Santiago, 14 de enero de 2005

PROGRAMA DE FORMACION
Recepcionado 17-1-05
Nº Ingreso 020

Sra.
María José Etchegaray
Jefa Programa de Formación y Promoción
Fundación para la Innovación Agraria
Presente

De mi consideración

Respecto a las observaciones realizadas al Informe Técnico Final de la propuesta "Introducción al Análisis Multivariado", código FIA-FR-V-2004-1-A-010, que dicen relación con profundizar los resultados del Curso y la Aplicabilidad de los conocimientos entregados por el Dr. Crossa, paso a complementar lo que se señala en dicho informe.

Uno de los aspectos más complejos de lograr en el análisis del comportamiento de variedades en diferentes medioambientes, es la adecuada interpretación de la interacción entre Genotipo y Medioambiente (GxM). Los conocimientos adquiridos en este curso permiten una mejor compresión de esta interacción. En primer lugar, diferentes variables que tienen algún grado de asociación se sintetizan en componentes principales, y luego a través de una técnica gráfica (construcción de Biplots) se logra asociar la síntesis de variables ambientales, fisiológicas u otras con el comportamiento de los genotipos estudiados. Esto permite incursionar en las causas de la interacción GxM y con ello precisar los requerimientos de adaptación de los genotipos a un determinado sitio, haciendo mucho más eficiente el proceso de fitomejoramiento. Al poner en términos explícitos (visibles) las interacciones complejas, se aprovecha en mucho mejor forma la información experimental que entregan los ensayos agronómicos, permitiendo disminuir notablemente el número de experimentos necesarios para lograr un determinado propósito.

Las herramientas del Análisis Multivariado entregadas por el Dr. Crossa son de aplicación más general y permiten su utilización en diferentes campos del conocimiento en que las interacciones dificulten la interpretación de los resultados. Así por ejemplo dentro de los 10 días siguientes al curso realizado por el Dr. Crossa se aplicaron estas técnicas a la interpretación del efecto de múltiples variables del suelo en el rendimiento de cultivos sometidos a diferentes tipos de manejo de suelo. Fue posible asociar los cambios de las variables de suelo a los tipos de manejo, y a su vez analizar el efecto del manejo en el rendimiento. Los resultados de este análisis se presentaron en el 55° Congreso de la Sociedad Agronómica de Chile, realizado a fines de octubre en la Universidad Austral de Valdivia. Las presentaciones en estas materias fueron seguidas con gran interés por la audiencia, por cuanto este tipo de interpretación es novedoso y de gran utilidad para el medio agronómico nacional.

Adjunto, además, en esta oportunidad la impresión del articulo "Linear-bilinear models for the analysis of genotype-environment interaction", el examen que envió el Dr. Crossa y el

(7)

listado de alumnos que cumplieron los requisitos y que, en consecuencia, aprobaron el curso. Este curso fue reconocido por el Programa de Doctorado CSAV de la Universidad de Chile (ver carta anexa). Estaré atento a cualquier otra inquietud que FIA tenga sobre el particular. Reitero mis agradecimientos a FIA por la ayuda financiera otorgada, en el convencimiento que actividades como ésta realmente contribuyen al desarrollo agrícola nacional al aumentar notablemente la eficiencia de la investigación y extensión agronómica.

Sin otro particular, le saluda atentamente

Prof. Edmundo Acevedo H.
Laboratorio de Relación Suelo-Agua-Planta
Facultad de Ciencias Agronómicas
Universidad de Chile

20 Linear–Bilinear Models for the Analysis of Genotype–Environment Interaction

J. Crossa[1] and P.L. Cornelius[2]

[1]Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Lisboa 27, Apdo. Postal 6-641, 06600 Mexico DF, Mexico; [2]Departments of Agronomy and Statistics, University of Kentucky, Lexington, KY 40546-0091, USA

Introduction

The presence of genotype–environment interaction (GEI) in a multi-environment trial (MET) is expressed either as inconsistent responses of some genotypes with respect to others due to the alteration of the ordering of the genotypes from one environment to another (GEI with rank change or crossover interaction (COI)) or as changes in the absolute differences between genotypes without rank change (GEI without rank change or non-crossover interaction (non-COI)).

Early approaches to the analysis of GEI included the conventional, two-way fixed-effects (FE2W) model, with sum to zero constraints, and the simple linear regression of genotype yields on the environment means. The FE2W model expresses the empirical mean, $\bar{Y}_{ij}$, of the ith genotype ($i = 1, 2, \ldots, g$) in the jth environment ($j = 1, 2, \ldots, e$) with n replications in each of the $g \times e$ cells as:

$$\bar{Y}_{ij} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \bar{\varepsilon}_{ij} \qquad (20.1)$$

where μ is the grand mean across all genotypes and environments and is estimated by $\bar{Y}_{..}$. τ_i is the additive effect of the ith genotype defined as a deviation of the genotype mean from the overall mean and is estimated by $\bar{Y}_{i.} - \bar{Y}_{..}$ which satisfies constraint $\sum_i \hat{\tau}_i = 0$. Similarly, δ_j is the additive effect of the jth environment estimated by $\bar{Y}_{.j} - \bar{Y}_{..}$ which satisfies constraint $\sum_j \hat{\delta}_j = 0$. Also, $(\tau\delta)_{ij}$ is the non-additivity, i.e. GEI, of the ith genotype and the jth environment estimated as the residual $\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$ after fitting the main effects. The $(\tau\delta)_{ij}$ satisfies constraints $\sum_i \sum_j (\tau\delta)_{ij} = \sum_i (\tau\delta)_{ij} = \sum_j (\tau\delta)_{ij} = 0$. The $\bar{\varepsilon}_{ij}$ term is the mean of the errors contributing to measurements on the ith genotype in the jth environment. The $\bar{\varepsilon}_{ij}$ are assumed NID(0, σ^2/n) where σ^2 is the pooled within-environment error variance, assumed to be homoscedastic. For the complete random effects two-way (RE2W) model, τ_i, δ_j and $(\tau\delta)_{ij}$ are assumed to be normally and independently distributed, with variances σ^2_τ, σ^2_δ and $\sigma^2_{\tau\delta}$, respectively. From the GEI perspective, the FE2W and RE2W models are unparsimonious, the analysis is uninformative and the $(g-1)\times(e-1)$ independent parameters of the GEI are difficult to interpret.

Fisher and MacKenzie (1923), who analysed data from an experiment evaluating 12 potato cultivars under each of six soil-fertilization treatments, were the first authors to propose breaking down the

CCAB International 2002. Quantitative Genetics, Genomics and Plant Breeding (ed. M.S. Kang)

response into a series of multiplicative terms fitted by least squares; however, this work was apparently forgotten for many years. Yates and Cochran (1938) proposed breaking down the GEI into one multiplicative term and a deviation therefrom, examining whether the GEI is a linear function of the additive environmental component, that is, $(\tau\delta)_{ij} = \xi_i\delta_j + d_{ij}$, where $1 + \xi_i$ is the linear regression coefficient of yields of the ith genotype on the environmental mean and d_{ij} is a deviation. This regression approach expresses the GEI simply as heterogeneity of slopes and was later used by Finlay and Wilkinson (1963) and slightly modified by Eberhart and Russell (1966).

Tukey (1949) proposed a test for the GEI in which the $(\tau\delta)_{ij}$ term is a constant multiplied by the product of the main effects of genotypes and environments, $(\tau\delta)_{ij} = \lambda\tau_i\delta_j$. Mandel (1961) generalized Tukey's (1949) model by letting the GEI term be the product $(\tau\delta)_{ij} = \lambda\alpha_i\delta_j$ for regression of genotype simple effects on environment main effects or $(\tau\delta)_{ij} = \lambda\tau_i\gamma_j$ for regression of environment simple effects on genotype main effects. Either of these consists of a "bundle of regression lines", which may be tested for concurrence (i.e. in the first case, whether the α_i are proportional to the τ_i or, in the second case, whether the γ_j are proportional to the δ_j) or non-concurrence. Proportionality of the α_i to the τ_i is a special case of regression of genotypes on the environmental mean, in which the regression lines all intersect at one point.

The Mandel (1961) tests for concurrent and non-concurrent regression lines partition the GEI into one degree of freedom (d.f.) for the concurrence of genotype (or environment) regressions on environment (or genotype) main effects (this is the same as Tukey's (1949) one d.f. for non-additivity), $g-2$ d.f. for the non-concurrence of genotype regressions, $e-2$ for the non-concurrence of site regressions and $(g-2)(e-2)$ d.f. for the remainder of the GEI. Cornelius *et al.* (1996) suggested Mandel's (1961) analysis as a diagnostic for choice of multiplicative model form.

Freeman (1973) cited Williams (1952) as the first researcher to link the FE2W model with principal-component analysis (PCA) and showed that the GEI term can be represented by the sum of eigenvalues of a matrix. Gollob (1968) and Mandel (1969, 1971) introduced the linear–bilinear model (LBM),

$$\bar{Y}_{ij} = \mu + \tau_i + \delta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij} \quad (20.2)$$

Where λ_k is a scale parameter or singular value for the kth bilinear (multiplicative) component. The λ_k are ordered, i.e. $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_t$. Further, the α_{ik} and γ_{jk} are elements of the left and right singular vectors, respectively, contributing to the kth bilinear (multiplicative) term. The α_{ik} represents genotypic sensitivities to a hypothetical environmental factor the level of which, in the jth environment, is represented by the element γ_{jk} in the right singular vector for the kth component. Elements of the singular vectors for genotypes and environments (α_{ik} and γ_{jk}) are subject to normalization constraints, $\Sigma_i \alpha_{ik}^2 = \Sigma_j \gamma_{jk}^2 = 1$, and to orthogonality constraints, $\Sigma_i \alpha_{ik}\alpha_{ik'} = \Sigma_j \gamma_{jk}\gamma_{jk'} = 0$ for $k \neq k'$. When Equation 20.2 is saturated, the number of bilinear terms is $t = \min(g-1, e-1)$ and, for any smaller value of t, the model is said to be 'truncated'. The word 'truncated' here is used not in the sense of having a truncated distribution, but, rather, in the sense of truncating the string of bilinear terms at something less than the number of terms that will saturate the model.

Mandel (1971) computed the number of degrees of freedom associated with the sum of squares (SS) due to each of the first three bilinear terms in Equation 20.2 by a Monte Carlo study, and Johnson and Graybill (1972) found that Mandel's (1971) results were close to the exact values. Gabriel (1978) showed that a least-squares (LS) solution for model parameters in Equation 20.2 can be obtained by taking the estimates of the bilinear terms as the t largest components of the singular value decomposition (SVD) of the matrix $\mathbf{Z} = [z_{ij}] = [\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}]$, with the additive (linear) effects μ, τ_i and δ_j estimated as we have previously given for their estimates in the FE2W model (Equation 20.1).

If the components of the SVD of $\mathbf{Z}$ are arranged in decreasing order with respect to the singular values, the first component

gives a rank-one matrix that, in an LS sense, approximates matrix **Z**; the first two components of the SVD give a rank-two matrix that approximates **Z**, etc.

Zobel *et al.* (1988) and Gauch (1988) named Equation 20.2 the "additive main effects and multiplicative interaction" (AMMI) model. They further introduced a data-splitting and cross-validation procedure for determining the number of multiplicative components to retain in a truncated AMMI model.

