Theor Appl Genet (2001) 102:327–334

V. Foucteau · M. El Daouk · C. Baril

Interpretation of genotype by environment interaction in two sunflower experimental networks

Received: 28 December 2000 / Accepted: 29 May 2000

Abstract Genotype by environment interaction was investigated for sunflower grain yield in French official trials from two different networks. Genotype by environment interaction was described using cross-product terms from factorial regression. This modelling provided parsimonious descriptions of interaction: up to 20% of the sum of squares of interaction was explained using less than 7% of the degrees of freedom. It also made biological interpretation easier. Genotypic earliness at flowering by environmental earliness, and genotypic oil content by water deficit during the flowering period, appeared to be two repeatable interactions: in both data sets, early genotypes performed better in early environments while late genotypes gave much better yields in late environments, and the yield advantage of low-oil genotypes over high-oil genotypes was larger when water supply during flowering was non-limiting than in the case of water stress. Combined analysis of both data sets allowed us to compare patterns of interaction in both networks and to select a more robust model.

Keywords GE interaction · Factorial regression · Covariates · Sunflower · Series of experiments

Introduction

Sunflower varieties submitted for registration in France have a wide range of earliness. They are evaluated in four different experimental networks, according to their earliness. In each network, experimental sites aim at being representative of the area to which the earliness of varieties is adapted.

Most of the submitted varieties are tested in the two networks of intermediate earlinesses: early to mid-Early (E/ME) and mid-early to mid-late (ME/ML). ME/ML varieties are adapted to the southern half of France; most of the experimental sites are located in the Bassin Aquitain (Fig. 1). E/ME varieties can be cultivated under a larger range of latitudes, in particular in the centre of France and in the south of the Bassin Parisien, where most of the experimental sites are located. Two genotype by environment (GE) data sets are investigated in this study. Each one is extracted from the results of trials belonging to one of the two networks described above.

GE interaction (GEI) is investigated through factorial regression (Denis 1980, 1988; van Eeuwijk et al. 1996). This model allows the use of additional information about genotypes and environments in the form of covariates depending on one of the two factors. It has been used in numerous studies (Baril 1992, Rameau and Denis 1992; Charmet et al. 1993; van Eeuwijk and Elgersam 1993) which provided a good biological interpretation of GEI. The ability of factorial regression to explain interaction obviously depends on the relevancy and availability of the covariates; in this study special attention has been paid to the climatic characterisation of the environments.

Cross-product terms are the terms of a factorial regression model that involve products of a genotypic by an environmental covariate. These terms are of direct interpretation from a biological viewpoint. Their sign depends on the signs of the centred genotypic and environmental covariates and the regression parameter. Hence their interpretation can be summed-up in a 2×2 table of signs (Baril et al. 1995). The parsimony principle states that, for a predictive aim, it is much better to retain a model with too few parameters than with too many: a small bias is better than a large variance. From this point of view, cross-products terms are interesting because they only use one degree of freedom each. The other terms use many more as they involve regression coeffi-

Communicated by H.C. Becker

V. Foucteau (✉) · M. El Daouk
INRA, Unité de Biométrie,
F-78352 JOUY-EN-JOSAS Cedex, France
e-mail:Vincent.Foucteau@jouy.inra.fr

