



**HAL**  
open science

## Statistical analysis and interpretation of line x environment interaction for biomass yield in maize

O Argillier, Y Hébert, Y Barrière

### ► To cite this version:

O Argillier, Y Hébert, Y Barrière. Statistical analysis and interpretation of line x environment interaction for biomass yield in maize. *Agronomie*, 1994, 14 (10), pp.661-672. hal-00885602

**HAL Id: hal-00885602**

**<https://hal.science/hal-00885602v1>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistical analysis and interpretation of line x environment interaction for biomass yield in maize

O Argillier \*, Y Hébert, Y Barrière

INRA, station d'amélioration des plantes fourragères, F86600 Lusignan, France

(Received 3 October 1994; accepted 12 January 1995)

**Summary** — The maize line x environment interaction for biomass dry matter yield was analysed using a multilocal factorial mating design. Various models, such as joint regression, biadditive model, factorial regression and structuring, were performed in order to partition and explain the interaction. Except for the joint regression model, which oversimplified the interaction pattern, all the models were effective in accounting for the line x environment interaction. Biological connections have been established between these models. The biological interpretation, using additional information, shows that the line x environment interaction for biomass yield in maize could to a large extent be due to earliness effects and yield-limiting factors, such as lodging susceptibility and water stress. The consequences of interaction modelling in plant breeding are discussed.

**genotype x environment interaction / factorial regression / biadditive model / pattern analysis / biomass yield / maize**

**Résumé** — **Analyse statistique et interprétation des interactions lignée x environnement pour le rendement en biomasse chez le maïs.** *Les interactions lignée x environnement pour le rendement en biomasse chez le maïs ont été étudiées à partir d'un plan factoriel multilocal. Différents modèles, comme la régression conjointe, la modélisation biadditive, la régression factorielle et la structuration, ont été utilisés dans le but de décomposer et d'expliquer ces interactions. En dehors de la régression conjointe, trop simplificatrice, les autres modèles sont tous efficaces pour rendre compte de l'interaction. Certaines connexions biologiques ont pu être mises en évidence entre les modèles. L'interprétation biologique, grâce surtout à la connaissance d'informations supplémentaires sur les milieux et lignées, montre que la plus grande part de l'interaction lignée x environnement pour le rendement en biomasse chez le maïs est due à des effets précocité et à des facteurs limitants du rendement, comme la sensibilité à la verse et le manque d'eau. Les conséquences en sélection sont aussi discutées.*

**interaction génotype x environnement / régression factorielle / modèle biadditif / structuration / rendement en biomasse / maïs**

\* Correspondence and reprints

## INTRODUCTION

In many trials in which a set of genotypes is grown over a range of environments, the genotypes have distinct differential responses. This phenomenon, known as genotype x environment interaction, presents serious problems in comparing the performances of genotypes over several environments and affects the extent of genetic progress through selection. Thus, when the genotype x environment interaction is significant, its nature, cause and implication must be carefully examined (Magari and Kang, 1993). The detection and characterization of genotype x environment interaction has been approached in various ways, which have been reviewed by Freeman (1973), Denis and Vincourt (1982) and Westcott (1986).

In maize, it has been shown that grain and biomass yield display significant high main effects and genotype x environment interactions (Vattikonda and Hunter, 1983; Geiger *et al*, 1986; Kang and Gorman, 1989; Crossa *et al*, 1990; Dhillon *et al*, 1990; Magari and Kang, 1993). The objective of maize breeders is to produce high-yielding and adapted genotypes for a wide range of environments. Genotype x environment interactions were studied for grain productivity (Kang and Gorman, 1989; Crossa *et al*, 1990), but not for biomass productivity.

Whole-plant dry matter yield is a major criterion for silage maize breeding. Data from a multi-local factorial mating design were used to study the effects of contrasting environments on the general combining ability of inbred lines. We focused on the interaction between the breeding value of lines and environments. Statistical models were used in order to characterize the lines and environments, and reveal the biological factors responsible for the interaction, thus analyzing the consequences for plant breeding. Several statistical models were used in order to see if their complementarity features could be of any help to the biological interpretation of interactions. Consequently, these models were compared on the basis of their effectiveness in accounting for line x environment interaction. The prediction ability of the models was not considered.

## MATERIALS AND METHODS

### *Experimental data*

The study was conducted within the framework of an agreement between various private breeding companies belonging to Promais<sup>1</sup>, the French research institute INRA and the French Ministry of Agriculture. Ten inbred lines (table I) were chosen as parents of a factorial mating design. These lines originate from various germplasm and present a large variability for yield. They were crossed to 4 tester lines (table I). The crosses were evaluated on 21 environments. These 21 environments corresponded to 2 years (1992 and 1993) and/or different locations, from south-west of France to the Netherlands (fig 1). The trials were randomized block designs with one replicate, each block corresponding to crosses with one tester line. Block effects and tester effects were thus confounded. The plant density was about 100 000 plants per hectare. The trials were harvested at silage stage and in each environment at the same date for all the genotypes. In addition to whole plant dry matter content and dry matter yield, the following traits were measured: mid-silking date (as the number of days after July 1st) and root lodging susceptibility (marked from 0, not lodged, to 5, all plants affected by severe lodging).

Meteorological data were also recorded in each environment: sum of air temperature from sowing to harvest (degree days, basis 6°C); average daily air temperature above 6°C; sum of rainfall from sowing to harvest (mm); and sum of rainfall during the months of June, July and August (mm), which correspond to the period of maximum sensitivity of maize to water stress.

