $J\times Q$ rank Q environmental component matrix, and $\mathbf{C}$ a $K\times R$ and R trait component matrix. $\mathbf{G}_{\langle 1\rangle}$ represents the $P\times(QR)$ rearrangement of the *core* matrix. Equivalently valid are the expressions $\mathbf{X}_{\langle 2\rangle}=\mathbf{B}\mathbf{G}_{\langle 2\rangle}(\mathbf{A}^T\otimes\mathbf{C}^T)$ and $\mathbf{X}_{\langle 3\rangle}=\mathbf{C}\mathbf{G}_{\langle 3\rangle}(\mathbf{B}^T\otimes\mathbf{A}^T)$.

For fixed ranks P, Q and R, estimates for the component matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, and the core matrix $\mathbf{G}$ are to be found from the minimization of the discrepancy function $\phi_3(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G})=\|\mathbf{Z}_{\langle 1\rangle}-\mathbf{A}\mathbf{G}_{\langle 1\rangle}(\mathbf{C}^T\otimes\mathbf{B}^T)\|^2$ with $\otimes$ the Kronecker product

(Kroonenberg and de Leeuw, 1980). The solution will not be unique as the component matrices can be multiplied by any non-singular matrix provided that the core matrix is multiplied by the inverse. This nonuniqueness can be used to impose orthonormality on the component matrices. As a consequence $\mathbf{G}_{\langle 1\rangle}\mathbf{G}_{\langle 1\rangle}^T$, $\mathbf{G}_{\langle 2\rangle}\mathbf{G}_{\langle 2\rangle}^T$, and $\mathbf{G}_{\langle 3\rangle}\mathbf{G}_{\langle 3\rangle}^T$, will be diagonal. Also $\mathbf{X}_{\langle 1\rangle}\mathbf{X}_{\langle 1\rangle}^T=\mathbf{A}\mathbf{G}_{\langle 1\rangle}\mathbf{G}_{\langle 1\rangle}^T\mathbf{A}^T$, $\mathbf{X}_{\langle 2\rangle}\mathbf{X}_{\langle 2\rangle}^T=\mathbf{B}\mathbf{G}_{\langle 2\rangle}\mathbf{G}_{\langle 2\rangle}^T\mathbf{B}^T$, $\mathbf{X}_{\langle 3\rangle}\mathbf{X}_{\langle 3\rangle}^T=\mathbf{C}\mathbf{G}_{\langle 3\rangle}\mathbf{G}_{\langle 3\rangle}^T\mathbf{C}^T$, So, within every one of its ways this three-way decomposition is equivalent to a two-way spectral decomposition (Weesie and van Houwelingen, 1983).

It is of some interest to know the number of degrees of freedom that accompany the model. This can be obtained by subtracting the number of constraints from the number of parameters to be estimated. The numbers of parameters to be estimated for an $I\times J\times K$ three-way array of K genotype by environment tables of non-additivity ($I\times J$ each) are $I\times P$ for **A**, $J\times Q$ for **B**, $K\times R$ for **C**, and $P\times Q\times R$ for **G**. Constraints amount to $P^2+Q^2+R^2$ for orthonormality and $P\times Q$ for the fact that the traits were corrected for genotypic and environmental main effects already.

Because of the chosen parametrization the squares of the core elements, g_{pqr}^2, are equal to the explained sums of squares for the combination of the p-th component in the first way, the q-th component in the second way, and the r-th component in the third way. Furthermore, an orthogonal decomposition of the total observed variation is possible; $z_{ijk}=\hat{z}_{ijk}+\hat{e}_{ijk}$, and

$$\sum_{ijk}\hat{e}_{ijk}^2=\sum_{ijk}(z_{ijk}-\hat{z}_{ijk})^2=\sum_{ijk}z_{ijk}^2-\sum_{ijk}\hat{z}_{ijk}^2,$$

where the summation can be over any number or combination of indices (ten Bere, de Leeuw and Kroonenberg, 1987).