The AMMI model and four other LBMs and their LS estimates were described and unified in one general methodology by Cornelius *et al.* (1996). These authors described various statistical tests for the significance of the bilinear terms and mentioned the possibility of using shrinkage estimates of these models for improving the prediction of the $g \times e$ cells. The four models that can be derived from the AMMI model are:

genotypes regression model (GREG)

$$\bar{Y}_{ij} = \mu_i + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$$

sites (i.e. environment) regression model (SREG)

$$\bar{Y}_{ij} = \mu_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$$

completely multiplicative model (COMM)

$$\bar{Y}_{ij} = \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$$

shifted multiplicative model (SHMM)

$$\bar{Y}_{ij} = \beta + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$$

The LS estimates of the additive effects of these models and the elements of the residual matrix **Z** are:

AMMI	$\hat{\mu} = \bar{Y}_{..}, \hat{\tau}_j = \bar{Y}_{.j} - \bar{Y}_{..}, \hat{\delta}_i = \bar{Y}_{i.} - \bar{Y}_{..}, z_{ij} = \bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$
SREG	$\hat{\mu}_j = \bar{Y}_{.j}, z_{ij} = \bar{Y}_{ij} - \bar{Y}_{.j}$
GREG	$\hat{\mu}_i = \bar{Y}_{i.}, z_{ij} = \bar{Y}_{ij} - \bar{Y}_{i.}$
COMM	$z_{ij} = \bar{Y}_{ij}$
SHMM	$z_{ij} = \bar{Y}_{ij} - \hat{\beta}, \hat{\beta} = \bar{Y}_{..} - \sum_{k=1}^{t} \lambda_k \bar{\alpha}_k \bar{\gamma}_k$

Seyedsadr and Cornelius (1992) developed the SHMM model, which is a reparameterization of the Tukey (1949) model for testing non-additivity. The singular vectors for genotypes and environments for the ordered components are called "primary effects" $[\alpha_{i1}, \gamma_{j1}]$, "secondary effects" $[\alpha_{i2}, \gamma_{j2}]$, and so on. The LS solution for $\hat{\beta}$ requires an iterative algorithm, because the solutions for the bilinear terms are the t largest components of the SVD of matrix $\mathbf{Z} = [z_{ij}]$, where, in this case, $z_{ij} = \bar{Y}_{ij} - \hat{\beta}$, but $\hat{\beta} = \bar{Y}_{..} - \sum_{k=1}^{t} \lambda_k \bar{\alpha}_k \bar{\gamma}_k$, where $\bar{\alpha}_k = g^{-1} \Sigma_i \hat{\alpha}_{ik}$, and $\bar{\gamma}_k = e^{-1} \Sigma_j \hat{\gamma}_{jk}$. Thus, **Z** depends on $\hat{\beta}$, but $\hat{\beta}$ depends on the SVD of **Z**. Consequently, the LS solution does not exist in closed form. Moreover, the value of $\hat{\beta}$ changes if the number of bilinear components, t, is changed.

Apparently, the SHMM model was the first LBM that, along with other statistical tools, was used for identifying subsets of genotypes or environments in which genotypic rank changes are negligible (Cornelius *et al.*, 1992, 1993b; Crossa and Cornelius, 1993; Crossa *et al.*, 1993, 1995). Later, the SREG model was suggested as a better model to use for identifying such subsets of environments (Crossa and Cornelius, 1997) (but not for identifying such subsets of genotypes). The SREG model is very appealing for breeders and agronomists because its multiplicative terms contain the main effects of genotypes plus the GEI, making it possible to assess both the general and the specific adaptation of genotypes. Crossa and Cornelius (1997) have used the SREG model for clustering sites without genotypic rank change under heterogeneity of within-site error variances. Yan *et al.* (2000) used the biplot of the first two bilinear components obtained from the SREG model to graphically identify specific 'winner' genotypes in certain subsets of environments.

The GREG model with $t = 1$ is a reparameterization and with $t > 1$ a generalization of the linear regression model of Yates and Cochran (1938), Finlay and Wilkinson (1963) and Eberhart and Russell (1966), except that we replace their estimator of δ_i with an LS solution and further impose orthonormality constraints, as in the AMMI model. Typically, the LS estimators of the γ_{jk} are very highly correlated with the

Finlay–Wilkinson/Eberhart–Russell estimator $\bar{y}_{.j} - \bar{y}_{..}$ of δ_j.

Cornelius and Seyedsadr (1997) defined the general linear–bilinear model (GLBM) as:

$$y_{ij} = \sum_{k=1}^{m} \beta_k x_{kij} + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \bar{\varepsilon}_{ij}$$

where the x_{kij} are known constants and the β_k parameters (regression coefficients) for the linear terms and the λ_k, α_{ik} and γ_{jk} in the bilinear terms are parameters to be estimated (α_{ik} and γ_{jk} subject to the previously defined orthonormality constraints).

In matrix notation, the GLBM can be expressed as:

$$\mathbf{Y} = \sum_{k=1}^{m} \beta_k \mathbf{X}_k + \mathbf{A}\Lambda\mathbf{G}' + \mathbf{E}$$

where $\mathbf{Y} = [\bar{y}_{ij}]$, $\mathbf{X}_k = [x_{kij}]$, $\mathbf{E} = [\bar{\varepsilon}_{ij}]$, $\Lambda = \text{diag}(\lambda_k, k = 1, 2, \ldots, t)$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_t$, $\mathbf{A} = (\alpha_1, \ldots, \alpha_t)$, $\mathbf{G} = (\gamma_1, \ldots, \gamma_t)$ and $\mathbf{A}'\mathbf{A} = \mathbf{G}'\mathbf{G} = \mathbf{I}_t$. Define $\mathbf{Z} = \mathbf{Y} - \sum_{k=1}^{m} \hat{\beta}_{k(t)} \mathbf{X}_k$, where $\hat{\beta}_{k(t)}$ is the LS estimate of β_k when the fitted model contains t bilinear terms ($t \leq \text{rank}(\mathbf{Z})$). Then the first t components of the SVD of $\mathbf{Z}$ provide the LS estimates of parameters in the bilinear terms. An LS solution for the linear effects (the $\hat{\beta}_k$) is given by any solution to the equation:

$$\mathbf{C}\hat{\beta} = \mathbf{T}$$

where the elements of $\hat{\beta}$ are the $\hat{\beta}_k$, the kth element of $\mathbf{C}$ is $C_{kk'} = \text{tr}(\mathbf{X}_k'\mathbf{X}_{k'}) = \sum_i \sum_j x_{kij} x_{k'ij}$ and the kth element of $\mathbf{T}$ is $T_k = \text{tr}\left[\mathbf{X}_k'(\mathbf{Y} - \hat{\mathbf{A}}\hat{\Lambda}\hat{\mathbf{G}}')\right] = \sum_i \sum_j x_{kij}\left(\bar{y}_{ij} - \sum_{\omega=1}^{t} \hat{\lambda}_\omega \hat{\alpha}_{i\omega} \hat{\gamma}_{j\omega}\right)$

Here tr[·] denotes the trace of the (square) matrix given as the argument.

An LBM is said to be 'balanced' (a BLBM) if $\mathbf{Z} = \mathbf{PYQ}$, where $\mathbf{P}$ and $\mathbf{Q}$ are projection matrices free of the bilinear effects. Under this condition T_k reduces to $\sum_i \sum_j x_{kij} \bar{y}_{ij}$ and, thus, in a BLBM, an LS solution for $\hat{\beta}$ ignoring the bilinear effects (i.e. for $t = 0$) is also a solution for $\hat{\beta}$, given any value for $t \leq \text{rank}(\mathbf{Z})$. Provided there are no missing cells, AMMI, CREG, SREG and COMM are BLBMs (COMM actually being without any linear terms at all), but SHMM is not.

In this chapter, we review the use of the SHMM and SREG models for finding clusters of environments with negligible genotypic COI and examine some of the unconstrained and constrained non-COI solutions for finding the 'distance' between pairs of environments. We also summarize results of 'shrinkage' estimators of LBMs developed as analogues of best linear unbiased predictors (BLUPs) and justified by a Bayesian argument, and further show empirical evidence that shrinkage estimators are usually better predictors than the best truncated LBM and sometimes better than BLUPs of a RE2W model with interaction.

SHMM for Assessing COI

Since the early 1990s, theoretical and practical studies have shown the utility of the SHMM model for identifying subsets of environments and genotypes without genotypic rank change (Cornelius *et al.*, 1992; Crossa and Cornelius, 1993; Crossa *et al.*, 1993, 1995, 1996; Abdalla *et al.*, 1997; Trethowan *et al.*, 2001). Cornelius *et al.* (1992), observing results obtained with SHMM_1 (SHMM with one multiplicative term), defined sufficient conditions for the absence of significant genotype 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. This implies that the multiplicative components, beyond the first, are not significantly different from zero.
2. The primary effects of environments, γ_{j1}, are all of like sign.

The SHMM model satisfying the above condition 2 has the following two proportionality properties:

1. Differences between genotypes in any single environment are proportional to genotype differences in any other environment.
2. Differences between environments with respect to the performance of any single genotype are proportional to environmental differences with respect to performance of

any other genotype (but, for environment differences, proportionality constants can be negative).

The second proportionality restriction is irrelevant for the case of genotypic non-COI and is relaxed in the $SREC_1$ model (Crossa and Cornelius, 1997).

When $SHMM_1$-predicted values, $\hat{y}_{ij} = \hat{\beta} + \hat{\lambda}_1\hat{\alpha}_{j1}\hat{\gamma}_{i1}$, are plotted against the primary effects of environments, $\hat{\gamma}_{i1}$, the graph consists of a set of regression lines, one for each genotype, all of which concur (i.e. intersect) at the point $(0, \hat{\beta})$. For a non-COI $SHMM_1$, the $\hat{\gamma}_{i1}$ are all of like sign (or zero) and thus the point of intersection is a point either at the boundary (if one $\hat{\gamma}_{i1} = 0$) or outside (left or right of) the region containing the plotted points.

If the $\hat{\gamma}_{i1}$ have different signs (some positive, some negative), then the point of intersection is within the region containing the plotted points and a complete reversal of rank order of genotypes is displayed on the right, as compared with the left, of the point of intersection. If the intersection point is far outside the region containing the plotted points, then the genotype regression lines appear very nearly parallel, implying that the data are essentially additive (provided that the $SHMM_1$ adequately fits the data).

SHMM clustering of environments with non-COI

Typically, when SHMM is fitted to the entire set of data from an MET, in addition to primary effects, one must include secondary and perhaps even higher-order effects if an adequate fit is to be achieved. The clustering strategy is to 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. In doing this clustering, the measure of 'distance' between two environments is taken as the residual mean square (RMS) after fitting $SHMM_1$ [$RMS(SHMM_1)$] to the data from the two environments subject to a non-COI constraint, namely, that both $\hat{\gamma}_{i1}$ must be either non-positive or non-negative. $RMS(SHMM_1)$ is obtained as $[RSS(SHMM_1)]/f$, where RSS stands for residual sum of squares and f is the d.f., namely, $f = g - 2 + v$, where v is the number of additional constraints imposed to achieve a non-COI solution.

It is a property of SHMM that, if $e < g$, $RSS(SHMM_{e-1}) = RSS(SREC_{e-1})$ in unconstrained LS solutions (Seyedsadr and Cornelius, 1992). Thus, for a subset of $e = 2$ environments, this property provides a closed-form solution for the distance, provided that the two $\hat{\gamma}_{i1}$ values are of like sign. Otherwise a constrained solution must be computed. The constraint, if needed, is imposed by putting $\hat{\gamma}_{i1} = 0$ for one of the two environments and $\hat{\gamma}_{i1} = \pm 1$ for the other environment. Two different methods for doing this have been devised, namely, a constrained LS method and a constrained SVD method. In the constrained LS method, we put $\hat{\gamma}_{i1} = 0$ for the environment that has the smaller value of $\sum(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})^2$ (moreover, this value becomes the distance value) and put $\hat{\gamma}_{i1} = \pm 1$ for the other environment. Other properties of the solution are $\hat{\beta} = \bar{y}_{j\cdot}$ and thus $\hat{y}_{ij} = \hat{\beta} = \bar{y}_{j\cdot}$ for the environment with $\hat{\gamma}_{i1} = 0$ and the quantities $\hat{\lambda}_1\hat{\alpha}_{j1}$ are chosen such that $\hat{y}_{ij} = \hat{\beta} + \hat{\lambda}_1\hat{\alpha}_{j1}\hat{\gamma}_{i1} = \bar{y}_{i\cdot}$ for i, now representing the environment for which $\hat{\gamma}_{i1} = \pm 1$.

