LINEAR, BILINEAR, AND LINEAR-BILINEAR MODELS FOR ANALYZING GENOTYPE × ENVIRONMENT INTERACTION

J. Crossa[1], F.A. van Eeuwijk[2], P.L. Cornelius [3], and M. Vargas[4]

[1] Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT) Apdo. Postal 6-641, 06600 Mexico DF, Mexico.[2] Wageningen University, Dep. of Plant Science, Laboratory of Plant Breeding, P.O.Box 386, 6700 AJ Wageningen, The Netherlands.[3] Dep. of Statistics and Dep. of Agronomy, University of Kentucky, Lexington, KY 40546-0091.[4] Universidad Autonoma de Chapingo, Chapingo, Mexico.

Key Words: Least Squares, Singular Value Decompostition, Environmental and Genotypic Covariables.

Introduction

The presence of genotype × environment interaction (G×E) in agriculture is expressed either as inconsistent responses of some genotypes relative to others due to genotypic rank change or as changes in the absolute differences between genotypes without rank change. For the description of the mean response of genotypes over environments and for studying and interpreting G×E in agricultural experiments, three classes of models are commonly used: (1) linear models; (2) bilinear models, and (3) linear-bilinear models. One class of linear models, namely factorial regression (FR) models, and one class of bilinear models, namely partial least square (PLS) regression, allow incorporation of external environmental and genotypic covariables directly into the model.

Early approaches for analyses of G×E included the conventional fixed effects two-way (FE2W) model with sum to zero constraints running over indices. The empirical mean response, $\bar{y}_{ij}$, of the i[th] genotype (i=1,2,…,I) in the j[th] environment (j=1,2,…,J) with n replications in each of the I×J cells is expressed as

$$\bar{y}_{ij} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \bar{\varepsilon}_{ij} \qquad \text{Eq. 1}$$

where μ is the grand mean over all genotypes and environments, τ_i is the additive effect of the i[th] genotype, δ_j is the additive effect of the j[th] environment, $(\tau\delta)_{ij}$ is the non-additivity interaction (G×E) of the i[th] genotype in the j[th] environment and $\bar{\varepsilon}_{ij}$ is the average error assumed to be NID $(0, \sigma^2/n)$ (where σ^2 is the within-environment error variance, assumed to be constant). Yates and Cochran (1938) introduced the model in which the G×E term is linearly related to the environmental main effect.

The purpose of this paper is to present parsimonious approaches other than the FE2W model to the analysis of G×E. Examples illustrating the use of various statistical models for analyzing G×E in the context of plant breeding, genetics, and agronomy are given.

Linear-bilinear models

Williams (1952) was the first author to link the FE2W model with principal components (PC) analysis by considering the model

$\bar{y}_{ij} = \mu + \tau_i + \lambda\alpha_i\gamma_j + \bar{\varepsilon}_{ij}$ where λ is the largest singular value of **ZZ′**and **Z′Z** (for **Z**= $\bar{y}_{ij}$ - $\bar{y}_{i.}$) and α_i and γ_j are the corresponding eigenvectors.

Gollob (1968) and Mandel (1969, 1971) extended Williams' (1952) work by considering the bilinear G×E term as $(\tau\delta)_{ij} = \sum_{k=1}^{t}\lambda_k\alpha_{ik}\gamma_{jk}$. Thus, the general formulation of the linear-bilinear model is

$$\bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t}\lambda_k\alpha_{ik}\gamma_{jk} + \bar{\varepsilon}_{ij} \qquad \text{Eq. 2}$$

