Euphytica **82:** 149–155, 1995.

Analysing genotype by environment interaction in Dutch potato variety trials using factorial regression *

C.P. Baril[1], J.-B. Denis[2], R. Wustman[3] & F.A. van Eeuwijk[4]
[1] *CIRAD-Foret, 94736 Nogent-sur-Marne Cedex, France;* [2] *INRA, F78026 Versailles Cedex, France;* [3] *PAGV, P.O. Box 430, 8200 AK Lelystad, The Netherlands;* [4] *CPRO-DLO, P.O. Box 16, 6700 AA Wageningen, The Netherlands*

Received 28 February 1994; accepted 16 February 1995

Key words: AMMI, biadditive model, factorial regression, multiplicative interaction, potato, variety trials, *Solanum tuberosum*

Summary

Genotype by environment interaction was investigated for yield data from the official Dutch Variety List trials for potato. The data set included 64 genotypes by 26 environments, where environments consisted of year by soil type combinations. Factorial regression models incorporating genotypic and environmental covariates in the interaction were used to analyse the data. The merits of factorial regression models were compared with those of biadditive models. Factorial regression models and biadditive models described comparable amounts of interaction, but factorial regression models provided a better basis for biological interpretation of the interaction.

Introduction

As part of the research programme concerned with the compilation of the Dutch Variety List of Field Crops a large number of potato genotypes are evaluated every year under circumstances thought representative of Dutch growing conditions. In the evaluation of the trials the phenomenon of genotype by environment interaction constitutes a recurring problem. Few attempts have been undertaken to analyse this interaction for the Dutch Variety List trials in general. In this paper genotype by environment interaction for yield in potato will be analysed. In companion papers of Kroonenberg, Basford & Ebskamp (1995), and van Eeuwijk, Keizer & Bakker (1995) variety by environment interaction in the Dutch Maize Variety Trials will be analysed. The method of analysis that will be used for the potato data is that of factorial regression (Denis, 1988; Baril, 1992), because this method seems the most suitable one for arriving at biologically interesting interpretations. Characteristic for factorial regression is that it incorporates genotypic and environmental covariates in the description of the genotype by environment interaction. The results obtained by factorial regression will be compared with those from the application of biadditive models (Denis & Gower, 1992, 1994). The latter may be better known under the name of AMMI models (Gauch, 1988). The utility of biadditive models has been amply demonstrated (Mandel, 1971; Bradu & Gabriel, 1978, Kempton, 1984; Crossa, Gauch & Zobel, 1990). Nevertheless, this model has the disadvantage that it is unable to accomodate additional information on genotypes and environments. We will show that the incorporation of genotypic and environmental covariates in factorial regression enhances biological interpretations of the genotype by environment interaction.

Material

To investigate the biological mechanisms underlying genotype by environment interaction for yield of potato in the Dutch Variety Trials, a selection of yield data was made covering 64 genotypes over a period of 16 years. During most, but not all, years the genotypes were grown at both sand and clay. Combinations of year and soil type, 26 in number, constituted the environmental dimension of a 64 by 26 genotype by environment table of mean yields. Though the origi-

* This article was previously published in Euphytica 82: 149–155.

nal experiments were replicated, only the means were available for analysis. From the additional information on the genotypes three covariates were selected on basis of a priori biological subject knowledge as possibly useful for the description of genotype by environment interaction. These genotypic covariates were: (1) a classification of genotypes indicating consumption or starch type (CS); (2) a rating on a 1 to 10 scale for early maturity (EM); (3) a rating on a 1 to 9 scale for leaf development (LD). The values for the genotypic covariates were obtained from the Variety List. For the environments five environmental covariates were selected on a priori grounds: (1) whether the soil type was sand or clay (S/C); (2) the number of frost days in the first half of April (F1); (3) the number of frost days in the second half of April (F2); (4) mean temperature over the growing season from April till September in °C (TP); (5) total radiation over the growing season corrected for light interception in J/cm^2 (RD). All covariates were centred.