The constrained SVD solution chooses $\hat{\beta}$ such that the first right singular vector $[\hat{\gamma}_{i1}]$ of $\mathbf{Z} = [z_{ij}] = [\bar{y}_{i\cdot} - \hat{\beta}]$ is either $(\pm 1, 0)'$ or $(0, \pm 1)'$. A sufficient condition for this is that the two columns of $\mathbf{Z}$ must be orthogonal to one another. There are two solutions for $\hat{\beta}$ that will satisfy this. For either solution, the right singular vectors are $(\pm 1, 0)'$ and $(0, \pm 1)'$, either in that order or in the reverse order. The final choice among the four combinations of one of the two possible $\hat{\beta}$ values and one of the two possible environments for which to put $\hat{\gamma}_{i1} = 0$ will be the combination for which the quantity $\sum(\bar{y}_{i\cdot} - \hat{\beta})^2$ is smallest. For further details, we refer the reader to Crossa *et al.* (1996). The residual

d.f. for either the constrained LS or constrained SVD solution fitted to data from two environments are $g-1$. We doubt that the choice of method for computing constrained solutions will ever be a critically important issue in clustering environments or genotypes into non-COI groups.

After the distances for all possible pairs of environments have been computed, a dendrogram is constructed using the complete linkage (furthest neighbour) clustering method. The final step is to analyse the subsets of data for each of the clusters suggested by branches of the dendrogram for adequacy of fit of $SHMM_1$ (constrained if necessary to obtain a non-COI solution).

A constrained SVD non-COI $SHMM_1$ solution for a subset containing more than two environments can usually be computed by iteratively alternating between computation of $\hat{\beta}=\dfrac{\sum_i \hat{\alpha}_i \bar{y}_{ih}}{\sum_i \hat{\alpha}_{ih}}$ and computation of the $\hat{\alpha}_{ih}$ as elements of the first left singular vector of $\mathbf{Z}=[z_{ij}]=\left[\bar{y}_{ij}-\hat{\beta}\right]$, where the subscript h denotes the environment to have its primary effect ($\hat{\gamma}_{hj}$) put equal to zero and the $\bar{y}_{ij}$ used in computing $\mathbf{Z}$ are only those for the particular subset being analysed. If what at first seems to be the most reasonable choice for environment h fails to give a non-COI solution (which occurs if the non-zero $\hat{\gamma}_{jh}$ in the constrained solution are not all of like sign), another choice for environment h may be tried. We doubt if it is possible to find a solution for $\hat{\beta}$ that will simultaneously constrain primary effects of more than one environment. If such a solution appears to be necessary, computation of a constrained LS solution may be mandatory. A Newton–Raphson algorithm for computing a constrained LS solution for any number of environments to have primary effects put equal to zero can be found in Crossa *et al.* (1996). Despite the more complicated algorithm and underlying mathematics, we prefer constrained LS solutions to constrained SVD solutions.

For a set (or subset) containing g genotypes and e environments, the d.f. of RSS ($SHMM_1$) is $ge-g-e$ in an unconstrained solution, $ge-g-e+1=(g-1)(e-1)$ in a constrained SVD solution and $ge-g-e+v$ in a constrained LS solution, where v is the number of environments with their $\hat{\gamma}_{jh}=0$ in the constrained LS solution.

The SREG Model and its Relationship with the COI

It has been shown that SREG with one multiplicative term ($SREG_1$) is a viable alternative to $SHMM_1$ as a model for identifying groups of environments without genotype COI, because $SREG_1$, like $SHMM_1$, also displays proportionality of genotype differences in different environments, but, unlike $SHMM_1$, $SREG_1$ does not impose proportionality of environmental differences with respect to performance of genotypes (Crossa and Cornelius, 1997). Proportionality of environmental differences with respect to different genotypes is not relevant to the issue of genotype COI and SREG's relaxation of these constraints may allow larger non-COI clusters to be obtained. Furthermore, SREG can be quite satisfactorily used to deal with heterogeneity of within-environment error variances by the simple device of rescaling the $\bar{y}_{ij}$ by dividing by $\sqrt{\dfrac{s_j^2}{n}}$, where s_j^2 is the error mean square within the jth environment. (So also can SHMM, but the result has the unpalatable property that SHMM fitted to the scaled data is no longer a SHMM when back-transformed to the original scale (Crossa and Cornelius, 1997).)

In the $SREG_1$ model, it is the deviations of genotype yields ($\bar{y}_{ij}$) from environment means $\bar{y}_{.j}$ that are modelled by the bilinear term. The fitted bilinear effects, $\hat{\lambda}_1\hat{\alpha}_{i1}\hat{\gamma}_{j1}$, can be plotted as a set of regression lines, one for each genotype, with the $\hat{\gamma}_{j1}$ as regressor variable and with zero intercepts. Because these regression lines all intersect at the zero point on the $\hat{\gamma}_{j1}$ scale, the graph, like the graph of $SHMM_1$, does not display genotype COI within the region of the plotted points $(\hat{\lambda}_1\hat{\alpha}_{i1}\hat{\gamma}_{j1}, \hat{\gamma}_{j1})$ if, and only if, the $\hat{\gamma}_{j1}$ are either all non-negative or all non-positive. Addition of the environment mean to the plotted

ordinates $(\hat{\lambda}_1\hat{\alpha}_i, \hat{\gamma}_{ji})$ gives the $SREC_1$ predicted response $\hat{y}_{ji} = \bar{y}_{.j} + \hat{\lambda}_1\hat{\alpha}_i\hat{\gamma}_{ji}$. If the $\hat{y}_{ji}$ are plotted against the $\hat{\gamma}_{ji}$, the plotted points for any given genotype no longer fall on a straight line, but, if plotted points for adjacent values of $\hat{\gamma}_{ji}$ are connected with line segments, the figure displays an overlaid set of broken-line graphs, one for each genotype, which, although they are not straight lines, nevertheless display no genotype COI within the region of plotted points.

For $SREC_1$ clustering of environments, the measure of distance between two environments is $RMS(SREC_1)$. For a subset of data with only two environments, $RSS(SREC_1) = RSS(SHMM_1)$ for both unconstrained and constrained LS non-COI solutions. Consequently, provided that constrained LS solutions are used for non-COI solutions (when needed), dendrograms for SREC and SHMM clustering are identical. Thus, subsets that the dendrogram suggests as groups to be evaluated for acceptability of fit of the one-term model are the same subsets whether $SREC_1$ or $SHMM_1$ is used, but the less parsimonious $SREC_1$ may sometimes give an acceptable fit to a subset to which $SHMM_1$ does not acceptably fit.

Unlike $SHMM_1$, the constrained LS non-COI $SREC_1$ solution for a subset containing more than two environments exists in closed form (Crossa and Cornelius, 1997). For a subset consisting of sub-subset S_1 containing e_1 environments that are to have their $\hat{\gamma}_{ji} = 0$ and sub-subset S_2 containing e_2 environments that are to have their $\hat{\gamma}_{ji}$ unconstrained, the constrained LS $SREC_1$ solution is to put $\hat{\mu}_j = \bar{y}_{.j}$ for all of the environments in the subset, but obtain $\hat{\lambda}_1$, $\hat{\alpha}_i$ and the nonzero $\hat{\gamma}_{ji}$ as the first component of the SVD of $\mathbf{Z}_2 = [z_{ij}] = [\bar{y}_{ij} - \bar{y}_{.j}]$, where the $\bar{y}_{ij}$ and $\bar{y}_{.j}$ used in computing $\mathbf{Z}_2$ are only those from environments in sub-subset S_2.

We have recently developed a constrained SVD solution for $SREC_1$ by putting $\hat{\mu}_j = \bar{y}_{.j} + \hat{\beta}$, where $\hat{\beta}$ is a constant chosen to force a particular $\hat{\gamma}_{ji} = 0$ (P.L. Cornelius and J. Crossa, unpublished result). The usefulness of such constrained SVD $SREC_1$ solutions for the SREC clustering problem has not been evaluated.

For a set (or subset) containing g genotypes and e environments, the d.f. of RSS $(SREC_1)$ is $ge - g - 2e + 2 = (g-2)(e-1)$ in an unconstrained solution, $(g-2)(e-1)+1$ in the above described constrained SVD $SREC_1$ solution and $(g-2)(e-1)+v$ in a constrained LS solution, where v is the number of environments with their $\hat{\gamma}_{ji} = 0$ in the constrained LS solution. Note that $v = e_1$, where e_1 is as previously defined.

SHMM Clustering of Genotypes with Non-COI

Cornelius *et al.* (1993b) used SHMM clustering to group 41 winter-wheat (*Triticum aestivum* L.) genotypes into non-COI clusters. The MET included seven environments. Thirty-five of the genotypes were grouped into nine clusters, leaving six genotypes unclustered because the procedure did not enter them into any cluster to which a non-COI $SHMM_1$ would give a satisfactory fit. Constrained non-COI solutions, when needed, were computed using constrained SVD solutions. Other examples of SHMM clustering of genotypes have been reported by Crossa *et al.* (1996) and Abdalla *et al.* (1997).

Crossa *et al.* (1996) compared constrained LS and constrained SVD solutions when constrained solutions were needed and found, for the particular example, that choice of method for computing constrained solutions made no difference with respect to the dendrogram or final acceptable clusters obtained. To our knowledge, this is the only published example in which consequences of the choice of method for computing constrained non-COI solutions has been studied.

SHMM clustering of genotypes is essentially by the same strategy as for clustering environments. The distance between two genotypes is defined as $RMS(SHMM_1)$, using a constrained solution, if necessary, when $SHMM_1$ is fitted to the subset of data deriving from only those two genotypes evaluated in the entire set of environments. Note that it is still the $\hat{\gamma}_{ji}$ (and not the $\hat{\alpha}_i$) that must be

either all non-positive or all non-negative to have a non-COI solution.

The unconstrained SHMM$_1$ solution for RSS(SHMM$_1$) can be obtained in closed form using the result RSS(SHMM$_1$) = RSS(GREC$_1$) if the set of data being analysed contains only two genotypes. Because when clustering genotypes the number of environments will always exceed two, constrained non-COI solutions will not exist in closed form. They can be computed as previously described for subsets containing more than two environments in the context of SHMM clustering of environments. The residual d.f. is $e-2$ for an unconstrained solution, $e-1$ for a constrained SVD solution and $e-2+v$ for a constrained LS solution, where v is the number of environments with their $\hat{\gamma}_{jk}=0$ in the constrained LS solution. Further details can be found in Crossa *et al.* (1996).