### *Statistical analyses*

#### *Preliminary analyses*

As the objective of this study was to analyze the interaction between the line breeding values and the environments, we considered the means of hybrids over the testers in each environment. Let  $Y_{ij}$  be the breeding value of line  $i$  in the environment  $j$ . A classical model can be written:

$$Y_{ij} = \mu + l_i + e_j + l_{ij}$$

where  $\mu$  is the grand mean,  $l_i$  the average breeding value of line  $i$ ,  $e_j$  the average effect of environment  $j$ , and  $l_{ij}$  the interaction studied between the breeding value of line  $i$  and the environment  $j$ .

<sup>1</sup> Members of Promais who joined the network program: CACBA, Cargill, Caussade Semences, Ciba Semences, Corn States International, Eurosemences, Limagrain, Maisadour, Northrup King Semences, Pioneer, Prograin Génétique, RAGT, Rhône-Poulenc Agro, Rustica, SDME, Semundo, SES, Van der Have.

**Table I.** The 10 inbred lines of the design and the 4 tester lines.

	Origin	Earliness and heterotic group
<i>Inbred lines</i>		
F271	Co125	Early dent
F252	Co125	Early dent
F244	US and European germplasms	Early dent
F288	F252 and F244	Early dent
Co125	Canadian germplasm	Early dent
W33	US germplasm	Early dent
Co255	INRA 258	Early flint x dent
F113	US germplasm	Medium early dent
F1772	French populations	Medium late flint
LH74	BSSS	Medium late dent
<i>Tester lines</i>		
F2	Lacaune	Early flint
F7	Lacaune	Early flint
MBS847	Iodent	Medium late dent
W117	US germplasm	Early dent

The data from the plots of all crosses between lines and testers in the different environments, provided an error term  $E_{ijk}$  (degrees of freedom = 540), confounding the triple interaction between line  $i$ , tester  $k$ , and environment  $j$ , and the random experimental error. Therefore, as we could not distinguish the triple interaction from the experimental error, since only one

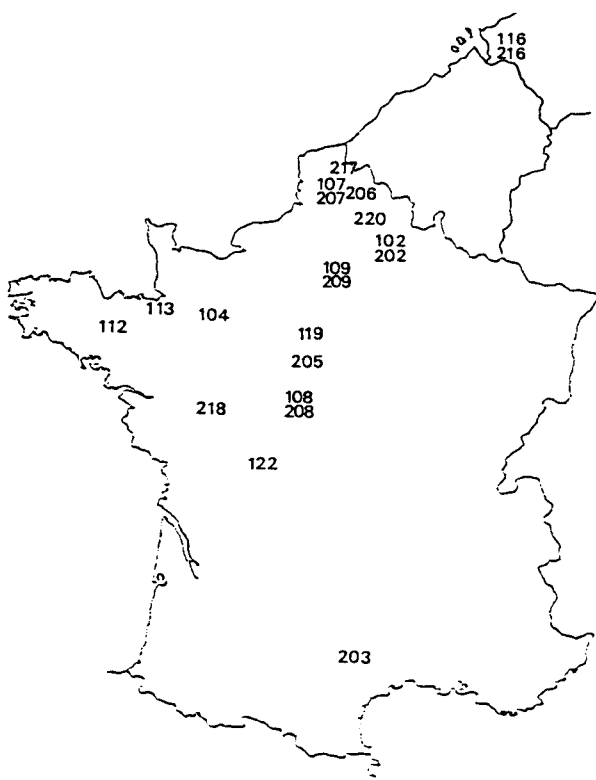
replicate was available, we had to suppose that the triple interaction was low compared to the experimental error, in order to perform the classical significance tests. To confirm this hypothesis, the  $E_{ijk}$  mean square was compared to a mean of error variance estimations provided by comparable trials including replicates. The  $F$ -ratio, equal to 1.12 was not significant ( $p$  value = 0.21). Therefore, we could consider that the estimated variance of  $E_{ijk}$  overestimated the true error variance by only a small amount. Consequently, this variance estimation has been used in the significance tests involved in interaction models. The estimated error terms were confirmed to meet the required assumptions of the analysis of variance (they were identically and independently distributed like  $\mathcal{N}(0, \sigma^2)$ ).

### Interaction modelling

Several linear and non-linear models were fitted in order to partition the sum of squares of the line x environment interaction ( $SS_I$ ) in different ways. Each model enables splitting  $I_{ij}$  into 2 parts. One accounts for variation due to interaction while the other is supposed to be a residual term. The analyses performed were: joint regression; biadditive model; pattern analysis; and factorial regression. All these models were regarded as fixed-effect models. They were of 2 types: the first 3 models used no extra-information, the latter included complementary characteristics of genotypes and/or of environments.

#### Joint regression

The most classical approach of joint regression was described by Yates and Cochran (1938), Tukey (1949) and Finlay and Wilkinson (1963). The interaction is assumed to be a linear function of the mean perfor-



**Fig 1.** Characterization of the 21 environments. The first figure of environment number corresponds to the year (1: 1992; and 2: 1993); the last 2 correspond to the number of the location.

mance of the lines in each environment. The model of the expectation of the line  $i$  in the environment  $j$  is:

$$E(Y_{ij}) = \mu + l_i + e_j + \gamma_i \hat{e}_j$$

where  $\gamma_i$  stands for the regression coefficient of the response of line  $i$  on the estimate  $\hat{e}_j$  of the environment main effect.