The problem of finding a multiplicative model for the three-way array **Z** can be solved by an iterative algorithm (Kroonenberg, 1983, Chpt. 4). First, note that in the model $\mathbf{G}_{\langle 1\rangle}=\mathbf{A}^T\mathbf{Z}_{\langle 1\rangle}(\mathbf{C}\otimes\mathbf{B})$, so that **G** can be calculated from **A**, **B**, **C** and **Z**, after **A**, **B** and **C** have been estimated. Next, substitution of $\mathbf{G}_{\langle 1\rangle}=\mathbf{A}^T\mathbf{Z}_{\langle 1\rangle}(\mathbf{C}\otimes\mathbf{B})$ in ϕ_3 gives us an expression that contains only the component matrices as unknowns. The iterative algorithm then consists in first solving for **A** given **C** and **B**, then for **B** given **A** and **C**, and finally for **C** given **B** and **A**. Within each cycle of the iterative process estimates for **A**, **B** and **C** are obtained as the eigenvectors of respectively

$$[\mathbf{Z}_{\langle 1\rangle}(\hat{\mathbf{C}}\otimes\hat{\mathbf{B}})]\,[\mathbf{Z}_{\langle 1\rangle}(\hat{\mathbf{C}}\otimes\hat{\mathbf{B}})]^T$$

$$[\mathbf{Z}_{\langle 2\rangle}(\hat{\mathbf{A}}\otimes\hat{\mathbf{C}})]\,[\mathbf{Z}_{\langle 2\rangle}(\hat{\mathbf{A}}\otimes\hat{\mathbf{C}})]^T$$

$$[\mathbf{Z}_{\langle 3\rangle}(\hat{\mathbf{B}}\otimes\hat{\mathbf{A}})]\,[\mathbf{Z}_{\langle 3\rangle}(\hat{\mathbf{B}}\otimes\hat{\mathbf{A}})]^T.$$

Starting values can be obtained as the eigenvectors of $\mathbf{X}_{\langle 1\rangle}\mathbf{X}_{\langle 1\rangle}^T$ for **A**, of $\mathbf{X}_{\langle 2\rangle}\mathbf{X}_{\langle 2\rangle}^T$ for **B**, and of $\mathbf{X}_{\langle 3\rangle}\mathbf{X}_{\langle 3\rangle}^T$ for **C**. The process must be continued until a convergence criterion is met (change in residual sums of squares or values of the component scores). After convergence **G** can be calculated.

This procedure would in the two-way case immediately lead to the solution. Postmultiply **Z** by **V**, Vcontaining the eigenvectors of $\mathbf{Z}^T\mathbf{Z}$, then, because $\mathbf{Z}=\mathbf{U}\Lambda\mathbf{V}^T$,

$\mathbf{ZV}=\mathbf{U\Lambda}$ and $\mathbf{U}$ can be obtained from the eigenvectors of $\mathbf{ZVV}^T\mathbf{Z}^T=\mathbf{U\Lambda}^2\mathbf{U}^T$. No iteration is necessary.

Results from two-way decompositions are often presented in the form of biplots (Gabriel, 1971; Kempton, 1984), simultaneous displays of row and column component scores. Comparable plots can be made for three-way decompositions (Kroonenberg, 1983, Chpt. 6). These plots are called joint plots. The relations between the component scores of two ways are displayed conditional on the values of the component scores in the third way. The third way defines the (two-way) slices that are taken from the core matrix. Suppose we want to visualize the relations between the rows and columns of $\mathbf{Z}$. Let $\mathbf{G}_r$ be r-th core slice as defined by the r-th components of the third way. The slice $\mathbf{G}_r$ is now decomposed by a singular value decomposition to give $\mathbf{U\Lambda V}_r^T$. Then the original component matrices $\mathbf{A}$ and $\mathbf{B}$ are transformed to $\mathbf{A}_r^*=\mathbf{AU}_r\mathbf{\Lambda}_r^{0.5}$ and $\mathbf{B}_r^*=\mathbf{BV}_r\mathbf{\Lambda}_r^{0.5}$, and $\mathbf{A}_r^*\mathbf{B}_r^*=\mathbf{AG}_r\mathbf{B}^T=\mathbf{Z}_r$. Now the rows of $\mathbf{A}_r^*$ and $\mathbf{B}_r^*$ contain the coordinates for points that can represent the rows and columns of $\mathbf{Z}_r$ in a joint plot.