Use of SREC as a model for identifying groups of genotypes without significant genotype COI is not practical because, for a pair of genotypes, the matrix $\mathbf{Z}=[z_{ij}]=[\bar{y}_{ij}-\bar{y}_{ij}]$ is of rank one and an unconstrained SREC$_1$ will fit the values exactly, resulting in the distance RSS(SREC$_1$) = 0. Thus, it is only for pairs of genotypes for which a constrained non-COI solution is needed that a non-zero distance will be obtained. Consequently, SREG$_1$ clustering of genotypes will, typically, not provide a unique starting-point for the clustering. When clustering genotypes, the d.f. of RSS (SHMM$_1$) in unconstrained and constrained solutions for a subset containing more than two genotypes will be by the same formulae as previously given for subsets containing more than two environments in the context of clustering environments.

Tests for Lack of Fit of SHMM$_1$ and SREG$_1$

Inadequacy of SHMM$_1$ or SREG$_1$ (either of these constrained, if necessary) for modelling data from a subset of environments and/or genotypes may be tested statistically using the F_R and/or F_{CH} (F_{CH1} or F_{CH2}) tests.

The F_R test (Cornelius *et al.*, 1992, 1996) of RMS(SHMM$_1$) or RMS(SREG$_1$) is from the fit of SHMM$_1$ or SREG$_1$ (constrained if necessary) to the subset against the MET's pooled error mean square. The d.f. of RMS(SHMM$_1$) or RMS(SREG$_1$) are as given earlier in this chapter.

The F_{CH} tests are sequential tests of the bilinear components. Significance of one or more components beyond the primary effects implies inadequacy of inclusion of only one bilinear term. Letting SS_k denote the sequential sum of squares (on an observation basis) due to the kth bilinear term, the F_{CH2} test is constructed as:

$$F_{CH2}=\frac{SS_k}{u_{1k}s^2}$$

where s^2 is the pooled error mean square and $u_{1k}=E(SS_k/\sigma^2 \mid \lambda_k=0,\ \lambda_{k-1}$ is large). The denominator d.f. are the pooled error d.f., and the numerator d.f. are approximated as $2u_{1k}^2/u_{2k}^2$, where $u_{2k}^2=V(SS_k/\sigma^2 \mid \lambda_k=0,$ λ_{k-1} is large).

A function that will closely approximate u_{1k} and u_{2k} for use in F_{CH} tests of bilinear components in SREC is given by Liu and Cornelius (2001). For doing so, always put u_{1k} and u_{2k} equal to their approximating functions for $E(\hat{\theta}_i)$ and $SD(\hat{\theta}_i)$, respectively, with their r and c defined as $r=\max(g-1,\ e)-k+1$ and $c=\max(g-1,$ $e)-k+1$. The approximating functions are valid for $r\le 199$ and $c\le 149$. For AMMI, SREC, GREC and COMM, the approximating functions given by Liu and Cornelius (2001) supersede the functions previously given by Cornelius *et al.* (1996).

For tests of bilinear components in SHMM$_1$ use the SHMM approximating functions for u_{1k} and u_{2k} given in the appendix of Cornelius *et al.* (1996) if $\max(g, e)<100$ and $\min(g, e)<20$. For cases that violate either of these bounds, use the formulas for $E(\hat{\theta}_i)$ and $SD(\hat{\theta}_i)$ from Liu and Cornelius (2001), with r and c defined as $r=\max(g, e)-k+1$ and $c=\max(g, e)-k+1$. Use of functions given by Liu and Cornelius (2001) for SHMM analysis tests the SHMM sequential bilinear terms as if they were bilinear terms in

COMM. This should give sufficiently accurate results for SHMM if r or c is large.

The F_{GH} derives from a method of moments approximation of the distribution of the quantity $1 + [SS_H/(\text{pooled error SS})]$ as the reciprocal of a beta random variable. The P value can be computed directly from the approximate beta distribution, but, because plant breeders will generally find the value of an F statistic more interpretable than the value of a beta statistic, we routinely transform the beta statistic to an F statistic (with d.f. equal to twice the values of the beta distribution). For details, see Cornelius *et al.* (1992, 1996) and Cornelius (1993). If the pooled error d.f. are large, as they generally are in a MET, P values for F_{GH} and F_{GH} are typically almost identical and thus there is ordinarily no need to compute both.

Example of the SHMM and SREG Clustering of Environments with Non-COI in Maize MET

The data come from an international maize (*Zea mays* L.) MET with nine genotypes ($g = 9$) evaluated in a randomized complete block design with four replicates in each of 20 environments ($e = 20$). The SHMM and SREG analyses showed that the first three components were statistically different from zero ($P < 0.05$) by the F_{GH} test (Table 20.1, results for all environments). Since the second and third components were statistically significant, $SHMM_1$ will not adequately fit the data from the entire set of environments. Moreover, even the fitted $SHMM_1$ (unconstrained) has its point of concurrence within the region containing the plotted points and thus the fitted $SHMM_1$ itself displays genotype COI. This is observed in Fig. 20.1, where three environments have $\hat{\gamma}_{j1} < 0$ and all others have $\hat{\gamma}_{j1} > 0$. Genotype 8 performed worst in the environment with the largest primary effect, but was one of the best two genotypes in the environment with the smallest (most negative) primary effect.

Figure 20.2 depicts the dendrogram of the 20 environments when RMS($SHMM_1$) was used as distance measurement and clustering was by the complete linkage (furthest neighbour) method. The dichotomous

Table 20.1. Probability values (P) for the F_R and F_{GH1} tests for the secondary and tertiary effects of the SHMM and SREG models for subsets of environments suggested by the dichotomous splitting of the dendrogram of Fig. 20.2.

		F_R		F_{GH1}	
Environments	Model form	Secondary effect	Tertiary effect	Secondary effect	Tertiary effect
All	SHMM	0.0300	0.0244	0.0000	0.0037
	SREG	0.0016	0.0685	0.0042	0.0322
{1, 3, 8, 10}	SHMM	0.0001	0.4757	0.0000	0.2532
	SREG	0.0552	0.4383	0.0481	0.2155
{1, 3, 10}	SHMM	0.6852	0.9884	0.4560	0.9884
	SREG	0.6585	0.9884	0.1245	0.4059
{2, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20}	SHMM	0.0000	0.0418	0.0000	0.0076
	SREG	0.0027	0.1042	0.0040	0.0287
{2, 6, 9, 12, 13, 18}	SHMM	0.1106	0.3641	0.2229	0.4686
	SREG	0.2134	0.7129	0.1455	0.6999
{4, 5, 7, 11, 14, 15, 16, 17, 19, 20}	SHMM	0.0254	0.7313	0.0004	0.4335
	SREG	0.2481	0.7067	0.1982	0.4839
{4, 5, 11, 14, 15, 16}	SHMM	0.3524	0.8925	0.1615	0.7020
	SREG	0.8113	0.9186	0.8642	0.8521
{7, 17, 19, 20}	SHMM	0.5342	0.9748	0.2717	0.9850
	SREG	0.4529	0.9581	0.2172	0.9717

F_R test of RMS($SHMM_J$) or RMS($SREG_J$) against the pooled error mean square. F_{GH1} test is a sequential test of the bilinear components.

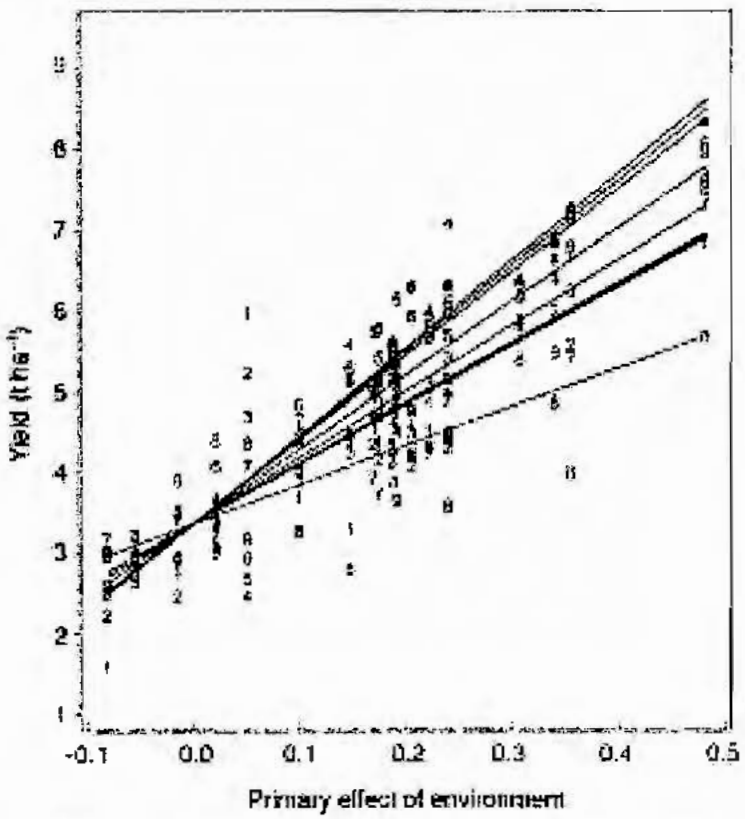

Fig. 20.1. SHMM$_1$ model fitted by least squares (unconstrained) to nine genotypes and 20 environments ($\hat{\beta}$=3.41). The scatter points are empirical cell means plotted using the digits identifying the genotype as plotting symbols and the regression lines, one for each genotype, plot the SHMM$_1$ predicted cell means. The regression lines from top to bottom in the region to the right of the point of concurrence $(0, \hat{\beta})$ are genotypes 4, 5, 6, 3, 2, 9, 7, 1 and 8. This rank order for SHMM-predicted yields is completely reversed to the left of the point of concurrence.

splitting suggested by Crossa *et al.* (1990) for finding subsets of environments with genotypic non-COI consisted of fitting SHMM$_1$ to subsets defined by the branches of the dendrogram, starting at the first split of the entire set of data into the two subsets of environments, [1, 3, 8, 10] and [2, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. For the first of these subsets, the lack of fit of SREG$_1$ is marginally significant ($P = 0.0552$) when tested by the F_R (Cornelius *et al.*, 1996) test, and the SREG secondary effect is significant at $P < 0.05$ when tested by the F_{GH1} test (Table 20.1). For the latter of these two subsets, there is highly significant lack of fit of both SHMM$_1$ and SREG$_1$, detected by both F_R and F_{GH1} tests (Table 20.1). Continuing with the dichotomous splitting, the adequacy of SHMM$_1$ for fitting subsets [1, 3, 10], [2, 6, 9, 12, 13, 18] and [4, 5, 7, 11, 14, 15, 16, 17, 19, 20] is tested. According to the F_R and the F_{GH1} tests, SHMM$_1$ is adequate for fitting the first two subsets, but not for the last one. However, SREG$_1$ did adequately fit all three subsets. Thus, we have here an example of a subset that is acceptably modelled by a non-COI SREG$_1$, but not by a non-COI SHMM$_1$.