### Biadditive model

Biadditive models of interaction effects were described by Gollob (1968), Mandel (1971), Crossa *et al* (1990) and Denis and Gower (1992). The basis of these models is to partition  $l_{ij}$  as a sum of multiplicative terms involving parameters which are specific to each of the interacting factors. The expectation of response can be written:

$$E(Y_{ij}) = \mu + l_i + e_j + \sum_{u=1}^r \lambda_{ui} \eta_{uj}$$

where  $r$  is the number of multiplicative terms,  $\lambda_{ui}$  and  $\eta_{uj}$  are, respectively, the parameters specific to the line  $i$  and environment  $j$  for the  $u$ th multiplicative term ( $\sum \lambda_{ui}^2 = \sum \eta_{uj}^2, \forall u$ ). This model shows analogies with principal component analysis. It is a bilinear model which does not need to assume a linear response of the genotypes over the environments.

Multiplicative terms were introduced as long as they significantly explained the interaction and until the residual term of the interaction was no longer significant.

### Interaction structuring

We chose the simultaneous agglomerative hierarchical clustering procedure based on the interaction term (Corsten and Denis, 1990). Groups of lines and groups of environments were simultaneously identified, in such a way that the interaction was mainly distributed between groups. The simultaneous clustering of lines and environments leads to the model:

$$E(Y_{ij}) = \mu + l_i + e_j + (LE)_{g(i)h(j)} + (Le)_{g(i)j} + (IE)_{ih(j)} + (Ie)_{ij}$$

where  $g(i)$  and  $h(j)$  designate the groups formed with the lines and environments, respectively. In this way, the line x environment interaction was split into 4 parts: variation between groups of lines – between groups of environments (BB), considered as the explained part of line x environment interaction; variation between groups of lines – within groups of environments (BW), variation within groups of lines – between groups of environments (WB); and variation within groups of lines – within groups of environments (WW). The last 3 together were considered to be the residual term of interaction.

As suggested by Baril *et al* (1994), we decided to stop the clustering process when the determination coefficient (defined as the ratio of the 2 main effects plus BB variation on the total variation of the model) was greater than 0.95.

### Factorial regression

The factorial regression model (Hardwick and Wood, 1972; Wood, 1976; Denis, 1980, 1988) uses concomitant genotypic and environmental information in order to split the line x environment interaction into biologically interpretable terms. This method can be considered as expensive, for it requires additional recordings, especially in an exploratory stage, when no precise source of variation can be suspected. However it is the only model that can directly lead to effective biological interpretations that are useful for growth prediction (Haun, 1982), environment potential characterization (Abou-El Fittouh *et al*, 1969) and plant breeding (Hardwick and Wood, 1972; Wood, 1976).

The mid-silking date (*SD*) and dry matter content (*DM*) were chosen to serve as genotypic and environmental covariates. Lodging susceptibility (*LS*) served only as genotypic covariate because of technical problems. For each trait, the estimations of the genotypic and environmental additive parameters respectively defined the genotypic and environmental covariates, as proposed by Baril (1992). The meteorological observations also served as environmental covariates.

The stepwise process proposed by Denis (1988) was applied. The factorial regression model was built by successive addition of the most significant covariates explaining the line x environment interaction. After finding the best single covariate model among all the possible 1-covariate models, the best 2-covariate model was looked for, given the first covariate, and so forth until the addition of a covariate brought no more significant information. For example, using 1 covariate ( $X$ ) for the line effect, and 1 covariate ( $Z$ ) for the environment effect, the decomposition of the different effects involved in the model is:

$$\begin{aligned} l_i &= \delta \cdot X_i + \alpha_i \\ e_j &= \omega \cdot Z_j + \beta_j \\ l_{ij} &= \varphi \cdot X_i \cdot Z_j + \rho_i \cdot Z_j + \tau_j \cdot X_i \end{aligned}$$

where  $\delta$  and  $\omega$  are the regression coefficients on main effects (line and environment, respectively),  $\alpha_i$  and  $\beta_j$  are the residual terms of the genotypic and environmental main effects, respectively,  $\varphi$  is the regression coefficient on the product of the 2 covariates,  $\rho_i$  is the genotypic regression coefficient on the environmental covariate, and  $\tau_j$  is the environmental regression coefficient on the genotypic covariate.

The models were compared regarding their ability to significantly explain the interaction. Classically, this is made by use of the coefficient of determination  $R^2$ , which is computed as the percentage of the interaction sum of squares accounted for by each model, namely:

$$R^2 = SS_M / SS_I$$

where  $SS_I$  and  $SS_M$  are the sums of squares of the total interaction and of the part of interaction explained by the model used, respectively.

We also took into account the number of degrees of freedom used by the models, because these were intrinsically different regarding their degrees of com-

plexity (number of parameters). We therefore computed  $R^2$  adjusted for the number of parameters in the model (Judge *et al.*, 1980), calculated as:

$$adjR^2 = 1 - \frac{(df_i - 1) \cdot (1 - R^2)}{(df_i - df_M)}$$

where  $df_i$  and  $df_M$  are the degrees of freedom of the total interaction and of the part of interaction explained by the model, respectively.

The statistical analyses were performed using the Intera software (Decoux and Denis, 1991) which provides least-square estimates of the parameters.

## RESULTS

Table II indicates that, in the joint regression on an estimate of the main environmental effect, the differences between regression coefficients  $\gamma_i$

accounted for 6% of the total interaction only. As a consequence the residual term of interaction was highly significant. The study of the residual terms of the interaction revealed that the model fitted the data badly for some particular lines like LH74, F1772 and Co125.

