Theor Appl Genet (1992) 83:1022–1026

Factor regression for interpreting genotype-environment interaction in bread-wheat trials

C. P. Baril

Research Station for Plant Genetics, INRA-CNRS, University of South Paris, Ferme du Moulon, F-91190 Gif-sur-Yvette, France

Received May 9, 1991; Accepted October 1, 1991
Communicated by A. R. Hallauer

Summary. The French INRA wheat (*Triticum aestivum* L. em Thell.) breeding program is based on multilocation trials to produce high-yielding, adapted lines for a wide range of environments. Differential genotypic responses to variable environment conditions limit the accuracy of yield estimations. Factor regression was used to partition the genotype-environment (GE) interaction into four biologically interpretable terms. Yield data were analyzed from 34 wheat genotypes grown in four environments using 12 auxiliary agronomic traits as genotypic and environmental covariates. Most of the GE interaction (91%) was explained by the combination of only three traits: 1,000-kernel weight, lodging susceptibility and spike length. These traits are easily measured in breeding programs, therefore factor regression model can provide a convenient and useful prediction method of yield.

Key words: Wheat – Genotype-environment interaction – Prediction – Yield trials – Factor regression

Introduction

A general feature of yield trials is the occurrence of a considerable and complex genotype-environment (GE) interaction. This phenomenon affects the amount of progress made during selection. The usefulness of yield data greatly depends on the accuracy of predicting yield.

The relative inadequacy of analysis of variance, principal components analysis (PCA) and linear regression (Mandel 1961; Finlay and Wilkinson 1963) has previously been illustrated (Zobel et al. 1988). A better predictive model seems to be the FANOVA (for factor ANOVA as denoted by Gollob 1968), which includes both additive and multiplicative components. This more general model, also called the AMMI model (as Additive Main effects and Multiplicative Interaction, Gauch 1988) is equivalent to the PCA method applied to the residual term of the additive model (Mandel 1969) and provides a powerful statistical tool for analysis of two-way data sets (Mandel 1971; Gabriel 1971; Bradu and Gabriel 1978; Kempton 1984; Crossa et al. 1990).

Yield trials conducted with many genotypes grown in several environments are usually combined with the measurement of agronomic traits. The wheat (*Triticum aestivum* L. em Thell.) testing program of INRA (France) requires the evaluation of lines selected in experimental research stations across a large range of environments. The objective of this study is to use the information on agronomic traits in a factor regression model, in order to provide some elements of a biological explanation of GE interaction for yield.

Materials and methods

Experimental data

A series of 34 winter wheat genotypes, grown in four environments in northern France in 1990, was used to assess GE interaction for yield. The lines were tested both in an extensive and an intensive yield trial in two locations (Le Moulon and La Minière). A randomized complete block design with two replications per agronomic treatment was used for all analyses. Plant density was higher in the yield trials at Le Moulon (300 plants/m^2) than in the yield trials at La Minière (280 plants/m^2). In Le Moulon, the difference between the two agronomic treatments was the amount of growth regulators and additive nitrate fertilizers applied in the intensive trial (equivalent to the highest national level of intensification). In addition, both intensive and extensive trials were protected by fungicides. In La Minière, the only difference between the two treatments was the fungicidal supply in the intensive trial. Each type of treatment applied in

each location was considered as one environment giving then four different environments.

Apart from grain yield (GY in q/ha), ten auxiliary traits were measured: heading date (HD in number of days), plant height (PH), spike weight (SW), spike length (SL), kernel weight per spike (KW/S), kernel number per spike (KN/S), spikelet number per spike (SLN/S), flower number per spike (FN/S), lodging susceptibility (LS noted from 1 to 9) and 1,000-kernel weight (TKW). All of the yield components were measured from a random sample of ten spikes per plot. Moreover, brown rust (*Puccinia recondita*) susceptibility (BRS noted from 1 to 9) and spike number per plant (SN/P) were evaluated in the two extensive yield trials.