From the preliminary analyses it was concluded that interaction occurred mainly on sand. Therefore, eight new environmental covariates were added, constructed from the original five. They represent the combined action of soil type with each of the remaining four environmental variables. The variables F1S, F2S, TPS, and RDS contain the values of respectively F1, F2, TP, and RD on sand, and the value of zero on clay. The variables F1C, F2C, TPC, and RDC were derived in the same way for the clay soils. Besides the three genotypic covariates and eight environmental variables defined above, the genotypic main effect (G) and the environmental main effect (E) were included in the list of covariates.

Methods

A simple model for a two-way table of genotypes by environments is the two-way analysis of variance (ANOVA) model with interaction,

$$E[Y_{ij}] = \mu + g_i + e_j + ge_{ij}.$$

On the left hand side we find the expectation of the random variable Y corresponding to genotype i (= 1. . .I) and environment j (= 1. . .J). On the right, μ is the general mean, g_i is the genotypic main effect, e_j the environmental main effect, and ge_{ij} represents the interaction (non-additivity). The two-way ANOVA model with interaction forms a general reference model.

In comparison with the two-way ANOVA model with a separate parameter for each combination of genotype and environment, the formulation for interaction in the biadditive model has been changed to a sum of products of genotypic scores, γ_{ri}, and environmental scores, δ_{rj}, scaled by proportionality constants θ_r. Or, $ge_{ij} = \sum_{r=1}^{r} \theta_r \gamma_{ri} \delta_{rj}$. The scores for genotypes and environments are obtained by a singular value decomposition of the matrix of ge_{ij}'s (Gabriel, 1978). The environmental scores can be interpreted as hypothetical environmental variables, the genotypic scores as sensitivities with respect to these hypothetical variables. The environmental scores have the property of maximizing the differential sensitivity of the genotypes, i.e. describe the maximum amount of interaction. The biadditive model is written in full as

$$E[Y_{ij}] = \mu + g_i + e_j + \sum_{r=1}^{r} \theta_r \gamma_{ri} \delta_{rj}$$

In the factorial regression model covariates can be incorporated for the genotypic as well as the environmental factor (Denis, 1988). The general form for a factorial regression model with K genotypic and H environmental covariates for the interaction reads

$$E[Y_{ij}] = \mu + g_i + e_j + \sum_{k=1}^{K} \sum_{h=1}^{H} \phi_{kh} x_{ki} z_{hj} + \sum_{h=1}^{H} \rho_{hi} z_{hj} + \sum_{k=1}^{K} x_{ki} \tau_{kj}$$

The parameters ϕ_{kh} represent coefficients with respect to cross-products of genotypic covariates, x_k, and environmental covariates, z_h. These coefficients are not dependent on either genotype or environment. They may be interpreted as a general correction for non-additivity (interaction) pertinent to the whole of the genotype by environment table. As such they provide the most simple means to deal with non-additivity. Often more elaborate models for interaction are necessary. The coefficients ρ_{hi} denote genotypic sensitivities to the environmental covariates z_h. The coefficients τ_{kj} denote environmental weighing constants with respect to the genotypic covariates x_k.

The popular regression on the mean or row regression model (Yates & Cochran, 1938; Mandel, 1961; Finlay & Wilkinson, 1963) can be interpreted as a special case of a biadditive model as well as a factorial regression model. The regression on the mean model can be formulated as

Table 1. Analysis of variance for a biadditive model with five terms for interaction. Sum of squares (SS), degrees of freedom (Df), mean square (MS), variance ratio (F), percentage of the interaction sum of squares (% SSI), residual corresponding to a model including this and higher terms, and residual degrees of freedom