C. Baril
GEVES, La Minière,
F-78285 GUYANCOURT Cedex, France

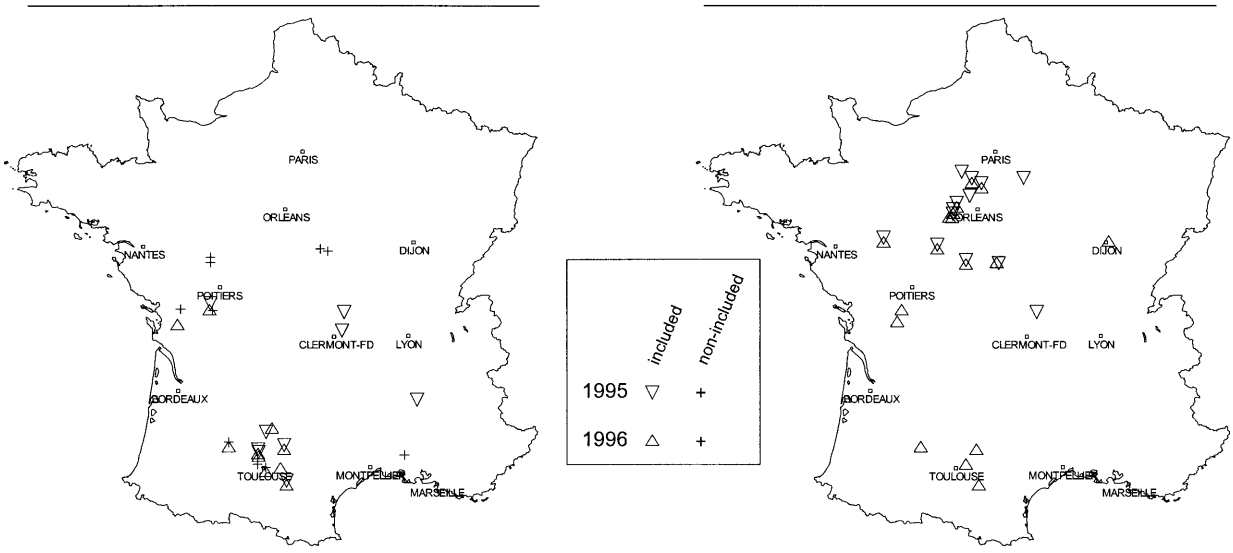

Fig. 1 Map of experimental sites for both networks

cients which depend on either genotypes or environments. They are second-order terms with respect to cross-products terms; so they can be neglected when looking for an approximation. Hence, for reasons of both convenience of interpretation and parsimony, only cross-product terms will be considered here.

The analysis of the two data sets allow us to compare interaction patterns in both networks. Combined covariate selection using both data sets may also lead to the choice of a more robust model.

Materials and methods

Experimental data

ME/ML and E/ME data sets respectively consist of a 13×19 and a 9×30 genotype by environment table. One genotype (a check variety) is common to both data sets. Each trial is considered as an environment. Both data sets involve 2 years (1995 and 1996). The E/ME data set includes all the experimental sites of the E/ME network while the ME/ML data set includes 19 of the 30 experimental sites of the ME/ML network (Fig. 1). The traits measured in each trial plot were: flowering date, number of plants per plot at harvest (NP), grain yield (GY) oil content (OC) and grain moisture content at harvest (MC). All the tables contain no missing values. From the three GE tables of NP, OC and MC, we derived three genotypic covariates (NP.G, OC.G and MC.G) and three environmental covariates (NP.E, OC.E and MC.E) by calculating the means by row and column respectively, as proposed by Baril (1992). The environments were also characterised by the number of days from sowing to the environmental mean-flowering date (ND.E).

The sunflower development cycle has been divided into five periods according to the physiology of the plant. Period 1 runs from sowing to emergence. The sowing date was available for all the trials whereas the emergence date was estimated by adding 90°C days to the sowing date. Period 2 extends from emergence to the B9 stage, which corresponds to the emergence of the 9th leaf, and was computed by adding 160° days to the emergence date. The flowering period (period 4) was delimited using the flowering dates of the earliest and latest genotype of each trial. The range in earliness at flowering was on average 10 days. In order to obtain a period of length comparable to the effective length of flowering for sunflower (approximately 20 days), we finally chose the earliest flowering date minus 5 days and the latest flowering date plus 5 days as boundaries for period 4. Period 3 goes from the B9 stage to the beginning of flowering as defined above. M3 is the maturity stage when grain moisture content decreases down to 15%; it matches with the end of physiological activity. The M3 stage was estimated from the harvest date using the empirical relationship:

$$M3=\text{Harvest Date}-(30-2MC.E),$$

where M3 and Harvest Date are expressed in number of days and MC.E is a percentage. Period 5 goes from the end of flowering to the M3 stage.