where the constant λ_k is the singular value of the k[th] multiplicative component that is ordered $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_t$; the α_{ik} are elements of the k[th] left singular vector of the true interaction and represents the genotypic sensitivity to hypothetical environmental factors represented by the k[th] right singular vector with elements γ_{jk}. The α_{ik} and γ_{jk} satisfy the ortho-normalization constraints $\sum_i \alpha_{ik}\alpha_{ik'} = \sum_j \gamma_{jk}\gamma_{jk'} = 0$ for k≠k′ and $\sum_i \alpha_{ik}^2 = \sum_j \gamma_{jk}^2 = 1$. When Eq. 2 is saturated the number of bilinear terms is t=min(I-1,J-1). Gabriel (1978) described the least squares fit of Eq. 2 and explained how the residual matrix of the G×E term, **Z**= $\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$, is subjected to a singular value decomposition (SVD) after adjusting for the

additive (linear) terms. Gauch (1988) called the Eq. 2 Additive Main Effects and Multiplicative Interaction (AMMI) model.

Other classes of linear-bilinear models, described by Cornelius et al. (1996), are Genotypes Regression Model (GREG) $\bar{y}_{ij} = \mu_i + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$, the Sites (environments) Regression Model (SREG) $\bar{y}_{ij} = \mu_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$, the Completely Multiplicative Model (COMM) $\bar{y}_{ij} = \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$, and the Shifted Multiplicative Model (SHMM) $\bar{y}_{ij} = \beta + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$.

The SHMM model was the first linear-bilinear model used for identifying subsets of genotypes or environments in which genotypic rank changes would be negligible (Cornelius et al., 1992, 1993; Crossa et al., 1993, 1995, 1996; Crossa and Cornelius, 1993). The SREG model is useful in plant breeding because the bilinear terms contain both the main effects of genotypes and G×E. (Crossa and Cornelius, 1997).

In matrix notation, these linear-bilinear models can be expressed as $\mathbf{Y} = \sum_{k=1}^{m} \beta_k \mathbf{X}_k + \mathbf{A\Lambda G'} + \mathbf{E}$ (Cornelius and Seyedsadr, 1997) where $\mathbf{Y}=[\bar{y}_{ij}]$, $\mathbf{X}_k=[x_{kij}]$, $\mathbf{E}=[\bar{\varepsilon}_{ij}]$, $\Lambda=\mathrm{diag}(\lambda_k,\ k=1,2,\ldots,t)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_t$, $\mathbf{A}=(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_t)$, $\mathbf{G}=(\gamma_1,\ldots,\gamma_t)$, and $\mathbf{A'A}=\mathbf{G'G}=\mathbf{I}_t$. The x_{kij} are known constants and β_k, λ_k, α_{ik}, and γ_{jk} are parameters to be estimated.

Linear models

The G×E is modeled directly using regression on environmental (and/or genotypic) variables. A useful linear model for incorporating external environmental (or genotypic) variables is the factorial regression (FR) model (Denis, 1988; van Eeuwijk et al., 1996). The FR models are ordinary linear models that approximate the G×E effects of Eq. 1 by the products of one or more of (1) genotypic covariables (observed) × environmental potentialities (estimated), (2) genotypic sensitivities (estimated) × environmental covariables (observed). For k=1,…,G genotypic covariables (centered) represented by $x_{i1}, \ldots, x_{iG}$, Eq. 1 becomes $\bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{g=1}^{G} x_{ig} \xi_{jg} + \bar{\varepsilon}_{ij}$, $G \leq I-1$, where ξ_{jg} represents an environmental factor (regression coefficient) with respect to the genotypic covariable, x_{ig}. Constraints on the parameters are $\sum_i \tau_i = \sum_j \delta_j = \sum_j \xi_{jg} = 0$. In matrix notation the expectation is

$$E(\mathbf{Y}) = \mu \mathbf{1}_I \mathbf{1}'_J + \tau \mathbf{1}'_J + \mathbf{1}_I \delta' + \mathbf{X}\Xi' \qquad \text{Eq. 3}$$

where $\mathbf{Y}=[\bar{y}_{ij}]$ is a $I \times J$ matrix; $\mathbf{1}_I$ and $\mathbf{1}_J$ are $I \times 1$ and $J \times 1$ vectors with all elements equal to one, respectively; $\tau=[\tau_I]$ is the $I \times 1$ vector of main effects of genotypes; $\delta=[\delta_j]$ is the $J \times 1$ vector of main effects of environments; $\mathbf{X}=[x_{ig}]$ is the $I \times G$ matrix of known genotypic covariables; $\Xi=[\xi_{jg}]$ is the $J \times G$ matrix of unknown environmental factors (regression coefficients).