4. ALTERNATIVE THREE-WAY MODELS

Various other three-way multiplicative models have been proposed. The so-called Tucker2 model can be obtained from the Tucker3 model by replacing the components matrix $\mathbf{C}$ by an identity matrix (Kroonenberg and de Leeuw, 1980):

$$z_{ijk}=\sum_{p=1}^{P}\sum_{q=1}^{Q} a_{ip}b_{jq}g_{kpq}+e_{ijk}.$$

This model does not reduce the number of parameters over the third way of the data array, as the number of components in the third way is equal to the number of levels of the classifying factor for the third way. For every layer there is separate core slice. Errors are again assumed to be i.i.d. with zero mean and constant variance.

Another proposal follows from the Tucker3 model by setting $P=Q=R$ and requiring the core matrix to be diagonal ($P\times P\times P$);

$$z_{ijk}=\sum_{p=1}^{P} a_{ip}b_{jp}c_{kp}g_{ppp}+e_{ijk},$$

This model is best known under the name of Parafac (Parallel factor) model (Harshman, 1970; Harshman and Lundy, 1984). An earlier discussion of the use of various three-way models in agriculture can be found in Basford, Kroonenberg, and DeLacy (1991).

Yet another three-way model for the multivariete analysis of genotype by environment interaction was developed by Denis and Moro (1995). The expectation structure of their model can be obtained from the Tucker2 model by requiring the core matrix to be diagonal, and from the Parafac model by requiring the component matrices $\mathbf{A}$ and $\mathbf{B}$ to be orthonormal. The variance-covariance structure of their model is more general than the rather basic diagonal structure of the other models.

5. AN APPLICATION TO SUGAR BEET RESISTANCE BREEDING DATA

5.1 PROBLEM, DATA, AND TWO-WAY ANALYSES

A disease that causes great yield loss in sugar beet is rhizomania, caused by infection of the beets with the beet necrotic yellow vein virus (BNYVV). Transmission of the virus occurs through a soil-borne fungus that invades the root system of the plant. The persistent character of the fungus in the soil makes breeding for resistance a preferred way to control the disease. Breeding programmes aim at the development of partially resistant genotypes. A few years ago RNA analyses of virus samples revaled that probably more than one type of BNYVV should be distinguished. This would have important consequences for breeding programmes. By recombination of virus strains new types of BNYVV might might develop that would make already introduced resistance genes ineffective.

In 1992 an international study was started to investigate the evidence for different types of virus in the field. Twenty-five trials were laid out over Europe (Austria, France, Germany, Greece, Italy, The Netherlands and Spain); 15 trials on infested soil and 10 control trials on noninfested soils (O). For the infested fields the type of virus was assessed during the growing season (A or B). There were 9 trials infested with the A type virus, 4 with the B type, and of 2 infested trials the virus type was unclear (?). Six beet genotypes were included (Accord, Rizor, Stratos, Roxane, Monodoro and C48) thought to represent a rather broad spectrum of resistance characteristics (genes). Measured were three yield characterisitics; yield, percentage of sugar (%Sugar), and sugar yield (SugarY), and three quality characteristics; concentrations of potassium (K), sodium (Na) and alpha-amino nitrogen (αN).