Climatic covariates were computed using meteorological data from the national meteorological network of Météo France. Each location was characterised by the nearest meteorological station measuring temperature, radiation, rainfall (R) or potential evapotranspiration (ETp). Average distances between locations and meteorological stations were respectively 6, 16, 4 and 14 km according to meteorological parameters. Water deficit (WD) in mm estimated as

$$WD=ETm-(R+I),$$

where I is the amount of water provided from irrigation and ETm is the maximum evapotranspiration, was computed for each of the five periods described above and called WD1 to WD5. ETm was calculated as $ETm=k\in Tp$ with k a coefficient equal to 0.6, 0.9, 1.25, 1.3 and 0.7 for periods 1 to 5 respectively (following Merrien 1992). The influence of low temperatures during flower differentiation (LT3) was quantified by summing daily minimum temperatures below 5°C over the 20 days following the B9 stage. In the same way, the effect of high temperatures during flowering and maturation (HT45) was obtained by summing daily maximum temperatures above 30°C over period 5. The sum of daily radiation from emergence to the M3 stage (RAD) was also computed. Mean temperature was calculated from sowing to flowering (MT). The sum of degree days (ST) based on 6°C (the sum of daily mean temperatures above 6°C) was computed over the same period for each genotype by environment combination; means by genotype and environment resulted in two additional covariates (ST.G and ST.E respectively). The latitude (LAT) and longitude (LONG) of

Table 1 Definition and summary statistics of the covariates

	Unit	Symbol	Network	Mean	Std	Min	Max
Quantitative genotypic covariates							
Oil content	%	OC.G	ME/ML	52.1	1.4	49.5	53.8
			E/ME	51.7	1.7	49.1	54.2
Moisture content at harvest	%	MC.G	ME/ML	9.1	0.7	7.7	10.3
			E/ME	10.6	1.1	9.5	12.7
Degree days based on 6°C from sowing to flowering	°C	ST.G	ME/ML	926	38	861	984
			E/ME	819	33	776	866
Sensitivity to *Phomopsis*	%	Ph	ME/ML	30	7	18	43
			E/ME	28	13	15	53
Sensitivity to *Sclerotinia*	%	Sc	ME/ML	31	9	18	46
			E/ME	31	10	13	46
Quantitative environmental covariates							
Oil content	%	OC.E	ME/ML	52.1	2.8	45.3	55.7
			E/ME	51.7	2.3	46.8	56.9
Moisture content at harvest	%	MC.E	ME/ML	9.1	2.2	6.1	14.5
			E/ME	10.6	3.2	5.4	19.4
Number of plants per plot		NP.E	ME/ML	6.4	0.5	5.7	7.4
			E/ME	6.6	0.7	4.9	8
Number of days from sowing to flowering		ND.E	ME/ML	83	9	74	108
			E/ME	85	8	73	105
Degree days based on 6°C from sowing to flowering	°C	ST.E	ME/ML	926	51	818	1006
			E/ME	819	66	714	941
Water deficit from sowing to emergence	mm	WD1	ME/ML	−11	21	−50	25
			E/ME	−3	16	−37	29
Water deficit from emergence to B9	mm	WD2	ME/ML	1	36	−73	48
			E/ME	8	22	−43	40
Water deficit from B9 to beginning of flowering	mm	WD3	ME/ML	−148	44	65	249
			E/ME	152	31	98	215
Water deficit during flowering	mm	WD4	ME/ML	72	34	12	134
			E/ME	78.3	30.7	0	138
Water deficit from end of flowering to M3	mm	WD5	ME/ML	37.6	60.3	−94	157
			E/ME	50	42	−48	130
RADiation days from sowing to M3	KJ/cm^2	RAD	ME/ML	247	24	210	298
			E/ME	236	19	199	280
LONGitude	Km	LONG	ME/ML	504	98	327	769
			E/ME	500	83	353	768
LATitude	Km	LAT	ME/ML	346	118	223	549
			E/ME	634	164	223	793
Mean temperature from sowing to flowering	°C	MT	ME/ML	17	1	15	19
			E/ME	16	1	13	18
Low temperatures from B9 to B9+20 days	°C	LT3	ME/ML	2	2	0	8
			E/ME	1	2	0	8
High temperatures from beginning of flowering to M3	°C	HT45	ME/ML	48	22	16	106
			E/ME	29	20	4	56
Qualitative environmental covariates		Symbol	Network		Levels		
Fungicide treatment		FT		Yes	No		
			ME/ML	13	6		
			E/ME	15	15		
Year		year		1995	1996		
			ME/ML	10	9		
			E/ME	17	13		