In Fig. 20.3, the consistent response of the nine genotypes across the ten environments of subset [4, 5, 7, 11, 14, 15, 16, 17, 19, 20] is clearly depicted through the overlaid broken line SREG$_1$ graphs that do not cross over. The residuals for the two models, both with unconstrained solutions, were RMS (SHMM$_1$) = RSS(SHMM$_1$)/$(ge - g - e)$ = 14,006,110/71 = 197,269 and RMS(SREG$_1$) = RSS(SREG$_1$)/$(g - 2)(e - 1)$ = 9,795,021/63 = 155,476. Further splitting of [4, 5, 7, 11, 14, 15, 16, 17, 19, 20] gives subsets [4, 5, 11, 14, 15, 16] and [7, 17, 19, 20], both of which can be adequately modelled by either

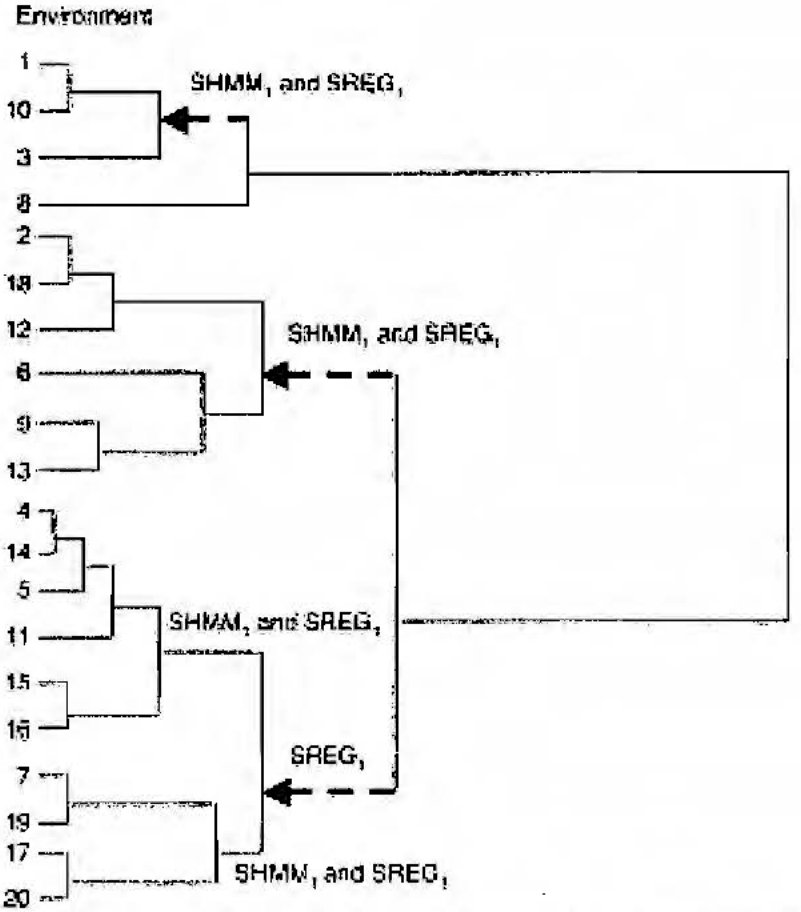

Fig. 20.2. Dendrogram of SHMM and SREG clustering of 20 environments using RMS(SHMM$_1$) = RMS(SREG$_1$) as distance measurement. Arrows denote the subsets of environments where SREG$_1$ gave an adequate fit. The subset indicated by the lower arrow on the figure had to be subdivided once more in order for SHMM$_1$ to give an adequate fit.

non-COI SHMM$_1$ or non-COI SREG$_1$ (Table 20.1).

The ten highest values of RMS(SHMM$_1$) (the distance) for pairs of environments with unconstrained and constrained LS and SVD non-COI solutions are shown in Table 20.2. As expected, based on the results obtained from the SHMM$_1$ and SREG$_1$ clustering (Fig. 20.2), environment 8 is the most frequently occurring environment in pairs of environments with large distance values.

To illustrate constrained and unconstrained solutions for a pair of environments needing a constrained solution for their distance, SHMM$_1$ fitted without constraint to the data from environment 8 ($\bar{y}_{.8}$ = 4027, $\hat{\gamma}_{8,1}$ = −0.7286) and environment 11 ($\bar{y}_{.11}$ = 5307, $\hat{\gamma}_{11,1}$ = 0.6849) has its point of concurrence within the region containing the plotted data points (Fig. 20.4). In the constrained LS solution (Fig. 20.5), the environment that had a positive primary effect ($\hat{\gamma}_{11,1}$ = 0.6849) in the unconstrained solution has its primary effect put equal to zero ($\hat{\gamma}_{11,1}$ = 0), thus moving the point of concurrence to the right boundary of the region containing the plotted points. In this constrained LS solution, $\hat{\gamma}_{8,1}$ = −1.

Both the SHMM and SREG clustering of environments have been extensively used for finding associations between international testing environments used by the International Maize and Wheat Improvement Center (CIMMYT) for maize, bread (*T. aestivum* L.) and durum (*Triticum turgidum* var. *durum*) wheat and triticale (*Triticosecale* Wittm.) METs. Abdalla *et al.* (1997) used SHMM clustering of environments and genotypes and found that durum-wheat genotypes with similar genetic backgrounds

formed non-COI clusters, but COI more frequently occurred with genotypes derived from different genetic backgrounds, especially those with different levels of resistance to specific diseases. Consequently, genotypes from diverse genetic

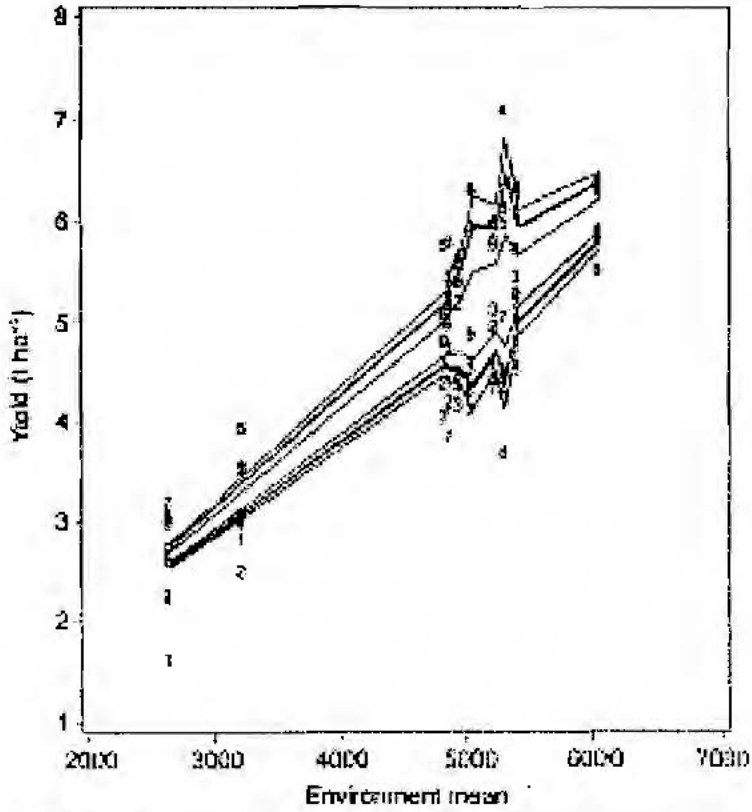

Fig. 20.3. SREG$_1$ model fitted to nine genotypes and a subset of 10 environments. The rank order of genotypes with respect to the overlaid broken-line graphs is 6, 4, 5, 7, 9, 3, 1, 2, 8.

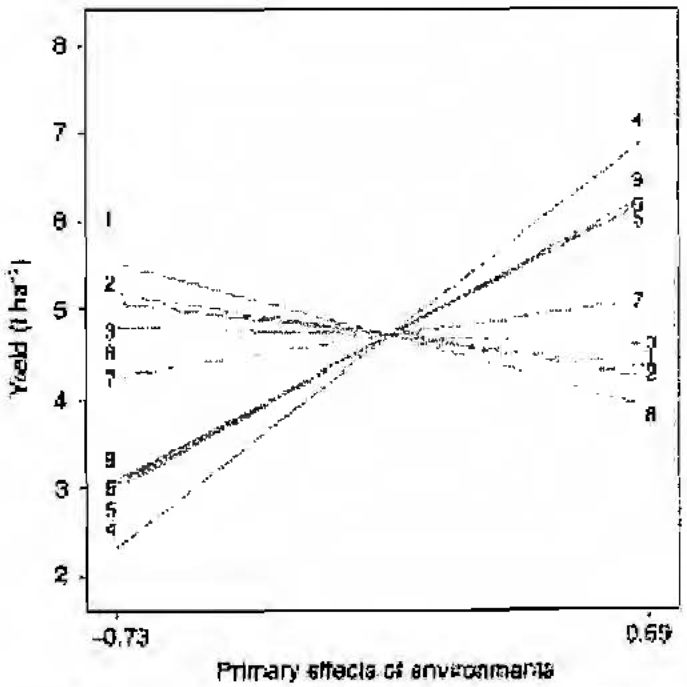

Fig. 20.4. Unconstrained SHMM$_1$ solution for nine genotypes in two environments (8 and 11).

backgrounds tended to cluster into different non-COI groups. Recently, Trethowan *et al.* (2001) used the SHMM and SREC clustering of environments to study long-term associations between international sites for a drought-tolerant bread-wheat MET. Results of this study demonstrated the usefulness of this approach for identifying key testing environments around the world.

Shrinkage Estimates of Linear–Bilinear Models Analogous to BLUPs

Simulation studies of Cornelius (1993) and Cornelius *et al.* (1996) for the AMMI model showed that the interaction mean squared errors (IMSE), $\sum_i \sum_j \left(\sum_{k=1}^{p} \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk} - \sum_{k=1}^{p} \lambda_k \alpha_{ik} \gamma_{jk}\right)^2$, where $p = \text{rank}(\mathbf{Z})$, can be reduced if the LS estimate of λ_k, i.e. $\hat{\lambda}_k$, is replaced by a shrinkage estimate of the form $S_k\hat{\lambda}_k$. The authors found that the IMSE of the shrinkage estimates were smaller than the IMSE of the best AMMI truncated model. These shrinkage estimators are of the form *signal* divided by (*signal* + *noise*) and are similar to the functions of variance components that occur in computation of empirical BLUPs of cell means in a RE2W model and can be justified by a Bayesian argument.

The conventional RE2W model with interactions for the mean of the ith genotype in the jth environment is given in Equation 20.1 with the τ_i, δ_j and $(\tau\delta)_{ij}$ considered random. Under normality, independence and a balanced data set, it is easy to compute empirical BLUPs of the realized performance levels, $\mu_{ij} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij}$, of the genotypes in the environments where they were tested (i.e. empirical BLUPs of cell means). The BLUP estimates of the main effects of τ_i, δ_j and $(\tau\delta)_{ij}$ for a balanced data set are:

$$\text{BLUP}(\tau_i) = \frac{n e \sigma_\tau^2}{\sigma^2 + n\sigma_{\tau\delta}^2 + ne\sigma_\tau^2}(\bar{Y}_{i.} - \bar{Y}_{..})$$

$$\text{BLUP}(\delta_j) = \frac{n g \sigma_\delta^2}{\sigma^2 + n\sigma_{\tau\delta}^2 + ng\sigma_\delta^2}(\bar{Y}_{.j} - \bar{Y}_{..})$$

$$\text{BLUP}\left[(\tau\delta)_{ij}\right] = n\sigma_{\tau\delta}^2\left[\frac{(\bar{Y}_{i.} - \bar{Y}_{..})}{\sigma^2 + n\sigma_{\tau\delta}^2 + ne\sigma_\tau^2} + \frac{(\bar{Y}_{.j} - \bar{Y}_{..})}{\sigma^2 + n\sigma_{\tau\delta}^2 + ng\sigma_\delta^2} + \frac{(\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} - \bar{Y}_{..})}{\sigma^2 + n\sigma_{\tau\delta}^2}\right]$$

Not only do the ordinary LS estimates of the main effects of genotypes $(\bar{Y}_{i.} - \bar{Y}_{..})$ and environments $(\bar{Y}_{.j} - \bar{Y}_{..})$ contribute to the BLUP of the ith genotypic main effect and of the jth environmental main effect, respectively, but each also contributes to the BLUP of the CEL Then, for $\hat{\mu} = \bar{Y}_{..}$, the BLUP of the cell mean is given by:

$$\hat{\mu} + \text{BLUP}(\tau_i) + \text{BLUP}(\delta_j) + \text{BLUP}\left[(\tau\delta)_{ij}\right] = \hat{\mu} + \frac{n\sigma_{\tau\delta}^2 + ne\sigma_\tau^2}{\sigma^2 + n\sigma_{\tau\delta}^2 + ne\sigma_\tau^2}(\bar{Y}_{i.} - \bar{Y}_{..}) +$$

Table 20.2. The ten largest distances (RMS ($SHHM_1$)) for unconstrained and constrained SVD and LS solutions. The degrees of freedom for the unconstrained solutions are the number of genotypes minus two, whereas for the constrained solution they are the number of genotypes minus one.