For the assessment of differences between the A and B type virus infestations bilinear models were fitted to the 6×25 genotype by environment tables of means for each of the 6 traits separately. For all 6 traits two bilinear terms were found sufficient for an adequate description of the interaction. Inspection of biplots of interaction scores should reveal differences between the virus types. The biplots were searched for clear separation of A, B, and control trials (O). Although infested soils were generally clearly distinct from non-infested soils, none of the 6 biplots exhibited an unequivocal separation of A and B types.

5.2 THREE-WAY ANALYSIS

It was expected that a simultaneous analysis of the interaction in the six traits might be more powerful than the separate analyses. The genotype by environment tables of non-additivity were first standardized for each trait separately by dividing them by the square root of the uncorrected mean square, i.e. $\sqrt{\Sigma_{ij} z_{ij}^2 / IJ}$, to give the traits comparable total interaction. The standardized genotype by environment

tables then formed the layers in a three-way array classified by genotypes, environments and traits. Our aim was to fit a Tucker3 model and then first make plots of the environmental component scores to see whether A, B and O types would be separable at all. If yes, subsequently joint plots of genotypes and environments should be constructed in which the separation is present too to see whether the separation can be motivated.

Table 1

Analysis of variance table for the Tucker3 model with 2 genotypic, 3 environmental and 2 trait components.

	Degrees of Freedom	Sum of Sguares	Mean Square
Model	89	3.87	0,0435
Residual	631	2.13	0.0034
Total	720	6.00	

A Tucker3 model was selected with 2 components for the genotypes, 3 for the environments and 2 for the traits. Table 1 gives the analysis of variance for this model. Because the nonadditivity was standardized per trait, the total sum of squares amounted to 6 for the 6 traits together. The model degrees of freedom are obtained from the general formula given in section 3, with $I=6$ (# genotypes), $J=25$ (# environments), $K=6$ (# traits), $P=2$, $Q=3$, $R=2$, degrees of freedom $=[(6\times2+25\times3+6\times2)+(2\times3\times2)]-[(2^2+3^2+2^2)+(2+3)]=89$.

A motivation for the chosen model can be based on Table 2. The average proportion of nonadditivity described by separate fits of bilinear models with two interaction terms amounted to 86%. However, these models use 324 degrees

Table 2

Proportion of the total variation explained in individual traits and in all traits jointly (model fit) by various models.
6x 2w2 = for each trait individually bilinear model with two terms;
T3 232 = Tucker3 with 2 genotypic, 3 environmental and 2 trait components;
T3 222 = Tucker with 2 components for each way;
T3 333 = idem with 3 components;
T2 23- = Tucker2 with 2 genotypic and 3 environmental components;
PF 2 = Parafac with 2 components.

	Model					
Trait	6x 2w2	PF 2	T3 222	T3 232	T3 333	T2 23-
SugarY	.92	.90	.90	.92	.93	.93
Yield	.92	.86	.86	.89	.90	.90
%Sugar	.91	.64	.64	.66	.71	.70
aN	.84	.55	.55	.66	.72	.65
Na	.80	.31	.31	.51	.61	.63
K	.78	.22	.22	.23	.25	.25
Overall	.86	.58	.58	.64	.69	.68

Table 3
Core matrix (before rotation) in $G_{<1>}$ form (A), and proportions of explained variation by combinations of components (B).

A

	Trait 1			Trait 2		
Environments	1	2	3	1	2	3
Genotype 1	-1.606	.009	-.108	-.221	-.445	.557
Genotype 2	-.026	.669	.257	.195	-.338	-.228