locations were employed as environmental covariates. The year of experimentation and the possible application of fungicide treatments were used through contrasts between 1996 and 1995, and between fungicide control and no fungicide control.

For some of the submitted varieties, sensitivity to two major pathogens of sunflower, *Sclerotinia* and *Phomopsis*, was assessed in artificially infested trials and is expressed as a percentage of infected plants per plot. Both data sets were selected so that information was available for all the genotypes. Table 1 indicates the summary statistics for all the covariates.

Statistical methods

Preliminary analyses

For each data set, a fixed anova model was fitted. All the trials were designed as a randomised complete block Δ; the mean intra-block error (weighted mean of the errors obtained from the analyses per trial, with weights equal to the number of degrees of freedom per trial, divided by the number of replicates which was the same for all trials) was used as an estimate of the error for all *F*-

tests. We also estimated variance components for genetic, environmental and interaction effects, considered as random, in order to compare the magnitude of the variation due to the various sources in both networks.

The factorial regression model

The factorial regression model (Denis 1980, 1988) allows multiple regression on covariates depending on either genotypes or environments:

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + \sum_{k=1}^{K}\sum_{h=1}^{H} X_{ik}\nu_{kh}Z_{jh} + \sum_{k=1}^{K} X_{ik}\tau_{jk} + \sum_{h=1}^{H} \rho_{ih}Z_{jh}, \quad \text{(model 1)}$$

$E(Y_{ij})$ the expectation of the random variable Y or the genotype i and the environment j; X_{ik} is the value of the k^{th} genotypic covariate (k=1, … , K) for genotype i; and Z_{jh} is the value of the h^{th} environmental covariate (h=1, … , H) for environment j. Covariates are centred.

Only the cross-product terms will be considered here. These terms consist of regressions on the products of a genotypic by an environmental covariate. Regressions on covariates will be extended to the main effects. So the model can be finally written as:

$$E(Y_{ij}) = \mu + \left(\sum_{k=1}^{K} X_{ik}\nu_{k0} + \alpha_i^*\right) + \left(\sum_{h=1}^{H} \nu_{0h}Z_{jh} + \beta_j^*\right) + \sum_{k=1}^{K}\sum_{h=1}^{H} X_{ik}\nu_{kh}Z_{jh}. \quad \text{(model 2)}$$

In the text, the parameter ν_{kh} will be named by replacing the subscripts k and h by the name of the k^{th} genotypic covariate and the h^{th} environmental covariate, for $k \geq 1$ and $h \geq 1$. With K genotypic covariates and H environmental covariates, the interaction is described using KH degrees of freedom. The efficiency of a model will be defined as the ratio of the percentage of the sum of squares of interaction explained by the model, by the percentage of degrees of freedom of interaction employed.

Analyses of variance for factorial regressions are described in two-way tables (see Tables 3–6), following the model proposed by van Eeuwijk et al. (1996). Genotypic and environmental covariates are given respectively in rows and columns. Each cell corresponds to a term of the model. Both types of covariates are ordered; each term is adjusted for the previous ones. The first row and the first column (labelled constant) correspond respectively to the genotypic and the environmental main effects; they are partially explained by the covariates. The cells belonging neither to the first row nor to the first column correspond to the interaction.