Only 2 multiplicative terms were introduced in the biadditive model and led to a non-significant residue; they accounted for 56% of the interaction sum of squares, using only 30% of the degrees of freedom of the interaction (table III). Figure 2 presents the plots of the first multiplicative parameters ( $\lambda_{1i}$  and  $\eta_{1j}$ ) against the additive parameters and the second multiplicative parameters ( $\lambda_{2i}$  and  $\eta_{2j}$ ) against the first multiplicative parameters, for the lines and environments. Figure 2a indicates the contrasting behavior of line LH74, which displayed the highest additive parameter and by far the highest  $\lambda_{1i}$  (this line is also the latest one). Except for lines F1772,

**Table II.** Variance analysis of joint regression on an estimate of the main environment effect.

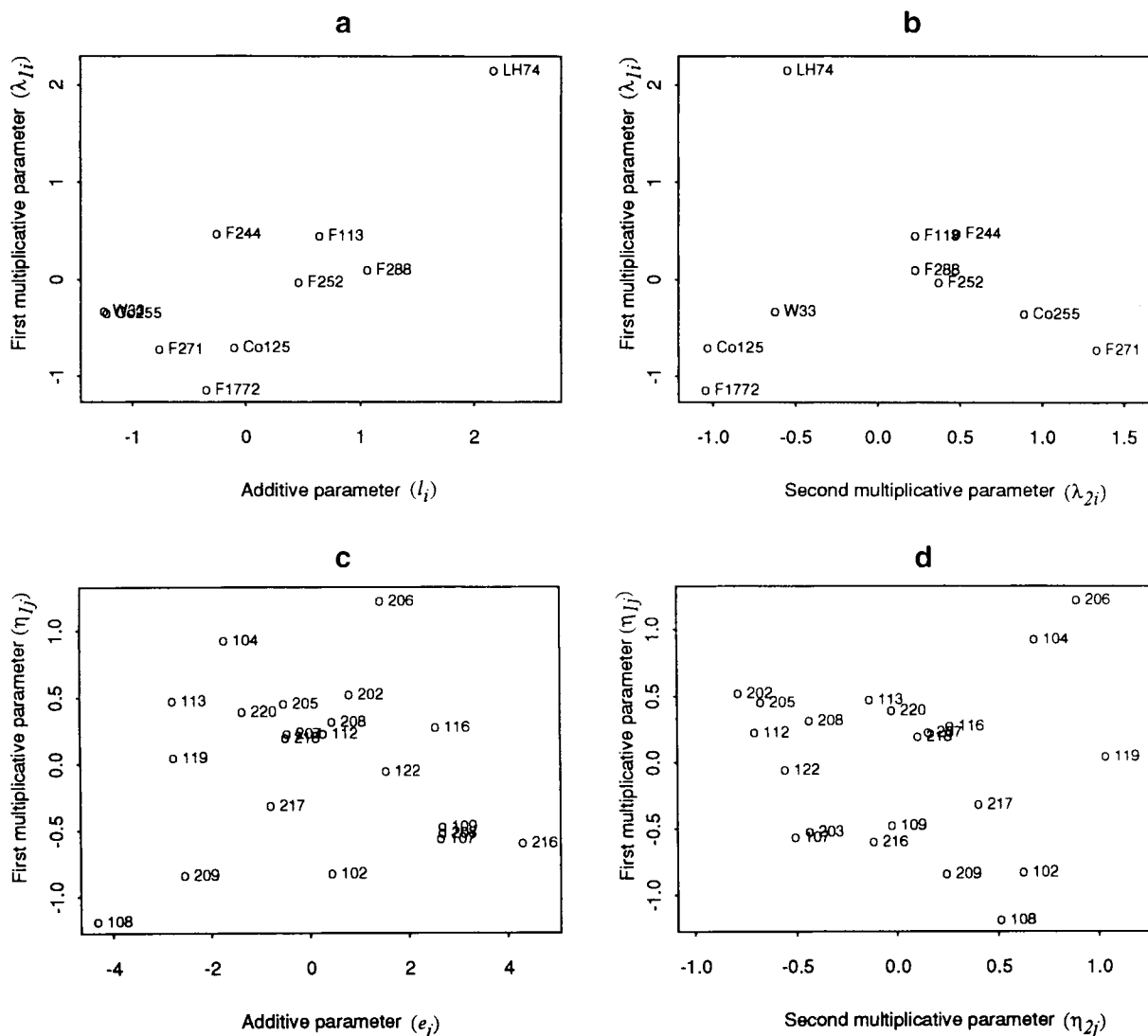
	SS	df	F	%SS <sub>i</sub>
Environment ( $e_j$ )	3 932.0	20	114.6	
Line ( $l_i$ )	862.2	9	55.8	
Interaction ( $l_{ij}$ )	651.6	180	3.6	100
Regression ( $\gamma_i \hat{e}_j$ )	38.1	9	2.5	6
Residual	613.4	171	2.1	94
Error	926.9	540		

Sums of squares (SS), degrees of freedom (df), F-ratios and percentage of interaction sum of squares accounted for by interaction modelling (%SS<sub>i</sub>). F-ratios are significant at the 1% probability level.

**Table III.** Variance analysis of biadditive model.

	SS	df	F	%SS <sub>i</sub>
Environment ( $e_j$ )	3 932.0	20	114.6	
Line ( $l_i$ )	862.2	9	55.8	
Interaction ( $l_{ij}$ )	651.6	180	3.6	100
MT1 ( $\lambda_{1i} \eta_{1j}$ )	229.2	28	4.8	35
MT2 ( $\lambda_{2i} \eta_{2j}$ )	136.6	26	3.1	21
Residual	285.8	126	1.3 ns	44
Error	926.9	540		

Sums of squares (SS), degrees of freedom (df), F-ratios and percentage of interaction sum of squares accounted for by interaction modelling (%SS<sub>i</sub>). F-ratios are significant at the 1% probability level, except for those marked ns.