Source of variation	SS	Df	MS	F	% SSI	Res. MS	Res. Df
Environment	2661388.7	25	106455.5				
Genotype	4675308.9	63	74211.3				
Interaction	3249550.1	1575	2063.2		100%	2063.2	1575
BT1[1)]	597074.3	87	6862.9	5.8	18.4%	1782.6	1488
joint regression	247512.3	63	3928.8	3.3	7.6%		
remainder	349562.0	24	14565.1	12.4	10.8%		
BT2	446416.5	85	5252.0	4.5	13.7%	1572.4	1403
BT3	358319.0	83	4317.1	3.7	11.0%	1399.8	1320
BT4	260752.7	81	3219.2	2.7	8.0%	1280.9	1239
BT5	223567.5	79	2830.0	2.4	6.9%	1175.4	1160
Residual	1363420.1	1160	1175.4		42.0%		

[1)] BT = biadditive interaction term.

$$E[Y_{ij}] = \mu + g_i + e_j + \beta_i e_j,$$

where the slope β_i can be read as a genotypic sensitivity to the environmental measure e_j. The regression on the mean model is obtained from a biadditive model by allowing only one multiplicative term for the interaction and imposing the constraint on the environmental scores of having to be equivalent to the environmental main effects. It is obtained as a factorial regression model by including only one environmental variable for the interaction, namely the environmental main effect.

For our potato yield data we considered two factorial regressions. Firstly, a factorial regression model was built using all genotypic covariates and environmental covariates described in the Material section. These covariates might be summarized as external covariates, because they are not derived from the genotype by environment table of yield itself. The second factorial regression model was obtained from the first by adding in the genotypic and environmental main effect, and test whether extra interaction was explained by this addition. The genotypic and environmental main effect might be called internal covariates as they are derived from the table itself.

For selection of the variables a stepwise procedure selecting the most significant covariates was used as described by Baril (1992). Each supplementary covariate is chosen to minimize the error mean square until the error stops decreasing. Calculations were performed using the computer package INTERA (Decoux & Denis, 1991).

The decomposition of the interaction degrees of freedom, sum of squares, and the associated model terms of either the biadditive or the factorial regression model can be presented in a two-way table as explained by Denis (1991). For the degrees of freedom for the interaction terms in the biadditive model the following rule due to Gollob (1968) can be used. The r-th biadditive interaction term receives (I - 1) + (J - 1) - (2r - 1) degrees of freedom. In the factorial regression model all ϕ_{kh}'s use one degree of freedom. For each set of genotypic sensitivities, ρ_{hi}, to an environmental covariate, z_h, I-1 degrees of freedom are available, with subtraction of one degree of freedom for each cross-product term $x_k z_h$ that was fitted anteriorly. Analogously, J-1 degrees of freedom are available for each set of environmental weights, τ_{kj}, with substraction of one degree of freedom for each cross-product term fitted before.

Results

Table 1 shows that the sum of squares for genotype by environment interaction amounted to 30.7% of the whole variability in yield. This two-way interaction could be re-expressed as the sum of five biadditive terms (BT1 till BT5), accounting for 58.0% of the interaction with 26.3% of the degrees of freedom, and a residual. This decomposition served as a reference

Table 2. Analysis of variance for factorial regression model. Abbreviations are as in Table 1

Source of variation	SS	DF	MS	F	% SSI
Environment	2661388.7	25	106455.5		
Genotype	4675308.9	63	74211.3		
Interaction	3249550.1	1575	2063.2		
Early maturity *G**	345620.5	25	13824.8	9.3	10.6%
Soil sand/clay *E*	245084.8	62	3953.0	2.7	7.5%
Frost days F1S *E*	246921.2	62	3982.6	2.7	7.5%
Consum./starch *G*	86131.3	23	3744.8	2.5	2.7%
Frost days F2S *E*	182014.5	61	2983.8	2.0	5.6%
Radiation RDS *E*	155625.1	61	2551.2	1.7	4.8%
Temperature TPS *E*	181733.3	61	2979.2	2.0	5.6%
Residual	1806419.3	1220	1480.7		55.6%

* *G* = genotypic covariate, *E* = environmental covariate.

base for the decomposition of the interaction sum of squares generated by the factorial regression model. The regression on the mean model can be thought of as being nested within the first biadditive term, BT1 (see Methods).