For h=1,…, H environmental covariables (centered) represented by $z_{j1}, \ldots, z_{jH}$, Eq. 1 is $\bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{h=1}^{H} \varsigma_{ih} z_{jh} + \bar{\varepsilon}_{ij}$, $H \leq J-1$, where ς_{ih} represents an genotypic sensitivity (regression coefficient) with respect to the to the environmental covariable, z_{jh}. Constraints on the parameters are $\sum_i \tau_i = \sum_j \delta_j = \sum_i \varsigma_{ih} = 0$. In matrix notation, the expectation is

$$E(\mathbf{Y}) = \mu \mathbf{1}_I \mathbf{1}'_J + \tau \mathbf{1}'_J + \mathbf{1}_I \delta' + \zeta \mathbf{Z}' \qquad \text{Eq. 4}$$

where $\mathbf{Z}=[z_{jh}]$ is the $J \times H$ matrix of known environmental covariables; $\zeta=[\varsigma_{ih}]$ is the $I \times H$ matrix of unknown differential genotypic sensitivities.

The FR model including genotypic and environmental covariables simultaneously is $\bar{y}_{ij} = \mu + \tau_i + \delta_j + \sum_{g=1}^{G} x_{ig} \xi_{jg} + \sum_{h=1}^{H} \varsigma_{ih} z_{jh} + \sum_{h=1}^{H} \sum_{g=1}^{G} x_{ig} \nu_{gh} z_{jh} + \bar{\varepsilon}_{ij}$ where ν_{kh} is a constant that scales the cross-product of the genotypic covariables, x_k, with the environmental covariables, z_h, and can be derived from the two previous FR models by imposing the restriction $\xi_{jg} = \nu_{gh} z_{jh}$ or $\varsigma_{ih} = x_{ih} \nu_{gh}$; each

cross product represents one degree of freedom in the G×E subspace. In matrix notation the expectation is $E(\mathbf{Y})=\mu\mathbf{1}_I\mathbf{1}'_J + \tau\mathbf{1}'_I + \mathbf{1}_j\delta' + \mathbf{X}\nu\mathbf{Z} + \mathbf{X}\Xi'+\zeta\mathbf{Z}'$ where the constraint $\mathbf{X}\Xi'=\zeta\mathbf{Z}'=\mathbf{0}$ (where $\mathbf{0}$ is a matrix H×G of zeros) is required. The model should be fitted for all possible combinations of genotypic covariables with environmental covariables.

When environmental (or genotypic) covariables show high collinearity, interpretation of the least squares regression coefficients is complicated because they are estimated very imprecisely. Consequently, a stepwise procedure for choice of the covariables to include could be useful for model construction. Noise on the response variable also complicates the interpretation of the FR parameters. Furthermore, least squares estimation of the parameters in the FR models are not unique when the number of covariables is larger than the number of observations, so an alternative estimation method is needed. Partial Least Squares (PLS) regression overcomes some of these problems and it can be used as an alternative estimation method.