Covariate selection

Considering the number of genotypes and environments in both data sets, we used K≤3 and H≤5. Covariate selection was performed by progressively adding to the additive model the best pairs of genotypic and environmental covariates until three covariates of each factor had been introduced or the best pair was non-significant. Then, additional environmental covariates could be eventually introduced, if significant and with H≤5.

This procedure was applied to each of the two data sets, leading to the selection of two subsets of covariates called ME/ML and E/ME. To obtain a more robust model, we performed a combined covariate selection using both data sets, providing a common subset of covariates named ME/ML+E/ME. As the sum of squares and degrees of freedom of interaction were very close from one data set to the other, the sums of squares accounted for by a pair of covariates (or an environmental covariate) in each of the two data sets were just added to quantify their relevance on both data sets. Each of the models including one of the three subsets of covariates was fitted to both data sets.

Table 2 Anova table for both data sets

Source of variation	SS	*df*	MS	*F*	Variance components
ME/ML					
Genotype	125	12	10.4	8.6	0.4
Environment	6188	18	344.8	284.9	26.2
Interaction	642	216	3.0	2.5	3.0
Error[a]		1529	1.2		
E/ME					
Genotype	143	8	17.9	15.9	0.5
Environment	9068	29	312.7	280.0	34.4
Interaction	624	232	2.7	2.4	2.7
Error[a]		2222	1.1		

[a] Mean intra-block error

Results

All the effects (genotype, environment and interaction) were significant at the 0.001 probability level (Table 2). Comparing variance components and error, it appears that the magnitude of GEI was slightly less than 3-times the error in both networks. Interaction was much more important than the genotypic effect and led to multiple genotypic rank changes.

Comparison of the three models

Covariate selection for each data set led to two models, namely ME/ML and E/ME. Model ME/ML fitted to data set ME/ML as well as model E/ME fitted to data set E/ME and provided a very satisfactorily explanation of GEI, accounting for 20% of the sum of squares of interaction (SSI) using less than 7% of the degrees of freedom (see Table 7). Model ME/ML fitted to data set E/ME, and model E/ME fitted to data set ME/ML, performed poorly even if their efficiency remained above 1. Model ME/ML+E/ME, resulting from the combined covariate selection using both data sets, performed nearly as well as model E/ME when fitted to data set E/ME, but distinctly less than model ME/ML when fitted to data set ME/ML.

The ME/ML subset of covariates included Sc, WC.G and Ph as genotypic covariates and WD1, RAD, WC.E, WD2 and FT as environmental covariates. The E/ME subset of covariates included ST.G, Ph and OC.G as genotypic covariates and ST.E, OC.E, WD4, WD3 and WC.E as environmental covariates. Finally, the ME/ML+E/ME subset included ST.G, OC.G and WC.G as genotypic covariates and ST.E, WD4, WD3, RAD, LT3 as environmental covariates. The ME/ML and E/ME subsets of covariates differed strongly: sensitivity to *Phomopsis* (Ph) and moisture content at harvest (MC.E) were the only common covariates. The ME/ML+E/ME subset of covariates mainly included covariates from the E/ME subset.

Table 3 Two-way anova table for model ML/ME fitted to the data set ML/ME. In each cell: mean squares, degrees of freedom, F-probability (***: $P<0.001$, ** $P<0.01$, * $P<0.05$ and ns: non-significant) and sign of estimated regression coefficients on cross-products (when significant)

	Constant	WD1	RAD	MC.E	WD2	FT	Remainder
Constant		429	1	345	954	865	277
		1	1	1	1	1	13
		***	ns	***	***	***	***
		–		+	–	+	
Sc	29	16	19	15	10	4	
	1	1	1	1	1	1	
	***	***	***	***	**	ns	
	–	+	–	–	+		
MC.G	39	1	15	5	3	1	
	1	1	1	1	1	1	
	***	ns	***	*	ns	ns	
	+		–	+			
Ph	1	10	0	0	11	14	
	1	1	1	1	1	1	
	ns	**	ns	ns	**	**	
		–			+	+	
Remainder	6						3
	9						201
	***						***