B

	Trait 1			Trait 2		
Environments	1	2	3	1	2	3
Genotype 1	43.0	0.0	0.2	0.8	3.3	5.2
Genotype 2	0.0	7.5	1.1	0.6	1.9	0.9

of freedom, far more than than our selected Tucker3 model. The latter model accounted for 64% of the total non-additivity, and gave good descriptions for the non-additivity of the individual traits, except for potassium. A Tucker3 model with 2 components for each way described 58% of the non-additivity, not so very much less than the chosen model, but now not only potassium was badly fitted, but also sodium. The latter provided a reason to prefer the Tucker3 model with 3 components for the environments. A Tucker3 model with 3 components for each way improved the fit slightly, but still did not provide an adequate fit for potassium. The same held for a Tucker2 model with 2 components for the genotypes and 3 components for the environments. The increase in the number of degrees of freedom did not lead to a substantially better fit. A 2 factor Parafac model had a fit that was almost identical to that of the Tucker3 model with 2 components for each mode.

The core matrix is given in Table 3A. The most important combination of components consist of the first component of each way, with a coefficient of -1.606. Note that core elements can be negative, this in contrast to the singular values of two-way matrices. The relative importance of combinations of components can easier be determined by taking the squares of the core elements and dividing them by the total variation, e.g. $(-1.606)^2/6=0.430$ (Table 3B). So 43.0% of the total variation is described by the combination of the first components in the three ways. The total amount of variation described by a component can be found by adding the squared core elements corresponding to the combinations of components in which the particular component occurred. For example, for the first trait component this was $43.0+0.0+0.2+0.0+7.5+1.1=51.8\%$. The second trait component described 12.7%. For the genotypic components this was 52.5% and 12.0%, respectively, and for the environments 44.5%, 12.7% and 7.3%. The sum of these percentages for one way always add up to the total of explained variation, 64.5%.

Plots of the environmental scores revealed that A, B and O type environments were separable. For an example see Figure 1 where the scores of the third

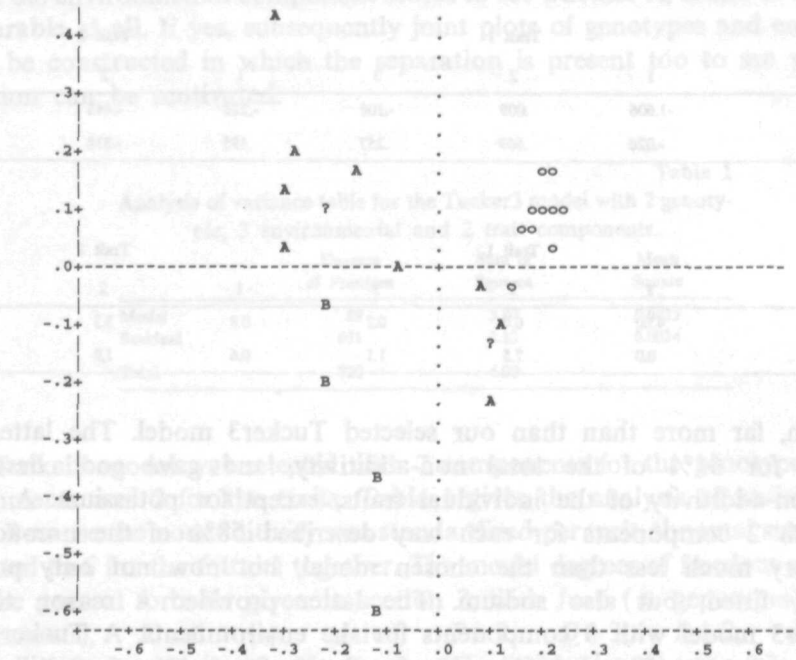

Fig. 1. Plot of environmental component 3 versus 1. A represents an environment (trial) with A type infestation of BNYVV, B an infestation of the B type, ? an infested trial with unknown infestation, while O stands for a non-infested control trial.

environmental component are plotted against those of the first. Reasons for the separation may be deduced from the two joint plots of genotypic and environmental scores, one for each trait component. However, the joint plots as obtained from the analysis dit not separate A nad B type trials (not shown). Still, the observed separation in the environmental plots implies that joint plots should exist that also show the separation. To that end the trait component matrix was rotated. The fit of a Tucker3 model is not changed by a rotation (multiplication with an orthogonal matrix) of a component matrix provided that the inverse rotation is applied to the core matrix. The trait component matrix

Table 4

Trait components before and after rotation.