Bilinear models

Multivariate Partial Least Squares (PLS) regression models (Aastveit and Martens, 1986; Helland, 1988) are a special class of bilinear models. When genotypic responses over environments (**Y**) are modeled using environmental covariables, then the *J*×*H* matrix **Z** of H (*h*=1,2,…,*H*) environmental covariables can be written in a bilinear form as

$$\mathbf{Z}=\mathbf{t}_1\mathbf{p}'_1+\mathbf{t}_2\mathbf{p}'_2+\ldots+\mathbf{t}_M\mathbf{p}'_M+\mathbf{E}_M=\mathbf{T}\mathbf{P}'+\mathbf{E} \qquad \text{Eq. 5}$$

where the matrix **T** contains the $\mathbf{t}_J$ *J*×1 vectors called latent environmental covariables or Z-scores (indexed by environments) and the matrix **P** has the $\mathbf{p}_1\ldots\mathbf{p}_H$ *H*×1 vectors called Z-loadings (indexed by environmental variables) and **E** has the residuals. Similarly, the response variable matrix **Y** in bilinear form is

$$\mathbf{Y}=\mathbf{t}_1\mathbf{q}'_1+\mathbf{t}_2\mathbf{q}'_2+\ldots+\mathbf{t}_M\mathbf{q}'_M+\mathbf{F}_M=\mathbf{T}\mathbf{Q}'+\mathbf{F} \qquad \text{Eq. 6}$$

where the matrix **T** is as in Eq. 5 and the matrix **Q** contains the $\mathbf{q}_1\ldots\mathbf{q}_I$ *I*×1 vectors called Y-loadings (indexed by genotypes) and **F** has the residuals. The relationship between **Y** and **Z** is transmitted through the latent variable **T**. The PLS algorithm performs separate (but simultaneous) principal component analysis of **Z** and of **Y** that allows reduction of the number of variables in each system to a smaller number of hopefully more interpretable latent variables.

Helland (1988) showed that a reduced number of PLS latent variables gives a low rank representation of the least squares estimates of the FR with environmental covariables because the expectation of $\mathbf{Y}'$ is

$$E(\mathbf{Y}')=\mathbf{Q}\mathbf{T}'=\mathbf{Q}(\mathbf{Z}\mathbf{W})'=(\mathbf{Q}\mathbf{W}')\mathbf{Z}'=\zeta\mathbf{Z}'=\sum_{h=1}^{H}\varsigma_{ih}z_{jh} \qquad \text{Eq. 7}$$

as in Eq. 4 where **T**, **Q**, and **Z** are defined as before and the vector **W** is *H*×1 and contains the Z-loadings (or weights) of the environmental covariables; ζ contains the PLS approximation to the regression coefficients of the responses in **Y** to the environmental covariables in **Z**. The matrices **T** (with *J* coordinates for environments), **Q** (with *I* coordinates for genotypes) and **W** (with *H* coordinates for environmental covariables) can be represented in the PLS biplot such that projecting the j^{th} environment (row) of **T** on the i^{th} genotype (row) of **Q** [$\mathbf{Y}'=(\mathbf{T}\mathbf{Q}')'$] approximates the G×E; projecting the h^{th} environmental covariable (row) of **W** on the i^{th} genotype (row) of **Q** ($\mathbf{Q}\mathbf{W}'=\zeta$) approximates the regression coefficient of the i^{th} genotype on the h^{th} environmental covariable (Vargas et al.,1999; van Eeuwijk et al., 2000). When genotypic covariables are used to model environmental responses over genotypes, then the latent genotypic covariables are $\mathbf{T}=\mathbf{X}\mathbf{W}$ where vector **W** is *G*×1 and contains the weights of the genotypic covariables. The expectation of **Y** is

$$E(\mathbf{Y})=\mathbf{T}\mathbf{Q}'=\mathbf{X}\mathbf{W}\mathbf{Q}'=\mathbf{X}\Xi'=\sum_{g=1}^{G}x_{ig}\xi_{jg} \qquad \text{Eq. 8}$$

as in Eq. 3 (van Eeuwijk et al., 2000; Vargas et al., 1999) where Ξ contains the PLS approximation to the regression coefficients of the responses in **Y** to the genotypic covariables in **X**. The matrices **T** (with *I* coordinates for genotypes), **Q** (with *J* coordinates for environments) and **W** (with *G* coordinates for genotypic covariables) can be represented in a PLS biplot such that projection of the i^{th} genotype (row) of **T** onto the j^{th} environment (row) of **Q** ($\mathbf{Y}=\mathbf{T}\mathbf{Q}'$) approximates the G×E; projection the g^{th} genotypic covariable (row) of **W** onto the j^{th} environment (row) of **Q** ($\mathbf{W}\mathbf{Q}'=\Xi$) approximates the regression coefficient of the j^{th} environment on the g^{th} genotypic covariable.