The first factorial regression model included two genotypic covariates (EM and CS) and five bioclimatic covariates (S/C, F1S, F2S, RDS, TPS). The corresponding analysis of variance table is presented in Table 2. This factorial regression model led to a residual mean square comparable to the residual mean square for a biadditive model with two or three interaction terms, while the number of degrees of freedom for interaction corresponded to four interaction terms in the biadditive model. Hence, this factorial regression model seems to be very satisfactory, acknowledging that biadditive interaction terms represent the theoretically best possible covariates for describing interaction.

The genotypic and environmental main effects can be included in a factorial regression model, just like other covariates. The environmental main effect can be interpreted as a biological indicator of the environmental circumstances, as is usually done in the context of row regression. Adding the genotypic and environmental main effect in our factorial regression model led to the inclusion of only the genotypic main effect (G). Therefore, the second factorial regression model involved three genotypic covariates (EM, CS and G) and five environmental covariates (S/C, F1S, F2S, RDS and TPS), of which four were only defined for sandy soils. Interaction occurred mainly on this type of

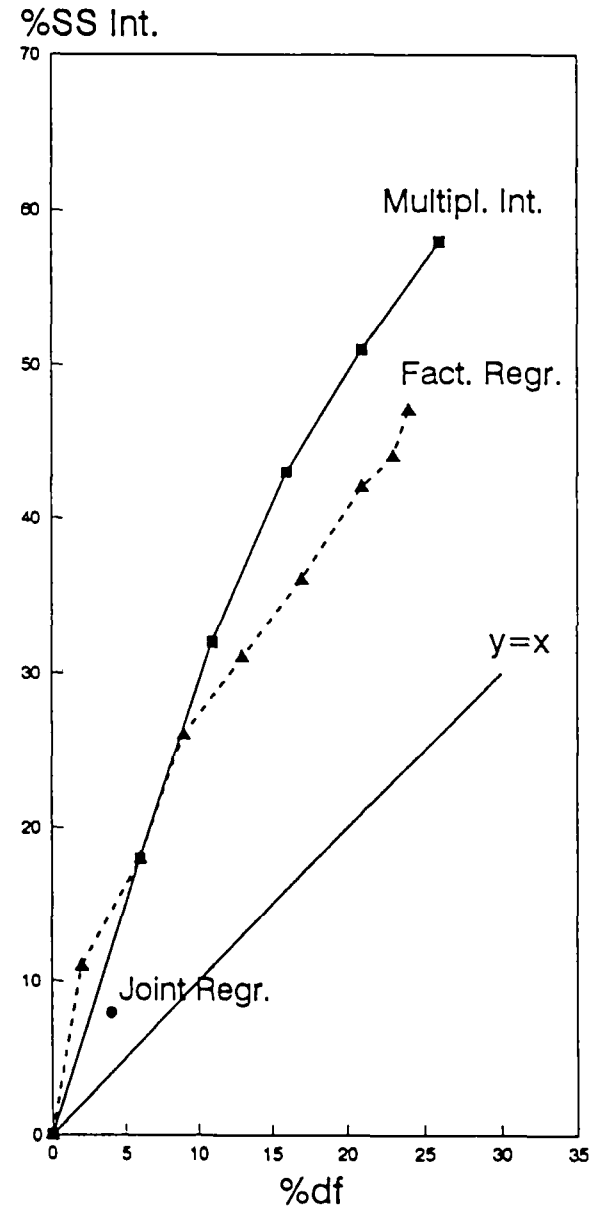

Fig. 1. Relationship between percentage of the interaction sum of squares explained and the percentage of the interaction degrees of freedom used by the joint regression model, the factorial regression model and the biadditive model.

soil. The decomposition of the interaction according to this second factorial regression model is illustrated in Table 3. The genotypic and environmental covariates are given in fitting order.