	Before rotation		After rotation	
Yield	.488	.443	-.140	.644
SugarY	.514	.359	-.054	.625
%Sugar	.458	-.120	.333	.337
Na	-.334	.457	-.563	-.061
aN	.320	-.668	.739	-.057
K	-.270	-.078	-.067	-.273

was the only matrix open to rotation as we wanted to look at the joint plots of genotypic and environmental scores. Rotations of the genotypic and environmental component matrix would have no effect, because the joint plots would stay the same.

Table 5
Core matrix (after rotation) in $G_{<1>}$ form (A), and proportions of explained variation by combinations of components (B).

A

	Trait 1			Trait 2		
Environments	1	2	3	1	2	3
Genotype 1	-.611	.391	-.537	-1.501	-.214	.184
Genotype 2	-.182	.627	.326	.075	.410	.109

B

	Trait 1			Trait 2		
Environments	1	2	3	1	2	3
Genotype 1	6.2	2.51	4.8	37.6	0.8	0.6
Genotype 2	0.6	6.5	1.8	0.1	2.8	0.2

Table 4 shows the scores for the two trait components before and after rotation. The trait component matrix was rotated such that it complied closer with the requirement of simple structure, i.e. traits were forced as much as possible to have substantial scores on only one of both axes.

Table 5A gives the core matrix after rotation. Rotation of the trait components should have no influence on the total amount of variation explained by individual genotypic and environmental components, but the distribution of the explained variation over the individual trait components will be changed. Table 5B is in agreement with this supposition.

The rotation was successful, because now the A, B and O type environments were separated in the joint plots (Figures 2A and 2B). The joint plots are interpreted as follows. First the genotypic and environmental points in the joint plots should be taken to indicate the end points of vectors starting at the origin. To get the interaction according to our Tucker3 model for the genotype Roxane in the trial at Yevre for the trait yield, take the inproduct between the vectors for Roxane and Yevre in Figure 2A and 2B, multiply the inproduct of Figure 2A by the score for yield in the first trait component (-.140), multiply the inproduct of Figure 2B with the yield score in the second trait component (.644), and finally sum both products.

The inproduct between two vectors is just the length of the orthogonal projection of one of both vectors upon the other, times the length of the vector upon which projection took place. For pairs of vectors under obtuse angles the result should be multiplied by -1.

Taking a closer look at the joint plots and the trait component scores we conclude that Figure 2A mainly contains information about sodium and alpha-amino nitrogen concentrations (high absolute trait scores on first component),

whereas Figure 2B represents information on especially yield and sugar yield, and to a lesser extent on percentage of sugar and potassium concentration. Figure 2A shows that Roxane had relatively high alpha-amino nitrogen concentrations at the locations with B type infestation (especially Muret and Yevre) against low