QTL and QTL× environment interaction analysis in genetics and plant breeding

In plant breeding much research is directed at locating the regions of the chromosomes that are involved in the physiological processes underlying phenotypical traits. These regions are called quantitative trait loci (QTL or QTLs). When these regions differ between genotypes in relation to changes in the environment, QTL×environment interaction occurs (QTL×E). The statistical problem can be interpreted as a multivariate multiple regression of phenotypic traits as observed over a set of environments on a set of genetic predictors. FR provides a suitable framework for QTL×E analysis. In

Crossa et al. (1999) examples are given of how FR and PLS can be used for assessing location and importance of QTL and QTL×E.

FR models of the form

$$\bar{y}_{ij} = \mu + \delta_j + \sum_{g=1}^{G} x_{ig}\xi_g + \bar{\varepsilon}_{ij} \text{ and}$$

$$\bar{y}_{ij} = \mu + \delta_j + \sum_{g=1}^{G} x_{ig}\xi_{jg} + \bar{\varepsilon}_{ij}$$ (van Eeuwijk, et al., 2000) are useful for studying QTL and QTL×E, respectively where x_{ig}'s are genotypic covariables, or genetic predictors, at specific locations of the chromosmes, whose values are functions of the neighboring genetic markers and the position at the chromosome. The ξ_g's represent the QTL main effects, which are indexed by environment, ξ_{jg}, represents the QTL×E. Following Haley and Knott (1992), the simplest QTL mapping analysis considers the regression on the genetic predictors at marker positions (individual marker regression) where the additive effects of the observed marker genotypes *MM*, *Mm* and *mm* are 1,0, and –1, respectively and the dominance effects for *MM*, *Mm* and *mm* are 0,1, and 0, respectively. Somewhat more advanced, simple interval QTL mapping analysis considers the regression on the genetic predictors not only at marker positions but also at regular intervals between markers. The additive effect and dominance effects can be computed.

In composite interval QTL mapping analysis a correction is added for the effects of QTLs at other positions in the genome. Let the position under evaluation be *p*, then other markers, called cofactors, *C*, are included in the model to reduce noise created by the effect of other QTLs, then the model is $\bar{y}_{ij} = \mu + \delta_j + \sum_{g \in C} x_{ig}\xi_g + x_{ip}\xi_p + \bar{\varepsilon}_{ij}$. Selection of the appropriate markers to be used as cofactors for correcting the effect of other QTL can be done by one of a few PLS-axes created by regressing the multivariate response on all genetic predictors outside the evaluation window in composite interval mapping, and then perform the mapping procedure with the corrected responses. Testing procedures for the presence of QTL and QTLxE at a certain position can be done by permutation tests (van Eeuwijk, at al., 2000).

Some results of the application of the methodology described in van Eeuwijk et al. (2000) now follow. The grain yield of F_2 (211) tropical CIMMYT maize lines lines was evaluated in eight environments that were contrasting in drought and nitrogen stress. As FR is essentially a regression method, QTLs and QTL×E can be located by the application of standard F-statistics. Plot of F-profiles over the first chromosome is in Fig. 1. Based on randomization studies thresholds should be applied of about 54, 4.5 and 9, respectively (α=0.05).