**Fig 2.** Plots of the main effect estimates, first and second multiplicative parameters of the biadditive model. **a)** Additive and first multiplicative parameters of the lines; **b)** first and second multiplicative parameters of the lines; **c)** environmental additive and first multiplicative parameters; and **d)** environmental first and second multiplicative parameters.

Co125 and F271, which were susceptible to root lodging and displayed quite a strong negative  $\lambda_{1i}$ , the other lines had low  $\lambda_{1i}$ . The plot of  $\lambda_{2i}$  against  $\lambda_{1i}$  (fig 2b), distinguishes F271 and Co255, which displayed a high positive  $\lambda_{2i}$ , and W33, F1772 and Co125, which had a high negative  $\lambda_{2i}$ . Figures 2c and 2d show a homogeneous distribution of the environments on both plots. Notably, there was no structure according to the location and/or the year.

Table IV recapitulates the results of the optimal factorial regression model, among the available covariates. With 3 covariates using 26% of the available degrees of freedom, 55% of the interaction sum of squares could be explained, the residual term of the interaction being non-significant at a 0.01 probability level. The genotypic covariates *SD* and *LS* explained 24 and 22% of

the  $SS_i$  respectively. Rainfall during the period from June to August (*RF*) explained 9% of the  $SS_i$ . We can notice that these covariates also explained a large part of the main effects. The 2 genotypic covariates *SD* and *LS* explained 56% of the yield variation among lines and the environmental covariate *RF* explained 28% of the yield variation among environments (table IV).

Table V gives the values of the covariates for the lines and the associated regression coefficients. Some environments displayed very high negative regression coefficients on the genotypic covariate *LS*, such as, for example, environments 206 and 104. In contrast, other environments exhibited high positive regression coefficients on the genotypic covariate *LS*, such as, for example, environments 209 and 122. Regarding the regression coefficient on the genotypic covariate

**Table IV.** Variance analysis of factorial regression.

	SS	df	F	%SS <sub>i</sub>
Environment ( $e_j$ )	3 932.0	20	114.6	
$\omega.RF_j$	1 118.0	1	651.2	
$\beta_j$	2 814.0	19	36.3	
Line ( $l_j$ )	862.2	9	55.8	
$\delta_1.SD_i$	392.8	1	228.8	
$\delta_2.LS_i$	86.8	1	50.4	
$\alpha_i$	382.6	7	31.9	
Interaction ( $l_{ij}$ )	651.6	180	3.6	100
$\varphi 1.RF_j.LS_i$	2.8	1	1.6 ns	$\epsilon$
$\varphi 2.RF_j.SD_i$	0.1	1	0.1 ns	$\epsilon$
$\tau_{1j}.SD_i$	154.5	19	4.7	24
$\tau_{2j}.LS_i$	140.6	19	4.3	22
$\rho_j.RF_j$	61.1	7	5.1	9
Residual	292.6	133	1.3 ns	45
Error	926.9	540		

Sums of squares (*SS*), degrees of freedom (*df*), *F*-ratios and percentage of interaction sum of squares accounted for by interaction modelling (%*SS<sub>i</sub>*). *F*-ratios are significant at the 1% probability level, except for those marked ns.  $\epsilon$  stands for %*SS<sub>i</sub>* < 1. The genotypic covariates are: mid-silking date (*SD*) and lodging susceptibility (*LS*); the environmental covariate is the sum of rainfalls during the months of June, July and August (*RF*).

*SD*, we can particularly distinguish 2 types of environments. Environments such as 108 and 209 displayed high negative regression coefficients. Environments such as 122 displayed high positive regression coefficients.

Table V also gives the values of the environmental covariate and the associated regression coefficients. Some lines such as F1772, Co255 and F244 displayed positive regression coefficients on the environmental covariate *RF*, whereas lines such as Co125, LH74, F288 and W33 exhibited negative regression coefficients.

By stopping the clustering process when the determination coefficient was equal to 0.95, 7 groups of environments and 6 groups of lines were obtained (table V). The *BB* term was highly significant and contained 59% of the *SS<sub>i</sub>* with only 17% of the total degrees of freedom (table VI). Moreover, the remainder term was not significant at the 0.01 probability level.

## DISCUSSION

The line x environment interaction can be analyzed by use of several models, such as joint regression, the biadditive model, factorial regression and cluster analysis. This analysis can help us to: i) propose a biological interpretation of the

interaction; ii) compare the different models regarding their statistical effectiveness, their similarity or complementarity; and iii) analyze their relative consequences for plant breeding.

### Biological interpretation

The factorial regression model provided an interesting partitioning of line x environment interaction of yield data into a sum of linear functions of genotypic and environmental covariates, which have the advantage of enabling biological explanations of interactions for yield. Based on our results, we can conclude that line x environment interaction for biomass dry matter yield is mainly due to earliness effects and yield-limiting factors such as lodging susceptibility and water deficiency, because the level of these stresses is variable among the environments and the responses of the genotypes are different.

The results of the regression coefficient on the genotypic covariate *SD* could be clarified considering the sum of temperatures between sowing and harvest, in each location, although this covariate was not statistically effective in an exhaustive model. We must notice that when the sum of temperatures was alone, it was a significant environmental covariate for explaining the



**Table V.** Values of the covariates of the lines and of the environments, and their associated regression coefficients, from the factorial regression and the groups obtained from cluster analysis.