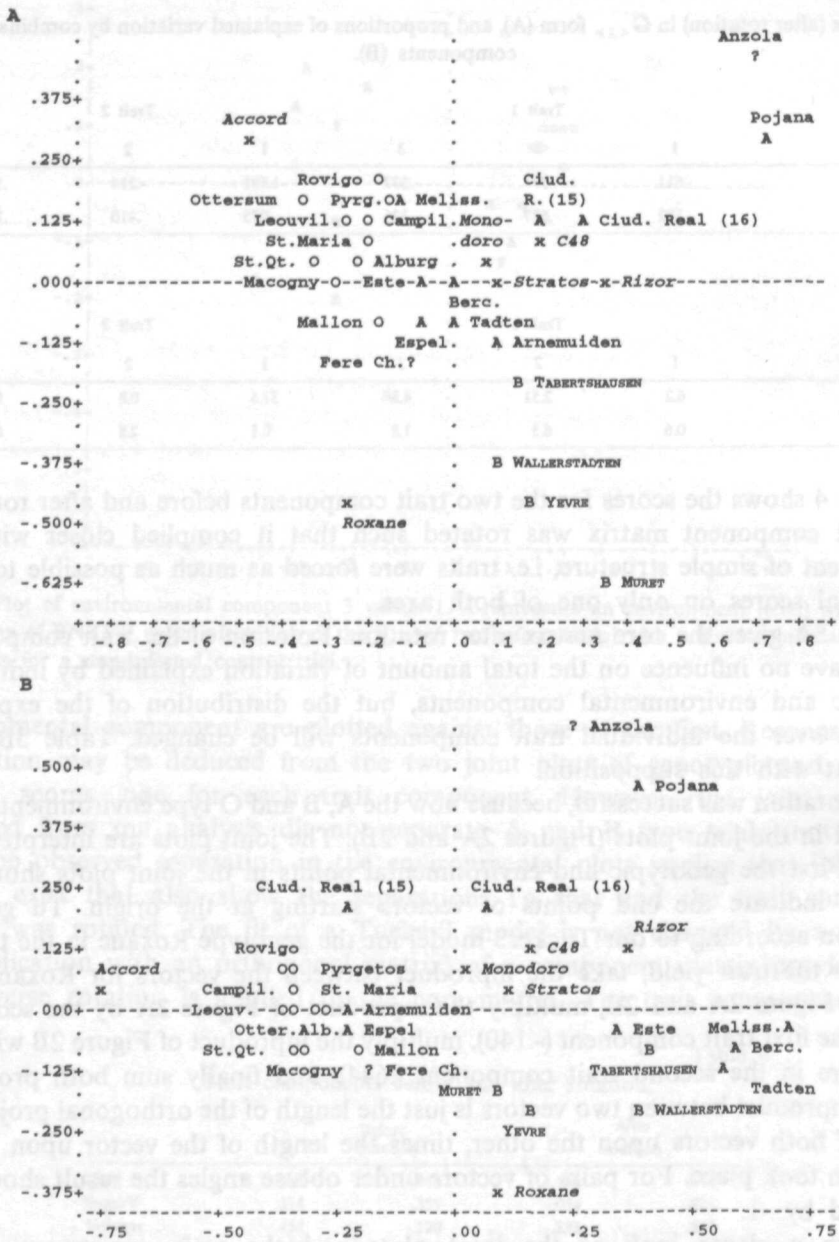

Fig. 2. Joint plot of genotypes (x) and environments (A, B, ? O) for the first trait component (A) and second trait component (B) (after rotation). Genotype names in italic, B type locations in small capitals.

alpha-amino nitrogen concentrations at some A type locations (most notably Pojana). For sodium Roxane behaved just opposite; relatively low concentrations at B type locations, and relatively high concentrations at some A type locations. Another interesting genotype in Figure 2A is Accord. At non-infested locations Accord had relatively high concentrations of alpha-amino nitrogen and low concentrations of sodium. Just the opposite response occurred at locations with B type infestation. Accord had average alpha-amino nitrogen and sodium concentrations at locations with A type infestation. The genotype Rizor had relatively low concentrations of sodium and high concentrations of alpha-nitrogen amino at infested locations independent of the type of infestation. At non-infested locations the roles of sodium and alpha-amino nitrogen were reversed. Similar interpretations are more difficult to obtain from Figure 2B, but the reader is invited to try his luck.

With the joint plots we were able to disclose important information about the differences in genetic background that underlie the resistance to BNYVV in sugar beets and to answer the question whether there is reason to distinguish between an A and B type of BNYVV (Yes). It would have been hard, if not impossible, to arrive at these conclusions by separate analyses of the genotype by environment interaction in individual traits.