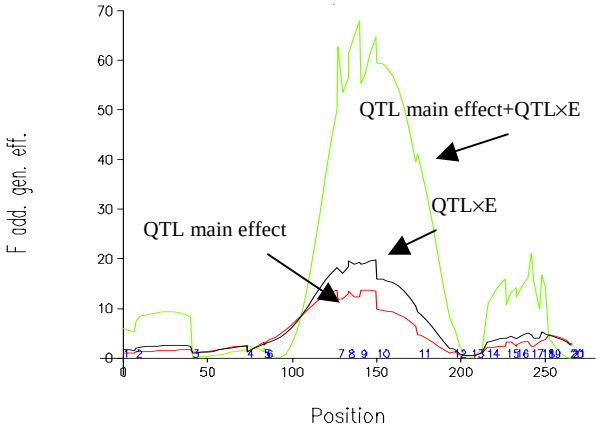

Figure 1. F-values for the different positions along the first chromosome for additive genetic QTL main effects, additive genetic QTL×E, and QTL main effects+QTL×E.

Treatment × environment interaction analysis in agronomy

A parsimonious description of the Treatment × Environment (T×E) existing in 24 agronomic treatments (1-24) [tillage, summer crop, manure, and nitrogen (N)] evaluated during 10 consecutive years (1988-97) was conducted by Vargas et al. (2001). Results of the final MFR were compared with those of a Partial Least Squares (PLS) to achieve extra insight in both the T×E and the final multiple factorial regression (MFR) model.

The MFR was applied on the six most important components of the T×E term: Year×Tillage, Year×Summer Crop, Year×Manure, Year×N, Year×Summer Crop×N, and Year× Manure× N. Results for the MFR of the 27 environmental covariables × tillage interaction showed that evaporation in December (EVD) × tillage sum of squares accounted for 68% of the whole year × tillage interaction. For year × summer crop, evaporation in April (EVA) accounted for 36% of the year × summer crop. For year × manure, covariables precipitation in December (PRD) and sun hours in February (SHF) contributed with 56% of the year × manure sum of squares. Year × Nitrogen (N) interaction determined the major part of year × treatment interaction sum of squares.

The PLS biplot separated the nine highest yielding treatments (9,19,21,17,11,12,10,23, and 18) from the nine lowest yielding treatments (1,2,3,4,5,6,7,8, and 16) (Fig. 3). The nine lowest yielding treatments had a

positive interaction with year 1995 that had high mTUF, mTF, and MTA but a negative interaction with year 1988 (opposite quadrant). The PLS biplot contains roughly five clusters of correlated environmental covariables. The order of inclusion of these covariables in the MFR with the stepwise procedure for each factor effect corresponds to selecting covariables for the different cluster groups depicted in Fig. 3.

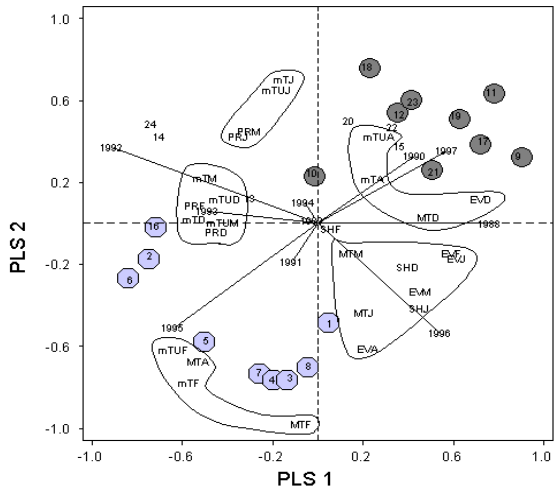

Figure 3. Biplot of the first and second PLS factors representing the Z-scores of the 10 years (1988-97), and the Y-loadings of the 24 practice treatments (1-24) enriched with the Z-loadings of 27 environmental variables: EV: total monthly evaporation, PR: total monthly precipitation, SH: sun hours per day, mT: mean minimum temperature sheltered, MT: mean maximum temperature sheltered, mTU: mean minimum temperature unsheltered; D: December, J: January, F: February, M: March, A: April; N: Nitrogen (from Vargas et al., 2001).