Pair of environments		Degrees of freedom	Residual mean square
SVD non-COI $SHMM_1$ solution			
5	8	8	534,929.83
3	4	8	538,867.47
8	17	8	546,730.47
8	12	8	621,662.28
3	11	8	630,314.30
4	8	8	636,836.50
8	16	8	1,118,858.50
8	13	8	1,197,429.00
8	18	8	1,278,760.50
8	11	8	1,687,992.90
LS non-COI $SHMM_1$ solution			
3	17	7	445,676.58
15	18	7	473,768.72
4	8	8	474,787.63
3	8	7	486,411.11
3	15	7	481,284.81
8	12	8	600,516.22
8	15	8	809,543.62
8	18	8	861,644.18
8	13	8	1,197,120.80
8	11	8	1,327,851.00

SVD, singular value decomposition constrained solution; LS, least-squares constrained solution.

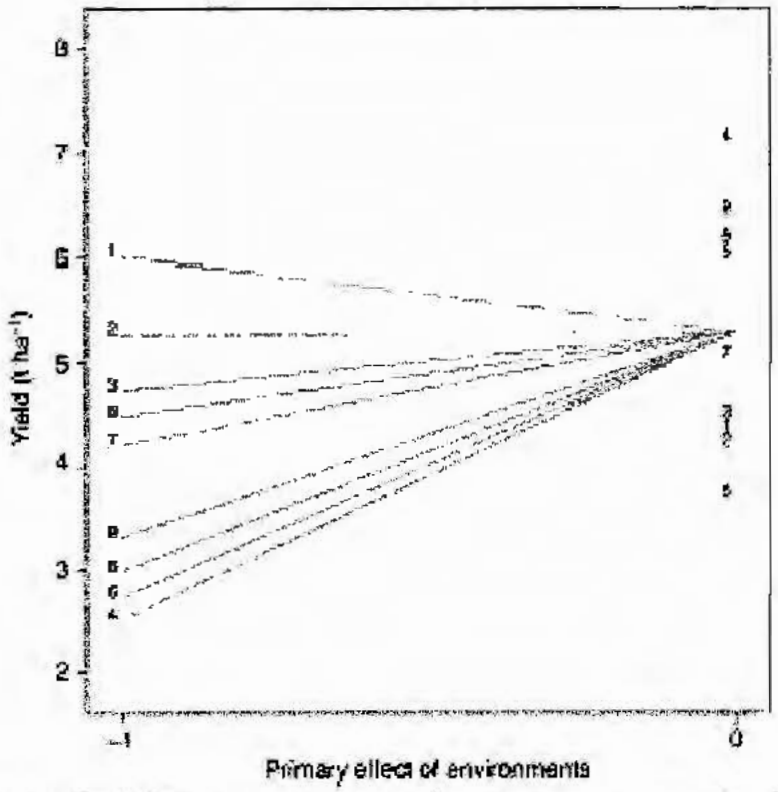

Fig. 20.5. Constrained LS SHMM$_1$ solution for nine genotypes in two environments (8 and 11).

$$\frac{n\sigma^2_{\alpha\beta}+n\beta\sigma^2_{\alpha}}{\sigma^2+n\sigma^2_{\alpha\beta}+n\beta\sigma^2_{\alpha}}(\bar{Y}_{i.}-\bar{Y}_{..})+\frac{n\sigma^2_{\alpha\beta}}{\sigma^2+n\sigma^2_{\alpha\beta}}$$
$$(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})=\bar{Y}_{...}+S_{\alpha}(\bar{Y}_{i..}-\bar{Y}_{...})+$$
$$S_{\beta}(\bar{Y}_{.j.}-\bar{Y}_{...})+S_{\alpha\beta}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}-\bar{Y}_{...})$$

In computing empirical BLUPs of cell means, the functions of the estimated variance components that multiply the LS estimates of the main effects of genotypes, environments and the GEI, namely $(\bar{Y}_{i..}-\bar{Y}_{...})$, $(\bar{Y}_{.j.}-\bar{Y}_{...})$ and $(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}-\bar{Y}_{...})$, respectively, are 'shrinkage' factors that can be estimated as:

$$S_{\alpha}=\frac{n\hat{\sigma}^2_{\alpha\beta}+n\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2+n\hat{\sigma}^2_{\alpha\beta}+n\hat{\sigma}^2_{\alpha}}=1-\frac{1}{F_{\alpha}}=$$
$$1-\frac{1}{\frac{\text{MS(Genotype)}}{\text{MS(Error)}}}=1-\frac{\text{MS(Error)}}{\text{MS(Genotype)}},$$

$$S_{\beta}=\frac{n\hat{\sigma}^2_{\beta}+n\beta\hat{\sigma}^2_{\beta}}{\hat{\sigma}^2+n\hat{\sigma}^2_{\alpha\beta}+n\beta\hat{\sigma}^2_{\beta}}=1-\frac{1}{F_{\beta}}=$$
$$1-\frac{1}{\frac{\text{MS(Environment)}}{\text{MS(Error)}}}=$$
$$1-\frac{\text{MS(Error)}}{\text{MS(Environment)}}$$

and

$$S_{\alpha\beta}=\frac{n\hat{\sigma}^2_{\alpha\beta}}{\hat{\sigma}^2+n\hat{\sigma}^2_{\alpha\beta}}=1-\frac{1}{F_{\alpha\beta}}=$$
$$1-\frac{1}{\frac{\text{MS(Genotype}\times\text{Environment)}}{\text{MS(Error)}}}=$$
$$1-\frac{\text{MS(Error)}}{\text{MS(Genotype}\times\text{Environment)}}$$

where MS = mean square.

For a saturated AMMI model, the estimates of the interaction parameters are such that $\sum_{k=1}^{t}\lambda_k\alpha_{ik}\gamma_{jk}=\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...}$, i.e. the BLUP of a cell mean is a shrinkage estimate of AMMI. It is reasonable to suppose, however, that an optimum strategy for obtaining shrinkage estimates of AMMI, COMM, SREC, CREC and SHMM should include different shrinkage factors for the different bilinear components.

Cornelius and Crossa (1995, 1999) and Cornelius *et al.* (1993a, 1996) used a procedure of shrinking each bilinear component by an estimate of the *signal* to (*signal* + *noise*) ratio of its eigenvalue, i.e. the kth bilinear component $\hat{\lambda}_k\hat{\alpha}_{ik}\hat{\gamma}_{jk}$ is shrunk as $S_k\hat{\lambda}_k\hat{\alpha}_{ik}\hat{\gamma}_{jk}$. Then the sum of shrunken bilinear terms is $\sum_k S_k\hat{\lambda}_k\hat{\alpha}_{ik}\hat{\gamma}_{jk}$, where $\hat{\lambda}_k$, $\hat{\alpha}_{ik}$ and $\hat{\gamma}_{jk}$ are the corresponding LS estimates. The shrinkage factor S_k is computed as:

$$S_k = \left(\hat{\lambda}_k^2 - \ddot{u}_k s^2/n\right)/\hat{\lambda}_k^2 = 1 - \frac{1}{\ddot{F}_k}$$

where s^2 is the error mean square, $\ddot{u}_k$ is a value equal to, or that estimates, the expected error variance absorption by the kth bilinear term, i.e. $nE\left[\left(\hat{\lambda}_k^2 - \lambda_k^2\right)/\sigma^2\right]$, and $\ddot{F}_k = n\hat{\lambda}_k^2/\ddot{u}_k s^2$, provided that $\ddot{F}_k > 1$ and therefore $S_k > 0$; otherwise, put $S_k = 0$. The $\ddot{F}_k$ is a statistic similar in structure to the F_{GH2} statistic that Cornelius *et al.* (1996) used to test the null hypothesis $H_0: \lambda_k = 0$, but F_{GH2} replaces $\ddot{u}_k$ with u_k, the latter defined as the conditional expectation $nE\left[\left(\hat{\lambda}_k^2/\sigma^2\right)|\lambda_k^2 = 0\right]$, i.e. $\ddot{u}_k$ is an estimate of $nE\left[\left(\hat{\lambda}_k^2|\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots\right)\right]$. The value of $\ddot{u}_k$ in S_k is appropriate for the alternative hypothesis $H_1: \lambda_k > 0$, but not for the null hypothesis. The above-described shrinkage estimators are appropriate for any of the balanced LBMs – AMMI, SREC, CREC and CDMM. See Cornelius and Crossa (1999) for a mathematical justification for the above-described shrinkage estimators and for details of a scheme for computing shrinkage estimates of SHMM.

In practice, we obtain initial values of $\ddot{u}_k$ as the number of independent parameters in the kth bilinear component (i.e. number of parameters minus number of constraints). These initial $\ddot{u}_k$ are used to obtain an initial set of shrinkage estimates of the λ_k. Then, these initial shrinkage estimates are used as supposed 'true' values of the λ_k in a simulation (parametric bootstrap) scheme to obtain more accurate values of the $\ddot{u}_k$, which, in turn, are used to obtain a new set of shrinkage estimates of the λ_k. The scheme can be iterated as often as desired. The scheme has been observed to move the shrinkage estimates into a rather stable neighbourhood in about five iterations. For our scheme for computing shrinkage estimates of SHMM, see Cornelius and Crossa (1999).

Prediction accuracy of the shrinkage estimates of linear–bilinear models

It has been suggested that shrinkage estimates of LBMs will provide better estimates of the realized values of the cell means (μ_{ij}) than will the empirical cell means or LS solutions of parsimonious models with the number of multiplicative terms chosen by cross-validation or any test of statistical significance (Cornelius *et al.*, 1993a, 1996; Cornelius and Crossa, 1995). Cornelius and Crossa (1999) analysed five MET data sets using random data splitting and cross-validation. They evaluated the predictive accuracy of the shrinkage estimates of the LBMs and compared them with: (i) the best LS fitted truncated LBM; (ii) the empirical BLUPs of the cell means based on the RE2W model; and (iii) the empirical cell means. The root mean square predictive difference (RMSPD) was computed for judging the best predictive model, i.e. the one having the lowest RMSPD. The authors used data adjusted by replicate differences within environments to reduce the noise in the modelling and validation data that otherwise occurs as a consequence of ignoring block differences when randomly splitting the data.

Results showed that the worst predictors of the genotypic performance were the empirical cell means (Table 20.3). In all five experiments, shrinkage estimates of LBMs were better predictors than the best truncated model fitted by LS and, except in one experiment, also better than the BLUPs of the cell means. These results suggest that shrinkage estimates of LBMs eliminate the need for testing hypotheses and cross-validation to select an optimum number of bilinear terms. Results also indicate that predictive accuracy differs little among the

five model forms if shrinkage estimators are used.

Experimental data set 3 of Cornelius and Crossa (1999) is the maize MET data, with nine genotypes evaluated in 20 environments, used earlier in this chapter to illustrate SHMM and SREG clustering of environments. For this MET, clearly

Table 20.3. RMSPD (kg ha^{-1}) values for the best truncated least-squares fitted model and shrinkage estimates for linear–bilinear model forms AMMI, GREG, SREG, COMM and SHMM, for the BLUPs of cell means and for empirical cell means, all obtained by cross-validation in five MET data sets (Cornelius and Crossa, 1999).