In order to build an optimal prediction model for grain yield variability, all of the other measured traits were used to define the relevant covariates explaining grain yield. A new method of definition of covariates is proposed in the following section.

Statistical analysis

A preliminary ANOVA carried out on the traits measured in the four environments revealed significant ($P=0.001$) GE interaction for grain yield and 1,000-kernel weight only. All of the following analyses were performed using the means of the two replications.

Unlike covariance analysis in which the covariates depend on the two factors, factor regression allows the partitioning of GE interaction into functions (see model 3 below) of only one factor each multiplied by a regression coefficient depending on the other factor. A breakdown of the GE interaction for yield requires the definition of functions (covariates) and the identification of the most efficient subset of covariates that can explain GE interaction.

Definition of covariates. The definition of covariates aims to split each agronomic trait into parameters dependant on only one factor. In the first step, the additive model was applied to each of the first nine agronomic traits to estimate additive parameters. The additive model is:

$$Y_{ij}=\mu+\alpha_i+\beta_j+R_{ij} \qquad (1)$$

where Y_{ij} is the mean for the trait measured in the two replications of the *ith* genotype in the *jth* environment; μ is the grand mean; α_i is the genotype mean deviation; β_j is the environment mean deviation; and R_{ij} is the residual. For each agronomic trait the genotypic additive parameter (α_i) defines the genotypic covariate, while the environmental additive parameter (β_j) defines the environmental covariate. The additive model has $I+J-1$ degrees of freedom (*df*) where I is the number of genotypes and J is the number of environments.

The multiplicative model was applied to 1,000-kernel weight because its GE interaction was significant in the ANOVA. The multiplicative model is:

$$Y_{ij}=\mu+\alpha_i+\beta_j+\lambda_i\cdot\theta\cdot\delta_j+r_{ij}\,. \qquad (2)$$

The decomposition of the additive model residual by multiplicative structuralization increases the global parametrical dimension of $I+J-3$ *df*. The additive (α_i) and multiplicative (λ_i) genotypic parameters are considered as genotypic covariates, while the additive (β_j) and multiplicative (δ_j) environmental parameters are considered as environmental covariates. θ is a scale parameter that does not include any factor effect (like μ in the additive part).

Optimal model construction. The factor regression model fitted to grain-yield data was built by progressive addition of the most significant covariates explaining yield variation to the basic additive model. The stepwise process proposed by Denis (1988) was applied to yield data using the following model:

$$Y_{ij}=\mu+\alpha_i+\beta_j+\sum_h \alpha'_{ih}\cdot E_{hj}+\sum_k \beta'_{jk}\cdot G_{ki}+\lambda_i\cdot\theta\cdot\delta_j+\varepsilon_{ij} \qquad (3)$$

where $k=1,\ldots,K\leq I$ and $h=1,\ldots,H\leq J$ are the number of genotypic and environmental covariates, respectively; α'_{ih} is the genotypic regression coefficient on environmental covariates (E_{hj}); and β'_{jk} is the environmental regression coefficient on genotypic covariates (G_{ki}). Taken together, the two sums of terms including the genotype and environmental covariates bring $K(I-1)+H(J-1)-HK$ *df*. The term ε_{ij} represents the residual.

A covariate is declared significant at a given probability level if its mean square exceeds the mean square of the sum of the multiplicative ($\lambda_i\cdot\theta\cdot\delta_j$ with $I+J-3-H-K$ *df*) and the residual (ε_{ij}) terms by the appropriate *F*-value. At each step in the analysis the significance of the multiplicative term relative to the residual term was tested.

After finding the best single covariate model from among all of the possible single covariate models ($K+H=1$), the best two-covariate model was looked for, knowing the first covariate, and so forth until the addition of the first multiplicative term brought no more significant information. Finally, a simple biological interpretation of the model obtained was provided. All these analyses were performed using the computer package INTERA (Decoux and Denis 1991), which provides least-squares estimates of parameters.