Crossover interaction analysis in plant breeding

Using linear-bilinear SHMM model, Cornelius et al. (1992) defined sufficient conditions for the absence of significant genotype crossover interaction (COI) in a set of environments and/or genotypes: (1) SHMM with t=1 ($SHMM_1$) must be an adequate model for fitting the data and (2) γ_{j1} are all of like sign. When $SHMM_1$ predicted values, $\hat{y}_{ij} = \hat{\beta} + \hat{\lambda}_1 \hat{\alpha}_{i1} \hat{\gamma}_{j1}$, are plotted against the primary effects of environments, $\hat{\gamma}_{j1}$, the graph consists of a set of regression lines, one for each genotype, all of which concur at the point (0, $\hat{\beta}$). For a non-COI $SHMM_1$, the $\hat{\gamma}_{j1}$ are all of like sign (or zero) and, thus, the point of intersection is a point either at the boundary (if one $\hat{\gamma}_{j1}$=0), or outside (left or right) of the region containing the plotted points. If the $\hat{\gamma}_{j1}$ have different signs, then the point of concurrence is within the region containing the plotted points and a complete reversal of rank order of genotypes is displayed on the right, as compared to the left, of the point of concurrence. For clustering environments, $SREG_1$ can be used instead of $SHMM_1$ and all the above properties still hold then the figure shows an overlayed set of broken lines (one for each genotype) that display no genotype COI within the region of plotted points.

When SHMM is fitted to the entire set of data, several components are necessary if an adequate fit is to be achieved. The procedure by a which subset of environments without COI is found consists in using a clustering strategy that will divide the environments into subsets such that significant variation captured as secondary, tertiary, etc., effects when SHMM is fitted to the entire data set, can be expressed as primary effects in separate analyses of data from the subsets. The measure of distance between two environments is the residual mean square after fitting $SHMM_1$ [$RMS(SHMM_1)$] to the data from the two environments subject to a non-COI constraint.

Data from a trial with g=9 genotypes evaluated in e=20 environments showed that $SHMM_1$ will not adequately fit the entire data and the fitted $SHMM_1$ itself displayed genotype COI. In Fig. 4, the consistant response of the nine genotypes across a subset of ten environments is depicted through the overlayed broken line $SREG_1$ that does not cross over.

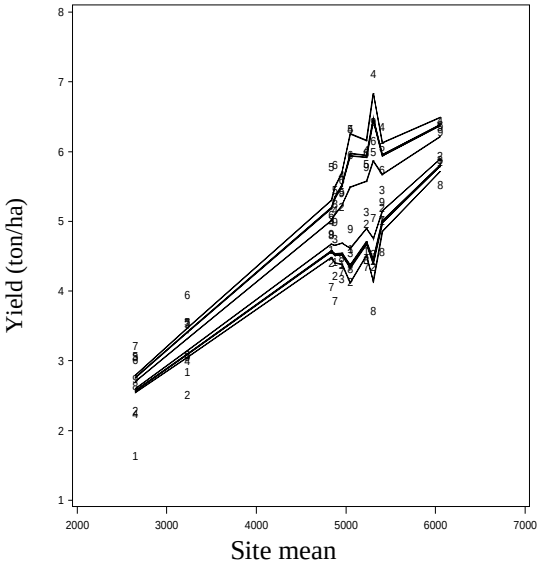

Figure 4. $SREG_1$ model fitted to nine genotypes and a subset of ten environments.

References

Aastveit, H. and H. Martens (1986). Anova interactions interpreted by partial leats squares regression. Biometrics 42:829-844.

Cornelius, P.L, Seyedsadr, M., and Crossa, J. (1992). Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theoretical and Applied Genetics 84:161-172.

Cornelius, P.L.,Van Sanford, D.A, and Seyedsadr, M. (1993). Clustering cultivars into groups without rank-change interactions. Crop Science 33:1193-1200.