Table 4 Two-way anova table for model E/ME fitted to the data set E/ME. In each cell data as a Table 3

	Constant	ST.E	OC.E	WD4	WD3	MC.E	Remainder
Constant		15	870	25	0	1315	285
		1	1	1	1	1	24
		***	***	***	ns	***	***
		–	+	–		+	
ST.G	7	39	10	1	11	2	
	1	1	1	1	1	1	
	*	***	**	ns	**	ns	
	+	–	–		+		
Ph	34	0	18	4	1	9	
	1	1	1	1	1	1	
	***	ns	***	ns	ns	*	
	–		+			–	
OC.G	22	0	0	21	7	6	
	1	1	1	1	1	1	
	***	ns	ns	***	*	*	
	–			+	–	–	
Remainder	16						2
	5						217
	***						***

Biological interpretation of cross-products

Many cross-product terms of the ME/ML and E/ME models were significant, respectively nine and eight over a total of 15 (Tables 3 and 4). Two cross-product terms from model ME/ML+E/ME are interpreted here. They appear as repeatable GE interactions since they were significant and followed the same pattern in both data sets. Their importance was greater in the E/ME than in the ME/ML network because of a larger range of environmental conditions in the former network.

ST.G by ST.E interaction

The cross-product term involving covariates ST.G and ST.E was the best one. It accounted for 6% of the SSI in the E/ME data set and for 1%, still being significant, in the ME/ML data set. ST.G is a commonly used measure of genotypic earliness at flowering. ST.E stands for some environmental earliness that must be clarified. In both data sets, ST.E was not significantly correlated with ND.E. In the ME/ML data set, ST.E was significantly correlated with MT (0.71), whereas in E/ME it was significantly correlated with LAT (-0.62) and RAD (0.46). In both cases, we will refer to environments with ST.E value below and above average as early and late respec-

Table 5 Two-way anova table for model ML/ME+E/ME fitted to the data set ML/ME. In each cell data as in Table 3

	Constant	ST.E	WD4	WD3	RAD	LT3	Remainder
Constant		493 1 *** –	379 1 *** +	92 1 *** +	1 1 ns	1368 1 *** –	297 13 ***
ST.G	71 1 *** +	8 1 * –	8 1 ** +	12 1 ** –	7 1 * +	12 1 ** –	
OC.G	4 1 ns	0 1 ns	7 1 * +	3 1 ns	3 1 ns	0 1 ns	
MC.G	14 1 ** +	2 1 ns	10 1 ** +	1 1 ns	11 1 ** –	4 1 ns	
Remainder	4 9 **						3 201 ***

Table 6 Two-way anova table for model ML/ME+E/ME fitted to the data set E/ME. In each cell data as in Table 3

	Constant	ST.E	WD4	WD3	RAD	LT3	Remainder
Constant		15 1 *** –	13 1 ** +	13 1 *** –	46 1 *** +	156 1 *** –	295 24 ***
ST.G	7 1 * –	39 1 *** –	4 1 * –	13 1 ** +	3 1 ns	0 1 ns	
OC.G	28 1 *** –	0 1 ns	22 1 *** +	5 1 * –	6 1 * –	10 1 ** +	
MC.G	16 1 *** +	1 1 ns	3 1 ns	4 1 * –	8 1 ** +	1 1 ns	
Remainder	14 5 ***						2 217 ***

Table 7 Comparisons of the three models. In each cell: the percentage of interaction sum of squares explained by model (a) in normal types, the percentage of interaction degrees of freedom used by model (b) in italics, and efficiency (a/b) in small type