Results

F-tests were carried out using all of the genotypic and environmental covariates defined in the Materials and methods. The significance of the introduction of each covariate into the additive model applied to yield data was tested with respect to the residual term of the corresponding linear single covariate model. Half of the traits gave significant *F*-test results (Table 1) for either the genotypic or the environmental parameter. Nevertheless, for each covariate the remaining multiplicative term of the GE interaction of model (3) was highly significant ($P=0.001$).

The multiplicative parameter of 1,000-kernel weight (TKW) was the only covariate that explained a large part ($P=0.001$) of grain-yield performance as both a genotypic and an environmental covariate. When the genotypic covariates were considered, the additive parameter for LS and the genotypic mean for BRS provided valuable explanations of grain-yield variation at the 0.001 and 0.01 probability levels, respectively. When the environmental covariates were taken into consideration, the additive parameters for TKW, KN/S and SW yielded significant information at the same probability level ($P=0.01$). The contribution of the environmental additive parameters for SL and KW/S was weak ($P=0.05$) when they were used as first covariates.

Therefore, the series of *F*-tests for all of the possible single covariate models allowed the retention of the genotypic additive parameter of LS in the first step. At each

Table 1. *F*-test results for each covariate introduced into a one-covariate model to explain grain yield. The genotypic and environmental covariates are derived from genotypic means and from multiplicative and additive parameters of auxiliary traits

Trait[a]	Genotypic covariate	Environmental covariate
	Genotypic mean	
BRS	4.6**	
	Multiplicative parameter	
TKW	7.9***	2.8***
	Additive parameters	
LS	8.8***	NS
TKW	NS	2.5**
KN/S	NS	2.4**
SW	NS	2.1**
SL	NS	1.9*
KW/S	NS	1.9*

*, ** and *** significant at the 0.05, 0.01 and 0.001 probability levels, respectively; NS, non-significant

[a] BRS, Brown rust susceptibility; TKW, 1,000-kernel weight; LS, lodging susceptibility; KN/S, kernel number per spike; SW, spike weight; SL, spike length, KW/S, kernel weight per spike

step, one covariate was selected from among the remaining ones, with the constraint of retaining the covariates that had already been introduced at steps previous.

Optimal model construction

The second covariate selected was the genotypic multiplicative parameter for TKW, while the genotypic mean of BRS ranked second for the second covariate. The third covariate selected was the genotypic additive parameter for SL, and the fourth one was the environmental multiplicative parameter for TKW.

The addition of this fourth covariate led to the threshold at which the multiplicative term of GE interaction lost any significance. Thus, it corresponds to the optimal model because the GE interaction term becomes unnecessary. However, the genotypic additive parameter for SLN/S was the last covariate to bring any additional significant explanation of grain yield to the optimal model ($P=0.05$).

With the order of the genotypic covariates known the ANOVA was performed (Table 2) according to the procedure proposed by Denis (1988). This analysis consists of progressively removing the less efficient covariates until the additive model is obtained. The accuracy of the *F*-tests, which were always based on a real error, leads to the preferred usage of this backward ANOVA over the forward ANOVA, which entails the progressive addition of more efficient new covariates, including all those already introduced.

Table 2. ANOVA of the optimal model with one multiplicative term

Source of variation[a]	*df*	Sum of squares	Mean squares	*F*-test	*P*-level ($\times 10^{-3}$)
Environment (E)	3	7,707	2,569	80.3	0*
Genotype (G)	33	10,877	330	10.3	0*
G: a.LS	3	2,138	713	22.3	0*
G: m.TKW	3	1,694	565	17.7	0*
G: a.SL	3	558	186	5.8	3*
E: m.TKW	30	2,904	97	3.0	2*
MT1	31	1,666	54	1.7	82
Residual	29	927	32		

* Significant at $P \leq 0.05$

[a] G and E, Mean genotypic and environmental covariates, respectively; a and m, mean additive and multiplicative parameters, respectively; MT1, the first multiplicative term; LS, lodging susceptibility; TKW, 1,000-kernel weight; SL, spike length