Cornelius, P.L., Crossa, J., and Seyedsadr, M.(1996). Statistical tests and estimators for multiplicative models for cultivar trials. In Kang, M.S., and Gauch, H.G., Jr., (Eds.), Genotype-by-Environment Interaction. Boca Raton: CRC Press, pp. 199-234.

Cornelius, P.L., and Seyedsadr, M. (1997). Estimation of general linear-bilinear models for two-way tables. Journal of Statistical Computation and Simulation 58: 287-322.

Crossa, J., Cornelius P.L., Seyedsadr, M., and Byrne, P. (1993). A shifted multiplicative model cluster analysis for grouping environments without genotypic rank-change. Theoretical and Applied Genetics 85:577-586.

Crossa, J., and Cornelius, P.L. (1993) Recent developments in multiplicative models for cultivar trials. In: Buxton, D.R., et al. (Eds.) International Crop Science I. Crop Science Society of America Madison, Wiscosin.

Crossa, J., Cornelius, P.L., Sayre, K., and Ortiz-Monasterio I.J. (1995). A shifted multiplicative model fusion method for grouping environments without cultivar rank change. Crop Science 35:54-62.

Crossa, J., Cornelius, P.L., and Seyedsadr, M.(1996). Using the shifted multiplicative model cluster methods for crossover genotype-by-environment interaction. In Kang, M.S. and Gauch, H.G., Jr., (Eds.), Genotype-by-Environment Interaction. Boca Raton: CRC Press, pp. 175-198.

Crossa, J., and Cornelius, P.L. (1997). Sites regression and shifted multiplicative model clustering of cultivar trials sites under heterogeneity of error variance. Crop Science 37: 405-415.

Crossa, J., M. Vargas, F.A. van Eeuwijk, C. Jiang, G.O. Edmeades, and D. Hoisington. (1999). Interpreting genotype×environment interaction in tropical maize using linked molecular markers and environmental covariables. Theoretical and Applied Genetics 99:611-625.

Denis, J-B. (1988). Two-way analysis using covariates. Statistics 19:123-132.

Gabriel, K.R. (1978). Least squares approximation of matrices by additive and multiplicative models. Journal of the Royal Statistical Society, Series B, 40:186-196.

Gauch, H.G. Jr. (1988). Model selection and validation for yield trials with interaction. Biometrics 44:705-715.

Gollob, H.F. (1968). A statistical model which combines features of factor analytic and analysis of variance. Psychometrika 33:73-115.

Helland, I.S. (1988). On the structure of partial least squares. Communications in Statistics, Part B Simulations and Computations 17:581-607.

Haley, C.S. and S.A. Knott (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315-324.

Mandel, J. (1969). The partitioning of interaction in analysis of variance. Journal of Research of the National Bureau of Standards, Series B, 73:309-328.

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

Vargas, M., J. Crossa, F.A. van Eeuwijk, M.E. Ramirez and K. Sayre. (1999). Using partial Least Squares, Factorial Regression and AMMI models for interpreting Genotype×Environment Interaction. Crop Science 39:955-967.

Vargas, M., J. Crossa, F.A. van Eeuwijk, K. Sayre, and M.P. Reynolds. (2001). Interpreting Treatment×Environment interaction in Agronomy Trials. Agronomy Journal (in press).

van Eeuwijk, F.A., J-B. Denis, and M.S. Kang. 1996. Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In S. Kang and H.G. Gauch (eds). Genotype-by-environment interaction, CRC Press, Boca Raton, Fl.

van Eeuwijk, F.A., J. Crossa, M. Vargas, J.M. Ribaut (2000). Variants of factorial regresión for analysing QTL by environment interaction. Proceedings of the 11th Meeting of the EUCARPIA Section Biometrics in Plant Breeding. In A. Gallais, C. Dillmann, and I. Goldringer (eds). Quantitative genetics and breeding methods: the way ahead. Paris, France.

Williams, E.J. (1952) The interpretation of interactions in factorial experiments. Biometrika 39:65-81.

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