<i>Lines</i>	<i>Covariate SD</i>	<i>Covariate LS</i>	<i>Regression coefficient on RF</i>	<i>Groups</i>
LH74	5.54	-0.59	-0.004	I
F271	-4.11	0.29	0.001	II
Co255	-1.11	-0.42	0.006	II
F252	-1.28	-0.71	-0.001	III
F288	-0.41	-0.94	-0.002	III
F244	-0.78	-0.75	0.002	III
F113	1.19	0.47	-0.001	IV
F1772	3.35	1.34	0.008	V
W33	-1.33	0.08	-0.002	VI
Co125	-1.05	1.27	-0.008	VI

<i>Environments</i>	<i>Regression coefficient on SD</i>	<i>Regression coefficient on LS</i>	<i>Covariate RF</i>	<i>Groups</i>
102	-0.17	0.46	87.4	I
209	-0.26	0.81	-49.6	I
217	-0.16	0.25	3.9	I
108	-0.41	0.38	-53.4	I
107	0.09	0.22	63.7	II
109	-0.04	0.03	3.6	II
203	-0.09	0.22	78.4	II
216	-0.12	0.52	83.9	II
122	0.25	0.56	71.9	III
104	0.09	-0.71	9.9	IV
113	0.20	-0.15	42.4	IV
116	0.08	-0.59	107.0	IV
207	-0.01	-0.45	0.5	IV
119	-0.18	-0.47	-24.0	IV
206	0.14	-1.50	0.5	V
112	0.10	0.45	-70.5	VI
202	0.20	0.21	-63.6	VI
218	-0.04	0.11	-47.2	VI
208	0.16	0.19	-91.5	VI
205	0.15	-0.09	-89.6	VII
220	0.02	-0.45	-63.6	VII

The first figure of environment number corresponds to the year (1: 1992; and 2:1993), the last 2 correspond to the number of the location. The 6 groups of lines and the 7 groups of environments were obtained from pattern analysis, using the coefficient of determination to stop the clustering process.

interaction, but that it lost its significance when it was taken together with the genotypic mid-silking date covariate, because of redundancy. Nevertheless, climatic conditions were worth considering. Some environments with high positive regression coefficients on the genotypic covariate *SD* (for example, environment 122) accentuated the effects of earliness, probably because the climatic conditions at the end of summer, in these environments (warmer with sufficient rainfall),

were favorable to the ripening of the latest lines. In contrast, we observed that the environments 108 and 209 had the highest negative regression coefficients. Thus, the latest lines were particularly disadvantaged there. These environments were characterized by the lowest sums of temperatures between sowing and harvest. Therefore, part of the line x environment interaction was due to the latest lines that displayed reduced yield in cold environments and, on the

**Table VI.** Variance analysis of structuration.

	SS	df	F	%SS <sub>i</sub>
Environment ( $e_j$ )	3 932.0	20	114.6	
Line ( $l_i$ )	862.2	9	55.8	
Interaction ( $l_{ij}$ )	651.6	180	3.6	100
$(LE)_{k(i)h(j)}$	383.7	30	7.5	59
$(Le)_{k(i)j}$	150.6	70	1.3 ns	23
$(lE)_{ik(j)}$	34.4	24	0.8 ns	5
$(le)_{ij}$	82.7	56	0.9 ns	13
Error	926.9	540		

Sums of squares (SS), degrees of freedom (df), F-ratios and percentage of interaction sum of squares accounted for by interaction modelling (%SS<sub>i</sub>). F-ratios are significant at the 1% probability level, except for those marked ns.

contrary, were advantaged in warm environments. This result was also found on wheat by Baril (1992).

The regression coefficients on the genotypic covariate *LS* revealed environments in which a large amount of selective lodging occurred (for example, environments 206 and 104) and environments with proportionally smaller lodging (for example, environments 209 and 122). The importance of lodging susceptibility in explaining the line x environment interaction for biomass yield, actually comes from the measure of yield. It is biased by severe lodging for some genotypes in some environments, since the harvests were generally made without manual straightening of the plants. This could have been suspected because: i) during the 2 years 1992 and 1993, a large amount of lodging occurred following severe local windstorms; ii) the locations were very distant from one another and thus great differences in average lodging were to be expected; and iii) among the 10 lines studied a large variability for lodging susceptibility existed.

The values of the regression coefficients on the environmental covariate *RF* reveal that some lines such as F1772, Co255 and F244 seemed to either be more susceptible to water deficiency during the critical period around flowering or to take better advantage of natural water availability. Therefore, the sum of rainfalls during the period of particular sensitivity of maize towards water deficiency (June, July and August) explained a significant part of the line x environment interaction for yield. This is in agreement with the work of Mohammad Saeed and Francis (1984) on grain sorghum yield. However, Kang and

Gorman (1989) and Magari and Kang (1993), in the USA, found that pre-season rainfall and rainfall during the growing season explained a very low amount of the interaction for grain maize yield.

From these results, it was noticed that 2 years of experiment in the same location always differed from one another. The absence of evident structure of environmental multiplicative parameters quoted previously in the results suggests that the effect of years on the magnitude of interaction could be as important as the effect of locations.

### Comparison of the models

The 4 models used in the present study, joint regression, biadditive models, factorial regression, and cluster analysis, accounted for 6, 56, 55 and 59% of the interaction sum of squares using 9, 54, 47, and 30 interaction degrees of freedom respectively (table VII) The adjusted  $F^2$  value provides an estimator of the variation explained, adjusted for the number of parameters in the model. This value allowed us to rank the models as joint regression, biadditive model, factorial regression and cluster analysis, respectively, from the least to the most efficient model (table VII). However, it must be pointed out that this statistic is probably biased and overestimated for the non-linear biadditive and clustering models.