Biological interpretation

The additive parameters used as either genotypic or environmental covariates are intuitively understood to reflect a sort of intrinsic value of either the genotypes or the environments. However, the multiplicative parameters, which are assessed by model (2) adjusted to the TKW, are not directly interpretable because environmental multiplicative parameter depends on the genotypic multiplicative parameter which is structurally associated to it. The observation of significant correlations between both the environmental covariate of the optimal model and all of the other environmental covariates and between the genotypic covariates of the optimal model and all of the other genotypic covariates provides us with some elements of interpretation (Table 3).

The environmental multiplicative parameter (δ_j) for TKW is highly correlated with the environmental additive parameter (β_j) for TKW ($\varrho=0.96$ at $P=0.005$). This multiplicative parameter is consistently highly correlated with the additive parameter of KN/S. The genotypic multiplicative parameter (λ_i) for the TKW is weakly correlated with the genotypic additive parameter (α_i) of TKW ($\varrho=0.47$ at $P=0.001$) and with the genotypic BRS ($\varrho=0.43$ at $P=0.005$). The genotypic additive parameter for SL is also correlated with the genotypic mean of BRS. With regard to the multiplicative parameters, only the absolute value of the correlation coefficients can be interpreted.

The second two-covariate model obtained with the step-by-step procedure included LS and the BRS as genotypic covariates. Since the genotypic mean of BRS is correlated with the genotypic multiplicative parameter of TKW and the genotypic additive parameter of SL, BRS was removed from subsequent models.

Table 3. Correlation coefficients and their probability level ($P \leq 0.05$) between the optimal model covariates and all the others

Traits[a]	Correlation coefficient	Probability level
Environmental covariate: (2 *df* Student's *t*-test)		
(mTKW,aKN/S)	$\varrho = -0.97$	0.001
(mTKW,aTKW)	$\varrho = +0.96$	0.005
(mTKW,aSW)	$\varrho = -0.92$	0.010
(mTKW,aKW/S)	$\varrho = -0.91$	0.010
(mTKW,aSL)	$\varrho = +0.87$	0.025
Genotypic covariates: (32 *df* Student's *t*-test)		
(aLS,aPH)	$\varrho = +0.37$	0.025
(mTKW,aTKW)	$\varrho = +0.47$	0.001
(mTKW,BRS)	$\varrho = +0.43$	0.005
(mTKW,SN/P)	$\varrho = +0.31$	0.050
(mTKW,aHD)	$\varrho = -0.31$	0.050
(aSL,BRS)	$\varrho = -0.33$	0.025
(aSL,SN/P)	$\varrho = -0.33$	0.025
(aSL,aSLN/S)	$\varrho = +0.31$	0.050

[a] a and m, Mean additive and multiplicative parameters, respectively; TKW, 1,000-kernel weight; KN/S, kernel number per spike; SW, spike weight; KW/S, kernel weight per spike; SL, spike length; LS, lodging susceptibility; PH, plant height; BRS, brown rust susceptibility; SN/P, spike number per plant; HD, heading date; SLN/S, spikelet number per spike

Discussion

The presence of GE interaction is the main obstacle in the estimation of genotypic yields from multilocation trials. The factor regression model provides an interesting partitioning of the GE interaction on yield data into a sum of linear functions of genotypic and environmental covariates. These covariates are derived from agronomic traits frequently measured in yield trials conducted by plant breeding programs and have the advantage of providing biological explanations of GE interaction for yield. In order to remove unnecessary terms from the predictive model, one can consider that the optimal model corresponds to the threshold at which the multiplicative term becomes non-significant relative to the residual term.