Model	Data set		
	ME/ML	E/ME	ME/ML+E/ME
ME/ML	19.3 *6.9* **2.8**	8.6 *6.5* **1.3**	14.1 *6.7* **2.1**
E/ME	10.0 *6.9* **1.4**	20.8 *6.5* **3.2**	15.3 *6.7* **2.3**
ME/ML+E/ME	13.5 *6.9* **2.0**	19.4 *6.5* **3.0**	16.4 *6.7* **2.4**

tively. Hence, in ME/ML network early environments were warm, from sowing to flowering, while in the E/ME they were located south and received much radiation. Symmetrically, in the ME/ML data set late environments were cool while in E/ME they were located north and received little radiation. In both networks early genotypes performed better in early than in late environments, while late genotypes performed better in late than in early environments. Such an interaction between genotype and environment earliness is frequent and is often referred to as a phenology related interaction. Moro et al. (1994) found a days to flowering by latitude interaction and a days to flowering by mean temperature over the growing season interaction in sunflower; these accounted respectively for 13% and 4% of SSI and showed the same pattern as in the present study. For sunflower, the timing of flowering in relation to water availability and high-temperature occurrence is crucial: late genotypes in early environments probably flowered during the drought season, penalising the setting of the number of kernels and grain filling.

The partitioning of genotypes into earliness groups was aimed at limiting these interactions. Despite this stratification, phenology related interaction was the major source of interaction in the E/ME network because of the large range of latitudes of the experimental sites.

OC.G by WD4 interaction

In both networks, water stress during flowering appeared to be more penalising for low- than for high-oil genotypes. The genetic correlation between oil content and grain yield is negative because lipid synthesis has a higher energetic cost than the synthesis of the other grain components (mainly protein). This correlation was not expressed in the sample of genotypes of the ME/ML data set ($\hat{v}_{OC.G,0}$ not significantly different from 0) but was strong in the case of E/ME. Fertilisation takes place during the flowering period. Hence, a water stress during period 4 (high positive value of the WD4 covariate) leads to a reduction of kernel number. When water supply was non-limiting, kernel number was maximum. High-oil genotypes could not fill all of them because of the higher cost of lipidogenesis, so their grain weight and finally their yield were lower than those of low-oil genotypes. In the case of water stress, the number of kernels to be filled was reduced, so high-oil genotypes were able to fill them more-satisfactorily. Hence, when water supply was limiting during flowering, the yield advantage of low-oil genotypes was smaller or even zero.

Discussion

Comparison of both networks

Disease-related covariates were all selected in the ML/ME model (Sc, Ph and FT) whereas only Ph was retained in the E/ME model. This result agrees with the positive relation between the earliness of genotypes and resistance to diseases: the ML/ME varieties are more likely to be susceptible than the E/ME varieties. Moreover, Bassin Aquitain, containing most of the ML/ME sites (see Fig. 1), has a higher *Phomopsis* risk than the south of the Bassin Parisien (CETIOM 1997), where most of the E/ME experimental sites were located. Hence, disease pressure must have been higher in the ME/ML network than in the E/ME network though two-thirds of the environments were treated in the ML/ME data set and only a half in E/ME. However, the non-selection of Sc in the E/ME model can be due to a stronger correlation between Ph and Sc in the E/ME (0.51) than in the ML/ME data set (0.16), even if none was significant.

The covariates of subset E/ME are more related to the end of the development cycle (MC.G for genotypic earliness, OC.E and WD4 for environmental covariates), whereas the covariates of subset ML/ME are related to early events (ST.G for genotypic earliness, WD1 and WD2 for environmental covariates). The end of the development cycle is more important for late genotypes, while the beginning of the development cycle is more important for early ones. Hence, events occurring during these periods are also important for GEI.

In the ML/ME network late genotypes at flowering gave better yields than early genotypes ($\hat{v}_{ST.G,0}$) because of a longer development cycle. This advantage of late genotypes was not expressed in the E/ME network ($\hat{v}_{ST.G,0}$ even when significantly positive at the 0.05 probability level). This can be due to the small number of genotypes involved in the E/ME data set. Alternatively, it could be an intrinsic difference between both networks: in E/ME the experimental sites were located too far north, or the tested genotypes were too early, compared to ML/ME network. Therefore, the allocation of the submitted genotypes to one or the other network could have changed their evaluation.