The use of the joint regression on the estimate of the main environmental effect did not allow a satisfactory description of the data in our study,

probably because this model assumes a linear response of the lines to the biological potential of the environments. This oversimplification of the joint regression model was previously mentioned by Hill (1975).

The 3 other models were able to significantly account for line x environment interaction, explaining more than half of the interaction sum of squares and the residual term was non-significant (at the 0.01 probability level) (table VII). In addition, the number of parameters included in the cluster analysis model was smaller than the others, and so this model showed a better effectiveness, regarding the adjusted  $R^2$  criterion (table VII). However, as already suggested, the adjusted  $R^2$  calculation favored non-linear models, such as cluster analysis, over linear models. Therefore, the effectiveness of cluster analysis must be tempered. On the other hand, these 3 methods have their own advantages and disadvantages, but do not appear as independent regarding the interpretation of the results.

Factorial regression is specific, since it includes extra information (*ie* the covariates). A number of types of covariates can be considered. Some can be derived from the traits studied, others are new characters recorded on the genotypes which could be involved in the trait studied, or data describing the environments (climatic data). In our case, data on earliness, lodging susceptibility of the genotypes in the environments and environment climatic characteristics such as rainfall, mainly explained the line x environment interaction for biomass yield. In this sense, factorial regression was more effective for the biological interpretation of the interaction. It allowed us to reveal some key factors of adaptation to the environment for maize: earliness (even though it is probably overestimated here) in relation to the temperature regime, susceptibility to lodging, and

sensitivity to water deficiency. It could therefore enable us to characterize the response of the genotypes to variable environments and assist in the choice of experimental locations so that they can better reveal possible defects of the evaluated genotypes.

Biadditive models are often considered as good models to partition the genotype x environment interaction (Crossa *et al*, 1990). The biplot display of parameters is also very useful in that it helps visualize the overall pattern of the data as well as the genotype x environment interaction, both in terms of the main effects and multiplicative components. Nevertheless, the interpretation of the parameters provided by biadditive models is not always obvious. It is made easier by the knowledge of additional information on genotypes and environments. Some associations between the results of biadditive models and those of factorial regression were thus highlighted. Lines (*eg*, LH74, F1772, Co125, F271) and environments (*eg*, 206, 104, 108, 209) with maximum and minimum MT1s in the biadditive model, were characterized by extreme values of some covariates and associated regression coefficients, estimated by factorial regression. The parameters computed from the biadditive model give an estimation of the contribution of the lines to the interaction for biomass productivity, and highlight lines of a high stability. Similar conclusions could be drawn from the environmental multiplicative parameters which could be used to discard non-interactive locations.

Cluster analysis is a relevant tool to classify genotypes and environments in order to decompose and interpret genotype x environment interaction. A crucial point of this method is to determine the criterion that allows us to cut the cluster procedure. We used the coefficient of determination, and chose to stop the clustering process

**Table VII.** Comparison of the different modellings in accounting for the line x environment interaction.

	SS	df	%SS <sub>i</sub>	adjR <sup>2</sup>
Interaction	651.6	180	100	
Joint regression	38.1	9	6	0.02
Biadditive model	365.8	54	56	0.38
Factorial regression	359.0	47	55	0.40
Structuring	383.7	30	59	0.51

Sums of squares (SS), degrees of freedom (df), percentage of interaction sum of squares accounted for by interaction modelling (%SS<sub>i</sub>), adjusted coefficient of determination  $R^2$  (adjR<sup>2</sup>).

when this was greater than 0.95. However, it led to a too large number of groups found (7 groups of environments, 6 groups of lines). The difficulty of interpretation of so many units together leads us to consider the possibility of obtaining fewer groups. *A priori* data on earliness, pedigree, or usual agronomic traits did not really help interpret the groups obtained. These characteristics depend little on the environment and have strong effects on yield. Such information is more likely to explain a part of main line effect. However, it could be observed that the groups corresponded to lines which were not very different regarding their multiplicative parameters. The interpretation of environment groups was made easier by the knowledge of additional information on the environments, because some associations between the results of clustering and those of factorial regression could be highlighted. The 2 groups including a single environment (122 and 206) corresponded to the environment with by far the highest positive regression coefficient on the genotypic covariate *SD* and to the environment with by far the highest negative regression coefficient on the genotypic covariate *LS*, respectively. The first group was composed of environments that displayed both positive regression coefficients on the covariate *LS* and negative regression coefficients on the covariate *SD*. The second group included environments which exhibited positive regression coefficients on the genotypic covariate *LS*, whereas in the fourth group, these coefficients were negative. The environments of the sixth group displayed high negative environmental covariate *RF* together with positive regression coefficients on the genotypic covariate *LS*. Therefore, some biological connections could be established between the results of the pattern analysis, factorial regression, and biadditive models. From a breeding point of view, the clustering method enables us to assess whether all the environments of the experimental network are really relevant to account for the genotype x environment interaction or if some of them could be removed. In fact, in our study we showed that the number of environments could be considerably reduced without losing too much information on the interaction for biomass yield. This could lead to rational savings in breeding programmes. However, our work clearly pointed out the magnitude of year effect in the interaction. As a consequence, the uncontrollable year factor should not be included in highly performing trial networks. It is a major limitation in the generalization of our conclusions.