Model form*	Truncated	Shrinkage	BLUP	Empirical cell mean
Experiment 1				
$AMMI_4$	637.45	627.65	–	–
$COMM_5$	636.40	626.92	–	–
$SREG_5$	641.12	629.04	–	–
$GREG_4$	638.90	629.17	–	–
$SHMM_5$	636.71	629.61		
	–	–	637.40	671.10
Experiment 2				
$AMMI_3$	1322.60	1273.67	–	–
$COMM_3$	1316.89	1272.27	–	–
$SREG_3$	1318.83	1275.44	–	–
$GREG_3$	1316.28	1272.21	–	–
$SHMM_3$	1315.54	1275.35	–	–
	–	–	1284.04	1331.26
Experiment 3				
$AMMI_1$	816.41	800.36	–	–
$COMM_2$	813.39	799.42	–	–
$SREG_1$	816.30	800.49	–	–
$GREG_2$	809.03	799.24	–	–
$SHMM_2$	810.16	799.40		
	–	–	817.64	849.45
Experiment 4				
$AMMI_{10}$	802.57	796.69	–	–
$COMM_{11}$	822.29	795.86	–	–
$SREG_8$	823.79	798.33	–	–
$GREG_{11}$	822.53	797.38	–	–
$SHMM_{11}$	823.61	799.56	–	–
	–	–	798.86	831.10
Experiment 5				
$AMMI_5$	677.49	669.50	–	–
$COMM_1$	675.79	668.14	–	–
$SREG_1$	684.56	671.60	–	–
$GREG_1$	675.09	668.09	–	–
$SHMM_1$	678.09	671.48	–	–
	–	–	663.95	715.84

*The subscripts on the model forms indicate the number of bilinear terms retained in the best truncated model. This subscript is not related to the shrinkage estimates.
RMSPD, root mean square predicted difference; AMMI, additive main effects and multiplicative interaction model; GREG, genotypes regression model; SREG, sites regression model; COMM, completely multiplicative model; SHMM, shifted multiplicative model; BLUP, best linear unbiased predictor; MET, multi-environment trials.

shrinkage estimates are better predictors than empirical cell means and BLUPs of cell means (Table 20.3). Crossa *et al.* (2002) computed $SREC_1$ clustering of the 20 environments in this MET, with the empirical cell means replaced with SREC shrinkage estimates as input to the procedure. Two of the final groups of environments obtained, namely, [2, 4, 5, 6, 9, 11, 12, 13, 14, 16, 18] and [7, 15, 17, 19, 20], were different from those in Fig. 20.2. Only the environmental group [1, 3, 10] was the same. Whereas two of these groups did not agree with the clustering computed from empirical cell means, they did agree with groups of environments delineated by sectors of an $SREC_2$ biplot computed from deviations of empirical cell means from site means.

This is a very intriguing empirical result, suggesting that the methodology deserves study in more examples. Because of the separation of pattern from random error that appears to be achieved by the shrinkage estimators, we believe SREC clustering using shrinkage estimates of cell means as input has considerable promise as a routine procedure for the study of interaction patterns in an MET.

Software

To compute SHMM and SREC clustering of environments or SHMM clustering of genotypes, we first use an SAS® program (SAS/IML) to obtain a dendrogram. Then, adequacy of fit of $SHMM_1$ or $SREC_1$ (constrained, if necessary) is evaluated using the Fortran program EIGADV. Enquiries concerning the availability of the software may be sent to P.L. Cornelius (cornelius@ms.uky.edu).

References

Abdalla, O.S., Crossa, J. and Cornelius, P.L. (1997) Results and biological interpretation of shifted multiplicative model clustering of durum wheat cultivars and test sites. *Crop Science* 37, 881–97.

Cornelius, P.L. (1993) Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop Science* 33, 1186–1193.

Cornelius, P.L. and Crossa, J. (1998) *Shrinkage Estimators of Multiplicative Models for Cultivar Trials*. Technical Report No. 352, Department of Statistics, University of Kentucky, Lexington, Kentucky, USA.

Cornelius, P.L. and Crossa, J. (1999) Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. *Crop Science* 39, 998–1009.

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.

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., Crossa, J. and Seyedsadr, M. (1995a) Tests and estimators of multiplicative models for variety trials. In: *Proceedings of the 5th Annual Kansas State University Conference on Applied Statistics in Agriculture, Manhattan, Kansas*, pp. 130–160.

Cornelius, P.L., Van Sanford, D.A. and Seyedsadr, M. (1995b) Clustering cultivars into groups without rank-change interactions. *Crop Science* 33, 1195–1200.

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

Crossa, J. and Cornelius, P.L. (1993) Recent developments in multiplicative models for cultivar trials. In: Buxton, D.R., Shibles, R., Forsberg, R.A., Blad, B.L., Asay, K.H., Paulsen, G.M. and Wilson, R.F. (eds) *International Crop Science I*. Crop Science Society of America, Madison, Wisconsin, pp. 571–577.

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

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., 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 Sayedsadr, 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*. CRC Press, Boca Raton, Florida, pp. 175–198.

Crossa, J., Cornelius, P.L. and Yan, W. (2002) Biplots of linear–bilinear models for studying crossover genotype × environment interaction. *Crop Science* (in press).

Eberhart, S.A. and Russell, W.A. [1966] Stability parameters for comparing varieties. *Crop Science* 6, 36–40.

Finlay, K.W. and Wilkinson, G.N. [1963] The analysis of adaptation in a plant breeding programme. *Australian Journal of Agricultural Research* 14, 742–754.

Fisher, R.A. and MacKenzie, W.A. [1923] Studies in variation II. The manurial response in different potato varieties. *Journal of Agricultural Science* 13, 311–320.

Freeman, G.H. [1973] Statistical methods for the analysis of genotype–environment interactions. *Heredity* 31(3), 339–354.

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.

Johnson, D.E. and Graybill, F.A. [1972] An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* 67, 862–868.

Liu, G. and Cornelius, P.L. (2001) Simulations and decimal approximations for the means and standard deviations of the characteristic roots of a Wishart matrix. *Communications in Statistics B. Simulation and Computation* (in press).

Mandel, J. [1961] Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association* 56, 878–888.

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.

Sayedsadr, M. and Cornelius, P.L. [1992] Shifted multiplicative models for non-additive two-way tables. *Communications in Statistics B. Simulation and Computation* 21, 807–822.

Trethowan, R.M., Crossa, J., vanGinkel, M. and Rajaram, S. [2001] Relationships among bread wheat international yield testing locations in dry areas. *Crop Science* 41, 1461–1469.

Tukey, J.W. [1949] One degree of freedom for non-additivity. *Biometrics* 5, 232–242.

Williams, E.J. [1952] The interpretation of interactions in factorial experiments. *Biometrika* 39, 65–81.

Yan, W., Hunt, L.A., Sheng, Q. and Szlavnics, Z. [2000] Cultivar evaluation and mega-environment investigation based on the GGE biplot. *Crop Science* 40, 597–605.

Yates, F. and Cochran, W.G. [1938] The analysis of groups of experiments. *Journal of Agricultural Science* 28, 556–580.

Zobel, R.W., Wright, M.J. and Gauch, H.G., Jr [1988] Statistical analysis of a yield trial. *Agronomy Journal* 80, 388–393.

INTRODUCCIÓN AL ANÁLISIS MULTIVARIADO

Examen

1) En el archivo TRIAL2.TXT tenemos un experimento de 64 genotipos evaluados en dos sitio y 2 repeticiones en cada sitio. El diseño de experimento es en bloques completamente al azar. El archivo tiene hileras y columnas.

1a) Hacer el análisis espacial en cada ambiente considerando repeticiones y genotipos como efecto aleatorio.

1b) Hacer el análisis espacial combinado a través de los dos sitios considerando a sitios y a genotipos y a su interacción como efectos aleatorios.

1c) Identifique el error estándar de la diferencia entre cualquier media (SED) para los efectos principales y la interacción correspondiente.

2) En el archivo INPUTAUDPC.CSV hay datos que corresponden a 9 genotipos evaluados en 3 ambientes con tres repeticiones por ambiente (no tomar en cuenta la columna de IB). La variable es AUDPC (area under the disease progress curve)

2a) Hacer un BIPLOT del modelo AMMI e interpretar los resultados

2b) Hacer un biplot del modelo SREG e interpretar los resultados

Los estudiantes se pueden comunicar con José Crossa al correo j.crossa@cgiar.org y hacer las preguntas que deseen. Pueden trabajar en grupos de manera de discutir los ejercicios, pero la presentación es individual.

El plazo de entrega es el día miércoles 3 de noviembre, y los trabajos deben ser enviados a mi correo electrónico (eacevedo@uchile.cl; edmundoacevedo@vtr.net).

Saludos cordiales

Edmundo Acevedo

Archivo Trial 2

site	rep	row	col	variety	yield
1	1	1	1	5	1.5318
1	1	1	2	19	2.2211
1	1	1	3	55	1.4589
1	1	1	4	23	1.2436
1	1	1	5	27	1.8989
1	1	1	6	38	1.3366
1	1	1	7	64	1.8966
1	1	1	8	44	1.4775
1	1	1	9	14	1.6483
1	1	1	10	13	1.5220
1	1	1	11	59	1.8581
1	1	1	12	25	1.9730
1	1	1	13	45	2.5290
1	1	1	14	49	2.3901
1	1	1	15	11	2.4703
1	1	1	16	43	3.5279
1	1	2	1	2	2.0066
1	1	2	2	36	1.4573
1	1	2	3	62	2.2574
1	1	2	4	29	1.4428
1	1	2	5	34	1.3521
1	1	2	6	40	1.8021
1	1	2	7	52	1.3945
1	1	2	8	54	1.5750
1	1	2	9	28	1.3825
1	1	2	10	26	1.6735
1	1	2	11	58	1.9658
1	1	2	12	42	2.1403
1	1	2	13	32	2.6153
1	1	2	14	6	2.8538
1	1	2	15	39	3.4978
1	1	2	16	53	2.4935
1	1	3	1	37	1.8915
1	1	3	2	3	1.9649
1	1	3	3	17	1.5145
1	1	3	4	24	1.3582
1	1	3	5	15	1.8466
1	1	3	6	50	1.7456
1	1	3	7	47	1.3042
1	1	3	8	56	1.7161
1	1	3	9	16	1.1509
1	1	3	10	30	1.4549
1	1	3	11	31	1.6291
1	1	3	12	60	1.4454
1	1	3	13	7	2.0237
1	1	3	14	18	2.8159
1	1	3	15	8	3.2111
1	1	3	16	21	3.2198
1	1	4	1	20	3.3420
1	1	4	2	35	2.6472
1	1	4	3	57	1.9362
1	1	4	4	63	3.1655
1	1	4	5	46	2.2962
1	1	4	6	12	2.7581
1	1	4	7	41	2.4225
1	1	4	8	10	2.6407
1	1	4	9	22	2.8679
1	1	4	10	1	2.8687