Every cell in Table 3 represents a model term for interaction. The effect of a specific genotypic (environmental) covariate in the interaction was split in a part due to interaction with specific environmental (genotypic) covariates and a part due to not further specified

Table 3. Decomposition of the interaction for second factorial regression model. In the cells; degrees of freedom ( ), mean square, and F-probability. I, J, K, H give number of varieties, number of environments, number of varietal covariates, and number of environmental covariates

Genotypes			Environments					
			J-1					
			H					(J-1)-H
			S/C	F1S	F2S	RDS	TPS	remainder
I-1	K	EM	(1)	(1)	(1)	(1)	(1)	(20)
			102.1	213.9	$50\ 10^3$	$6\ 10^3$	$53\ 10^3$	$12\ 10^3$
			NS	NS	***	*	***	***
		CS	(1)	(1)	(1)	(1)	(1)	(20)
			$5\ 10^3$	20	$4\ 10^3$	$2\ 10^3$	181.7	$4\ 10^3$
			NS	NS	NS	NS	NS	***
		G	(1)	(1)	(1)	(1)	(1)	(20)
			$2\ 10^3$	$12\ 10^3$	717.0	107.3	$4\ 10^3$	$5\ 10^3$
			NS	**	NS	NS	NS	***
	(I-1)-K	remainder	(60)	(60)	(60)	(60)	(60)	(1200)
			$4\ 10^3$	$4\ 10^3$	$3\ 10^3$	$3\ 10^3$	$3\ 10^3$	1425
			***	***	***	***	***	

* = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$, and NS is non-significant.

interaction with environments (genotypes). Using the size of the variance ratio, Table 3 points to three interesting products of covariates: EM.TPS, EM.F2S and G.F1S. In the discussion below we will present a biological explanation for these results.

The model explained 47.4% of the interaction sum of squares consuming 23.8% of the interaction degrees of freedom. Figure 1 shows the percentage interaction explained (% SSI) by the nested biadditive models and the nested factorial regression models in relation to the percentage of degrees of freedom (% df) used. Both families of models were satisfactory as their curves are situated far above the bisecting line. For the first three covariates, the factorial regression model possessed better or comparable explanatory power. The further introduced covariates, however, were not too far from the biadditive terms.

Discussion

Biological interpretation of the most important parameters, ρ_{kh}

As an aid to understanding the meaning of the interactions accounted for by the cross-products of genotypic and environmental covariates as revealed by Table 3, Table 4 gives a summary of the signs of the cross-products in relation to the signs of the centred variables constituting the products, plus the signs of the coefficient ϕ_{kh}. For early maturity we know that ratings are higher as genotypes mature earlier. So, after centring, early genotypes score positively whereas late genotypes score negatively. In the same way, high yielding genotypes score positively and low yielding genotypes score negatively. Concerning the environmental covariates; more frost days than average leads to a positive value after centring, and less than average leads to a negative value. With regard to temperature, higher than average temperatures give positive values, while lower than average temperatures give negative values. Combining the signs of the cross-products with those of the ϕ_{kh} parameters constitutes a first step towards a more biological interpretation of the interaction. Table 5 gives an overview of the biological interpretations.

Interpretation interaction EM.TPS

Early genotypes benefit from high average temperatures by stronger growth. Later genotypes in principle also benefit from higher average temperatures, but on sandy soils high temperatures can easily create drought stress later in the season, which can have negative effects on the production of the later genotypes. Low average temperatures cause slow growth, which is most influential in earlier genotypes. Later genotypes can

Table 4. Signs for the cross-products of centred covariates and the corresponding coefficients in the factorial regression model. EM stands for early maturity, G for genotypic main effect, TPS for mean temperature on sandy soils, F2S and F1S for number of frost days during the second and the first half of April on sandy soils