Based on these results from wheat trials, one can conclude that GE interaction is essentially due to yield-limiting factors occurring at the grain-filling stage. One could hazard the conjecture that the multiplicative parameter for TKW could be thought of as reflecting some physiological traits like grain-filling kinetics. Lodging was particularly widespread in 1990 and occurred during grain maturation. Hence, lodging susceptibility, though having no correlation with the 1,000-kernel weight, reduced yield, probably by acting on the kernel number per square meter.

One could argue that it may have been more biologically meaningful to retain the genotypic mean of BRS even if the statistical approach led to the elimination of this covariate ($P = 0.01$ instead of $P = 0.001$ for the genotypic multiplicative parameter of TKW). But the step-by-step procedure performed with the two covariates model, including the genotypic additive parameter of LS and the genotypic mean of BRS until the treshold at which the first multiplicative term lost any significance, leads to a five covariates model. This model includes the following three additional covariates: the genotypic multiplicative parameter of TKW, the genotypic additive parameter of SLN/S and the environmental multiplicative parameter of TKW. Hence, given the data used in this study the optimal model seems to be the four covariates model presented previously. The present results stress the importance of genotypic BRS for interpreting GE interaction for yield via the multiplicative parameter for TKW. This observation fits in with the covariance analysis performed by Baker (1971), which illustrated that genotype-site interaction was due primarily to a simply inherited trait, such as rust resistance.

As a basis for comparing the factor regression model with the AMMI model, it should be noted that the latter first fits additive effects for genotypes and environments and then fits multiplicative effects for GE interaction by principal components analysis (PCA). The first PCA axis usually explains the largest fraction of the GE interaction and often appears to be sufficient (Gauch and Zobel 1988; Zobel et al. 1988; Crossa et al. 1991). As a consequence, the multiplicative modeling in the factor regression method could be reduced to the first multiplicative term without significant loss of information. Indeed, the determination coefficient, which is the proportion of the variability explained by the AMMI model that includes only one multiplicative term relative to the whole variability, equals 85% in this example. Nevertheless, the determination coefficient of the model splitting the GE interaction (which makes up 35% of the whole variability of yield) into a linear combination of relevant agronomic covariates is equal to 91% without the first multiplicative term (MT1) and to 97% when MT1 is included into the factor regression model.

Even if the difference between the AMMI model and the factor regression model cannot be statistically tested, given these two models are not nested, in addition to improving the prediction of yield stability, this alternative method especially provides a direct biological interpretation of GE interaction.

Acknowledgements. The author thanks Dr. J. B. Denis for his guidance in biometrical analysis and has appreciated helpful comments and suggestions for improving the manuscript from Dr. P. Brabant, Dr. J. B. Denis, Dr. A. Gallais and the referees. This work was supported by the cereal unit of INRA (led by Dr. G. Doussinault) and the French public company Agri-Obtentions.

References

Baker RJ (1971) Effects of stem rust and leaf rust of wheat on genotype-environment interaction for yield. Can J Plant Sci 51:457–461

Bradu D, Gabriel KR (1978) The biplot as a diagnostic tool for models of two-way tables. Technometrics 20:47–68

Crossa J, Gauch HG, Zobel RW (1990) Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Sci 30:493–500

Crossa J, Fox PN, Pfeiffer WH, Rajaram S, Gauch HG (1991) AMMI adjustment for statistical analysis of an international wheat yield trial. Theor Appl Genet 81:27–37

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

Decoux G, Denis JB (1991) INTERA. Logiciels pour l'interpretation statistique de l'interaction entre deux facteurs. Biométrie INRA route de Saint. Cyr F 78026 Versailles France: 175 p

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

Gabriel KR (1971) The biplot graphic display of matrices with application to principal components analysis. Biometrika 58:453–467

Gauch HG (1988) Model selection and validation for yield trials with interaction. Biometrics 88:705–715

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

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

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

Mandel J (1961) Non-additivity in two-way analysis of variance. Am Stat Assoc J 56:878–888

Mandel J (1969) The partitioning of interaction in analysis of variance. J Res Nat Bur Stand Sect B 73B:309–328

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

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