1	1	4	11	51	3.1765
1	1	4	12	61	2.5182
1	1	4	13	33	2.7195
1	1	4	14	48	3.4329
1	1	4	15	4	3.8140
1	1	4	16	9	4.1069
1	2	5	1	58	3.7139
1	2	5	2	44	3.6386
1	2	5	3	15	2.8465
1	2	5	4	25	2.4823
1	2	5	5	18	2.6002
1	2	5	6	53	2.5417
1	2	5	7	5	1.6013
1	2	5	8	22	2.1818
1	2	5	9	55	3.5086
1	2	5	10	36	3.4147
1	2	5	11	33	3.1765
1	2	5	12	29	2.8818
1	2	5	13	12	3.1165
1	2	5	14	47	3.5868
1	2	5	15	8	3.5954
1	2	5	16	1	3.5711
1	2	6	1	41	2.7717
1	2	6	2	17	2.9044
1	2	6	3	63	3.0646
1	2	6	4	37	3.2219
1	2	6	5	38	2.4959
1	2	6	6	19	3.7729
1	2	6	7	26	2.7808
1	2	6	8	21	2.5957
1	2	6	9	28	2.3549
1	2	6	10	4	3.0692
1	2	6	11	6	3.2589
1	2	6	12	45	3.4685
1	2	6	13	56	2.9773
1	2	6	14	46	3.6739
1	2	6	15	59	4.4576
1	2	6	16	42	4.7826
1	2	7	1	14	2.9164
1	2	7	2	9	2.2630
1	2	7	3	62	1.9820
1	2	7	4	48	2.6628
1	2	7	5	34	2.7700
1	2	7	6	31	2.7649
1	2	7	7	60	2.0307
1	2	7	8	13	2.3102
1	2	7	9	20	2.2525
1	2	7	10	7	2.6594
1	2	7	11	64	2.3498
1	2	7	12	3	2.8951
1	2	7	13	39	3.5951
1	2	7	14	40	3.6799
1	2	7	15	10	3.8023
1	2	7	16	23	4.2910
1	2	8	1	24	2.2748
1	2	8	2	52	2.1081
1	2	8	3	11	2.5377
1	2	8	4	54	1.4670
1	2	8	5	51	1.6475
1	2	8	6	57	2.7363
1	2	8	7	35	2.2646
1	2	8	8	2	2.6282

1	2	8	9	50	1.7699
1	2	8	10	27	2.1685
1	2	8	11	30	2.1797
1	2	8	12	61	2.7998
1	2	8	13	43	2.3903
1	2	8	14	49	3.2857
1	2	8	15	32	3.8995
1	2	8	16	16	3.5976
2	1	1	1	5	3.5318
2	1	1	2	19	4.2211
2	1	1	3	55	5.4589
2	1	1	4	23	6.2436
2	1	1	5	27	7.8989
2	1	1	6	38	2.3366
2	1	1	7	64	3.8966
2	1	1	8	44	4.4775
2	1	1	9	14	1.6483
2	1	1	10	13	2.5220
2	1	1	11	59	3.8581
2	1	1	12	25	4.9730
2	1	1	13	45	5.5290
2	1	1	14	49	6.3901
2	1	1	15	11	7.4703
2	1	1	16	43	3.5279
2	1	2	1	2	9.0066
2	1	2	2	36	1.4573
2	1	2	3	62	6.2574
2	1	2	4	29	2.4428
2	1	2	5	34	3.3521
2	1	2	6	40	2.8021
2	1	2	7	52	3.3945
2	1	2	8	54	4.5750
2	1	2	9	28	5.3825
2	1	2	10	26	6.6735
2	1	2	11	58	7.9658
2	1	2	12	42	3.1403
2	1	2	13	32	4.6153
2	1	2	14	6	3.8538
2	1	2	15	39	7.4978
2	1	2	16	53	8.4935
2	1	3	1	37	9.8915
2	1	3	2	3	9.9649
2	1	3	3	17	9.5145
2	1	3	4	24	9.3582
2	1	3	5	15	1.8466
2	1	3	6	50	2.7456
2	1	3	7	47	3.3042
2	1	3	8	56	4.7161
2	1	3	9	16	4.1509
2	1	3	10	30	4.4549
2	1	3	11	31	4.6291
2	1	3	12	60	4.4454
2	1	3	13	7	4.0237
2	1	3	14	18	5.8159
2	1	3	15	8	6.2111
2	1	3	16	21	3.2198
2	1	4	1	20	4.3420
2	1	4	2	35	5.6472
2	1	4	3	57	2.9362
2	1	4	4	63	1.1655
2	1	4	5	46	1.2962
2	1	4	6	12	1.7581

2	1	4	7	41	2.4225
2	1	4	8	10	3.6407
2	1	4	9	22	4.8679
2	1	4	10	1	3.8687
2	1	4	11	51	4.1765
2	1	4	12	61	3.5182
2	1	4	13	33	4.7195
2	1	4	14	48	5.4329
2	1	4	15	4	5.8140
2	1	4	16	9	6.1069
2	2	5	1	58	7.7139
2	2	5	2	44	7.6386
2	2	5	3	15	8.8465
2	2	5	4	25	8.4823
2	2	5	5	18	8.6002
2	2	5	6	53	9.5417
2	2	5	7	5	7.6013
2	2	5	8	22	6.1818
2	2	5	9	55	5.5086
2	2	5	10	36	4.4147
2	2	5	11	33	3.1765
2	2	5	12	29	2.8818
2	2	5	13	12	2.1165
2	2	5	14	47	3.5868
2	2	5	15	8	4.5954
2	2	5	16	1	5.5711
2	2	6	1	41	6.7717
2	2	6	2	17	7.9044
2	2	6	3	63	8.0646
2	2	6	4	37	9.2219
2	2	6	5	38	7.4959
2	2	6	6	19	6.7729
2	2	6	7	26	5.7808
2	2	6	8	21	4.5957
2	2	6	9	28	3.3549
2	2	6	10	4	3.0692
2	2	6	11	6	4.2589
2	2	6	12	45	4.4685
2	2	6	13	56	5.9773
2	2	6	14	46	6.6739
2	2	6	15	59	7.4576
2	2	6	16	42	8.7826
2	2	7	1	14	9.9164
2	2	7	2	9	2.2630
2	2	7	3	62	1.9820
2	2	7	4	48	7.6628
2	2	7	5	34	5.7700
2	2	7	6	31	4.7649
2	2	7	7	60	5.0307
2	2	7	8	13	6.3102
2	2	7	9	20	7.2525
2	2	7	10	7	3.6594
2	2	7	11	64	4.3498
2	2	7	12	3	4.8951
2	2	7	13	39	2.5951
2	2	7	14	40	3.6799
2	2	7	15	10	4.8023
2	2	7	16	23	5.2910
2	2	8	1	24	6.2748
2	2	8	2	52	7.1081
2	2	8	3	11	8.5377
2	2	8	4	54	8.4670

2	2	8	5	51	9.6475
2	2	8	6	57	9.7363
2	2	8	7	35	8.2646
2	2	8	8	2	7.6282
2	2	8	9	50	6.7699
2	2	8	10	27	5.1685
2	2	8	11	30	4.1797
2	2	8	12	61	5.7998
2	2	8	13	43	3.3903
2	2	8	14	49	4.2857
2	2	8	15	32	5.8995
2	2	8	16	16	5.5976

Archivo INPUTAUDPC.CSV

Env	Genotype	IB	Rep	AUDPC
1	1	2	1	545
1	1	9	2	652.5
1	1	17	3	712.5
2	1	2	1	1732.5
2	1	9	2	1392.5
2	1	17	3	1662.5
3	1	1	1	1508.5
3	1	7	2	1575
3	1	4	3	1575
1	11	4	1	242.5
1	11	11	2	277.5
1	11	17	3	237.5
2	11	4	1	757.5
2	11	11	2	792.5
2	11	17	3	792.5
3	11	6	1	882
3	11	1	2	637
3	11	4	3	882
1	12	4	1	172.5
1	12	7	2	147.5
1	12	15	3	170
2	12	4	1	897.5
2	12	7	2	517.5
2	12	15	3	692.5
3	12	7	1	322
3	12	1	2	287
3	12	3	3	287
1	14	4	1	170
1	14	9	2	137.5
1	14	13	3	102.5
2	14	4	1	697.5
2	14	9	2	452.5
2	14	13	3	485
3	14	3	1	175
3	14	7	2	490
3	14	1	3	420
1	15	4	1	105
1	15	7	2	137.5
1	15	13	3	135
2	15	4	1	412.5
2	15	7	2	312.5
2	15	13	3	310
3	15	6	1	315
3	15	4	2	210
3	15	2	3	245
1	17	4	1	67.5
1	17	7	2	67.5
1	17	13	3	70
2	17	4	1	93.5

2	17	7	2	65.5
2	17	13	3	93.5
3	17	7	1	420
3	17	3	2	350
3	17	2	3	350
1	19	4	1	85
1	19	9	2	85
1	19	15	3	102.5
2	19	4	1	277.5
2	19	9	2	260
2	19	15	3	277.5
3	19	5	1	462
3	19	3	2	497
3	19	4	3	707
1	29	6	1	272.5
1	29	9	2	242.5
1	29	17	3	277.5
2	29	6	1	382.5
2	29	9	2	382.5
2	29	17	3	417.5
3	29	1	1	532
3	29	2	2	742
3	29	6	3	595
1	30	6	1	687.5
1	30	11	2	747.5
1	30	15	3	735
2	30	6	1	1532.5
2	30	11	2	1562.5
2	30	15	3	1602.5
3	30	6	1	1648.5
3	30	5	2	1578.5
3	30	6	3	1680

INTRODUCCION AL ANALISIS MULTIVARIADO

DR. JOSE CROSSA R.
Biometrics and Statistics Unit
International Maize and Wheat Improvement Center (CIMMYT)

12 al 14 de octubre de 2004
Campus Antumapu, Facultad de Ciencias Agronómicas. Universidad de Chile

Listado de Asistentes

Alumnos que aprobaron el curso

Nombre	Institución
Edmundo Acevedo Hinojosa	U. de Chile
Paola Silva Candia	U. de Chile
Herman Silva Robledo	U. de Chile
Eduardo Martínez	U. de Chile
Susana Valle Toledo	U. de Chile
Mauricio Ortiz Lizana	U. de Chile
Javiera González Cruz	U. de Chile
Claudia Harcha Cortes	U. Austral
Patricio Sandaña G.	U. Austral
Gaston Delard Rodríguez	Semillas Pioneer
Miguel Ibáñez Vial	Semillas Pioneer
Andrea Salinas Rubio	Semillas Pioneer
Juan Eduardo Zarhi Salim-Hanna	Semillas Pioneer
Sybil Herrera Foessel	CIMMYT

Alumnos que no aprobaron el curso

Nombre	Institución
Alberto Mancilla Martínez	U. de Chile
Juan Barrios	U. de Chile
Carlos Montes Verdugo	U. de Chile
Marcela Opazo Illanes	U. de Chile
Daniel Calderini	U. Austral
Rodrigo Carvacho Baillon	Semillas CIS
Erika Salazar Suazo	INIA
Jorge Saavedra DelReal	INIA
Carolina Rivera Montoya	U. de Chile
Verónica Muñoz Muñoz	U. de Chile

SANTIAGO, 17 de enero de 2005

A QUIEN CORRESPONDA

EL CURSO INTRODUCCION AL ANALISIS MULTIVARIADO, DICTADO POR EL PROFESOR JOSE CROSSA, ESTA RECONOCIDO EN CALIDAD DE CURSO ELECTIVO EN EL PROGRAMA DE DOCTORADO EN CIENCIAS SILVOAGROPECUARIAS Y VETERINARIAS, DEL CAMPUS SUR, DE LA UNIVERSIDAD DE CHILE.

Universidad de Chile
Comité Académico
Doctorado en Ciencias
Silvoagropecuarias y Veterinarias
Campus Sur

CECILIA DE LA MAZA A.
Secretaria Académica
Programa DSAVP

Sta Rosa 11.315 Casilla 1004 Correo Central, Santiago, Chile/ fono 678 5864 Fax 56-2-678 5752 e mail: cdelamaz@uchile.cl