Use of cross-product terms

Cross-product terms provide a simple way to deal with interaction. But more-elaborate models are often needed, such as factorial regression models including, besides regressions on cross-products, regressions on genotypic covariates for each environment and regressions on environmental covariates for each genotype (model 1). In Baril et al. (1995), three genotypic covariates and five environmental covariates were selected based on model 1 to explain genotype by environment interaction for yield in potato crops. Cross-product terms accounted for only 4% of the SSI but the whole model explained 47%. Nevertheless some examples of single meaningful cross-product can be found in the literature: the days to flowering by latitude interaction cited before (Moro et al. 1994), or an isozyme allelic frequency by minimum temperature interaction accounting for 28% of SSI in ryegrass spring vigour (Balfourier et al. 1997).

Combined covariate selection procedure

One genotype is common to both data sets. But fitting a factorial regression to a unique data set obtained for ME/ML and E/ME data sets was not possible since a necessary condition for the model estimability is the presence of at least two common genotypes.

Combined covariate selection was performed by adding the sums of squares of interaction explained by the model in each data set. This simple procedure was sensible because both the sum of squares and the degrees of freedom of interaction were very similar from one data set to the other. If this condition is not fulfilled, several solutions can be thought of. To divide the sum of squares of interaction accounted for by the model by the total sum of squares of interaction in each data set, amounts to giving them the same weight. One can suggest to weigh the contributions of each data set by the degrees of freedom of interaction or by the error. We prefer the former solution because the degrees of freedom are known whereas the error has to be estimated.

Conclusion

Covariate selection for each data set led to the choice of two very different subsets of covariates. Some differences between both subsets were predictable from the known specificity of each network (stronger genotypic earliness by environmental earliness interaction in the E/ME than in the ME/ML network, or importance of the covariates related to the beginning of development cycle for the E/ME network while related to the end of development cycle for ME/ML).

The use of only cross-product terms, which are much more parsimonious than the full factorial regression model, and the combined analysis of two data sets led to select more robust model. But the use of only cross-product terms requires the relevant covariates for both factors.

Acknowledgements The authors thank J.-B. Denis for his helpful comments for improving the manuscript. This study was supported by the GEVES and the département de Biométrie et Intelligence Artificielle de l'INRA. The authors declare that the experiments conducted for this publication comply with the current laws of France.

References

Baril CP (1992) Factor regression for interpreting genotype by environment interaction in bread-wheat trials. Theor Appl Genet 83:1022–1026

Baril CP, Denis J-B, Wustman R, van Eeuwijk FA (1995) Analysing genotype by environment interaction in Dutch potato variety trials using factorial regression. Euphytica 82:149–155

Balfourier F, Oliveira JA, Charmet G, Arbones E (1997) Factorial regression analysis of genotype by environment interaction in ryegrass populations, using both isozymes and climatic data as covariates. Euphytica 98:37–46

CETIOM (1997) Le tournesol en 1997. CETIOM

Charmet G, Balfourier F, Ravel C, Denis J-B (1993) Genotype by environment interactions in a core collection of French perennial ryegrass populations. Theor Appl Genet 86:731–736

Denis J-B (1980) Analyse de régression factorielle. Biométrie-Praximétrie 20:1–34

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

Denis J-B (1998) BIAREG: Splus functions to perform biadditive regressions. Report INRA, Versailles

Eeuwijk FA van, Denis J-B, Kang MS (1996) Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables. In: Kang MS, Gauch HG (eds) Genotype-by-environment interaction. CRC-Press, Boca Raton, Florida, USA, pp 15–49

Eeuwijk FA van, Elgersam A (1993) Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. Heredity 70:447–457

Merrien A (1992) Physiologie du tournesol. Les points techniques du CETIOM. CETIOM

Moro J, Abad A, Rodriguez A, Denis J-B (1994) Interaction variedad×localidad en ensayos de valor agronomico de girasol. Investigacion agraria : produccion y control vegetales 9:35–52

Rameau C, Denis J-B (1992) Characterization of environments in long-term multi-site trials in *Asparagus*, through yield of standard varieties and use of environmental covariates. Plant Breed 109:183–191