## CONCLUSION

Joint regression failed to describe the line x environment interaction satisfactorily, probably because of an oversimplification. However, the other 3 methods used in this study (biadditive model, pattern analysis and factorial regression) were able to significantly account for the interaction.

In agreement with van Eeuwijk (1992), we have shown that various methods led to similar results and interpretation, and that biological connections could be established between the results of factorial regression, pattern analysis and the biadditive model. However, even though each of these models explained slightly more than half of the interaction, they were not strictly identical, because part of their results did not agree with each other. Therefore, they probably also exhibited some complementarity, which could be very useful.

On the basis of our results, we concluded that the line x environment interaction for biomass dry matter yield in maize could essentially be due to differences in line earliness and lodging susceptibility (and to the different ability of the environments to reveal them), and to differences in environment rainfalls (and to the variable susceptibility of the lines to water stress).

Interaction modelling has been shown to be useful for maize breeding. It enables us to evaluate the contribution of the genotypes to the interaction, to highlight some key factors of adaptation to the environment for maize, and to optimize the construction of the experimental networks.

## ACKNOWLEDGMENTS

The authors are indebted to JB Denis and C Bastien for useful discussions and comments on this manuscript. The authors thank the Ministry of Agriculture for their grant, the members of Promais and INRA research stations who conducted the experiments. We also acknowledge Météo-France for environment climatic data.

## REFERENCES

- Abou-EI Fittouh HA, Rawlings JO, Miller PA (1969) Genotype by environment interactions in cotton — their nature and related environmental variables. *Crop Sci* 9, 377-381

- Baril C (1992) Factorial regression for analyzing genotype x environment interaction in bread wheat trials. *Theor Appl Genet* 83, 1022-1026
- Baril C, Denis JB, Brabant P (1994) Selection of environments using simultaneous clustering based on genotype x environment interaction. *Can J Plant Sci* 74, 311-317
- Corsten LCA, Denis JB (1990) Structuring interaction in 2-way ANOVA tables by clustering. *Biometrics* 46, 207-225
- Crossa J, Gauch HG, Zobel RW (1990) Additive main effects and multiplicative interaction analysis of 2 international maize cultivar trials. *Crop Sci* 30, 493-500
- Decoux G, Denis JB (1991) INTERA : logiciels pour l'interprétation statistique de l'interaction entre 2 facteurs. Technical Report, Laboratoire de Biométrie, INRA-Versailles, France, 175 p
- Denis JB (1980) Analyse de la régression factorielle. *Biom Praxim* 20, 1-34
- Denis JB (1988) Two-way analysis using covariates. *Statistics* 19, 123-132
- Denis JB, Vincourt P (1982) Panorama des méthodes statistiques d'analyse des interactions genotype x milieu. *agronomie* 2, 219-230
- Denis JB, Gower J (1992) Biadditive models. Report, Laboratoire de Biométrie, INRA-Versailles, France, 33 p
- Dhillon BS, Gurrath PA, Zimmer E, Wermke M, Pollmer WG, Klein D (1990) Analysis of diallel crosses of maize for variation and covariation in agronomic traits at silage and grain harvest. *Maydica* 35, 297-302
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in plant-breeding program. *Aust J Agric Res* 14, 742-754
- Freeman GH (1973) Statistical methods for the analysis of genotype x environment interactions. *Heredity* 31, 339-354
- Geiger HH, Melchinger AE, Schmidt GA (1986) Analysis of factorial crosses between flint and dent maize inbred lines for forage performance and quality traits. In: *Breeding of Silage Maize* (O Dolstra, P Miedema, eds), Pudoc, Wageningen, Germany, 147-154
- Gollob HF (1968) A statistical model with combined features of factor analytic and analysis of variance techniques. *Psychometrika* 33, 73-115
- Hardwick RC, Wood JT (1972) Regression methods for studying genotype x environment interactions. *Heredity* 28, 209-222
- Haun JR (1982) Early prediction of corn yields from daily weather data and single predetermined seasonal constants. *Agric Meteorol* 27, 191-207
- Kang MS, Gorman DP (1989) Genotype x environment interaction in maize. *Agron J* 81, 662-664
- Hill J (1975) Genotype x environment interaction — a challenge for plant breeding. *J Agric Sci (Camb)* 85, 447-493
- Judge GG, Griffiths WE, Hill RC, Lee T (1980) *The Theory and Practice of Econometrics*. John Wiley and Sons, Inc, New York, USA
- Magari R, Kang MS (1993) Genotype selection via a new yield-stability statistic in maize yield trials. *Euphytica* 70, 105-111
- Mandel J (1971) A new analysis of variance model for non-additive data. *Technometrics* 13, 1-18
- Mohammad Saeed, Francis CA (1984) Association of weather variables with genotype x environment interactions in grain sorghum. *Crop Sci* 24, 13-16
- Tukey JW (1949) One degree of freedom for non-additivity. *Biometrics* 5, 232-242
- van Eeuwijk FA (1992) Interpreting genotype-by-environment interaction using redundancy analysis. *Theor Appl Genet* 85, 89-100
- Vattikonda MR, Hunter RB (1983) Comparison of grain yield and whole-plant silage production of recommended corn hybrids. *Can J Plant Sci* 63, 601-609
- Westcott B (1986) Some methods of analysis genotype x environment interaction. *Heredity* 56, 243-253
- Wood JT (1976) The use of environmental variables in the interpretation of genotype x environment interaction. *Heredity* 37, 1-7
- Yates F, Cochran WG (1938) The analysis of groups of experiments. *J Agric Sci (Camb)* 28, 556-580