6. CONCLUSION

Three-way multiplicative models can be a powerful tool for the analysis of three-way data arrays in general and in particular for the simultaneous analysis of genotype by environment interaction in a number of traits. The Tucker3 model as presented in this paper has a transparent structure and is a straightforward generalization of already existing and widely used two-way multiplicative models. The elaborated example of the sugar beet resistance shows that the Tucker3 model may clarify structures that are inaccessible via two-way methods. Because theoretical demands do not exceed those of two-way models plant breeders should in future be able to apply three-way models to their benefit.

ACKNOWLEDGMENT

The authors wish to thank the IIRB (International Institute for Sugar Beet Research at Brussels) for their permission to use the sugar beet data.

REFERENCES

Basford, K.E., Kroonenberg, P.M., and DeLacy, I.H. (1991). Three-way methods for multiattribute genotype × envirinment data: An illustrated partial survey. *Field Crops Research* **27**, 131-157.

Denis J.-B., and Moro J. (1995). Multivariate generalizations for modeling two-way interaction; Part I: Defining and estimating models. *Biuletyn Oceny Odmian* (*Cultivar Testing Bulletin*), this issue.

Eckart, C. and Young, G. (1937). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211-218.

Gabriel, K. R. (1971). Biplot display of multivariate matrices with application to principal component analysis. Biometrika **58**, 453-467.

Gauch, H. G. Jr. (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Amsterdam: Elsevier.

Harshman R. A., (1970). Foundations of the PARAFAC procedure: Models and methods for an „explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics.* **16**, 1-84.

Harshman, R. A., and Lundy, M. E. (1984). *The PARAFAC model for three-way factor analysis and multidimensional scaling.* In: *Research methods for multimode data analysis* Eds. H. G. Law, C. W. Snyder Jr., J. A. Hattie and R. P. McDonald. New York: Praeger. 122-215.

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

Kroonenberg, P. M. (1983). *Three-mode principal components analysis. Theory and applications.* Leiden: DSWO-Press.

Kroonenberg, P. M., and De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **45**, 69-97.

Paul, H., van Eeuwijk, F. A., and Heijbroek, W. (1993). Multiplicative models for cultivar by location interaction in testing sugar beets for resistance to beet necrotic yellow vein virus. *Euphytica* **71**, 63-74.

ten Berge, J. M. F., de Leeuw, J., and Kroonenberg, P. M. (1987). Some new results on principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **52**, 183-191.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **37**, 3-27.

van Eeuwijk, F. A., Denis, J.-B., and Kang, M. S. (1995). *Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables.* In: *Genotype by environment interaction: New Perspectives.* Ed. M. S. Kang and H. G. Gauch Jr., Boca Raton: CRC-Press, 15-49.

Weesie, J., and van Houwelingen, J. H. (1983). *GEPCAM user's manual: Generalized principal components analysis with missing values.* Technical report, Institute Mathematical Statistics, University of Utrecht.

ANALIZA INTERAKCJI GENOTYPOWO-ŚRODOWISKOWEJ DLA KILKU CECH JEDNOCZEŚNIE PRZY UŻYCIU TRÓJKIERUNKOWEGO MULTIPLIKATYWNEGO MODELOWANIA

Streszczenie: Dla opisu interakcji genotypowo-środowiskowej w przypadku pojedynczych cech odpowiednie okazały się dwukierunkowe modele analizy. W celu przeprowadzenia analizy interakcji genotypowo-środowiskowej dla większej liczby cech jednocześnie zaproponowano trójkierunkowe modele multiplikatywne. Omówiono podstawy teoretyczne dla takich modeli i zilustrowano danymi dotyczącymi odporności buraka cukrowego.

Słowa kluczowe: Interakcja; Model multiplikatywny; Hodowla roślin; Dekompozycja pojedynczej wartości; Burak cukrowy; Model trójkierunkowy; Tablica dwukierunkowa.