TPS

		High(+)	Low(-)	
EM	Early(+)	+	-	$\rho_{EM,TPS}$ -
	Late(-)	-	+	

F2S

		Many(+)	Few(-)	
EM	Early(+)	+	-	$\rho_{EM,F2S}$ -
	Late(-)	-	+	

F1S

		Many(+)	Few(-)	
G	High(+)	+	-	$\rho_{G,F1S}$ +
	Low(-)	-	+	

Table 5. Biological interpretations of some genotype by environment interactions. Abbreviations are as in Table 4

TPS

		High temp.	Low temp.
EM	Early	Good No problem	Bad No compensation
	Late	Bad < Marg. mean	Good Compensation

F2S

		Many days	Few days
EM	Early	Bad No compensation	Good No problem
	Late	Good Compensation	Bad <Marg. mean

F1S

		Many days	Few days
G	High	Relatively bad	Very good
	Low	Relatively good	Relatively bad

compensate through a longer growing season without drought stress.

Interpretation interactions EM.F2S

Many frost days in the second half of April is an indication of a cold period. This can delay come-up and lead to a shorter growing season. Especially the earlier genotypes will suffer from production loss. Later genotypes have a possibility to compensate later on in the growing season. When the second half of April is warmer, early genotypes perform relatively better than expected on basis of the marginal means, in contrast to the later genotypes, which do relatively worse.

Interpretation interaction G.F1S

A cold start of April delays planting. High yielding genotypes are relatively stronger affected by the delay than low yielding genotypes. When early planting is possible, because of good weather, high yielding genotypes take more advantage of these circumstances than low yielding genotypes.

General conclusions

Type of soil is a main determinant of genotype by environment interaction for yield in the Dutch Variety List trials for potato. Factorial regression models are useful statistical tools for finding biological interpretations of genotype by environment interaction.

Acknowledgements

Jan Bakker and Henk Bonthuis are thanked for comments and discussion, Paul Keizer for data and manuscript preparation.

References

Baril, C.P., 1992. Factor regression for interpreting genotype-environment interaction in bread wheat trials. Theor. Appl. Genet. 83: 1022–1026.

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

Crossa, J., H.G. Gauch & R.W. Zobel, 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. Crop Sci. 30: 493–500.

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

Denis, J.B., 1991. Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. Rev. Stat. Appl. XXXIX: 5–24.

Denis, J.B. & J.C. Gower, 1992. Biadditive models. Technical report. Laboratoire de Biométrie, INRA, Route de Saint-Cyr F78026, Versailles, France.

Denis, J.B. & J.C. Gower, 1994. Biadditive models. Letter to the editor. Biometrics 50: 310–311.

Decoux, G. & J.B. Denis, 1991. INTERA. Logiciels pour l'interpretation statistique de l'interaction entre deux facteurs. Laboratoire de Biométrie, INRA, Route de Saint-Cyr F78026, Versailles, France. 175 pp.

Finlay, K.W. & G.N. Wilkinson, 1963. The analysis of adaptation in a plant-breeding programme. Aust. J. Agric. Res. 14: 742–754.

Gabriel, K.R., 1978. Least squares approximation of matrices by additive and multiplicative models. J. R. Stat. Sc. B. 40: 186–196.

Gauch, H.G., 1988. Model selection and validation for yield trials with interaction. Biometrics 88: 705–715.

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

Kempton, R.A., 1984. The use of bi-plots in interpreting variety by environment interactions. J. Agric. Sci. C. 103: 123–135.

Kroonenberg, P.M., K.E. Basford & A.G.M. Ebskamp, 1995. Three-way cluster and component analysis of maize variety trials. Euphytica, this issue.

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

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

van Eeuwijk, F.A., L.C.P. Keizer & J.J. Bakker, 1995. Linear and bilinear models for the analysis of multi-environment trials: II. An application to data from the Dutch Maize Variety Trials. Euphytica, this issue.

Yates, F. & W.G. Cochran, 1938. The analysis of groups of experiments. J. Agric. Sci. C. 28: 556–580.