Theor Appl Genet (2001) 103:371–382

M. Brancourt-Hulmel · C. Lecomte · J.-B. Denis

Choosing probe genotypes for the analysis of genotype-environment interaction in winter wheat trials

Received: 19 June 2000 / Accepted: 15 November 2000

Abstract Genotype-environment interaction was analyzed in French multi-environment wheat *(Triticum aestivum* L.*)* trials using probe genotypes and bi-additive factorial regression. Probe genotypes are specific genotypes in which the comparisons of yield components to reference values describe the most-important environmental factors that limited grain yield. The time-period until flowering was described by the deviation of kernel number from a threshold number while the grain-filling period was described by the reduction of thousand-kernel weight from a potential value. The aim of this paper was to determine the convenient number and the characteristics of probe genotypes to include in wheat breeding trials.

Two sets of genotypes were used to model genotype-environment interaction: set 1 with 12 varieties tested in 18 environments and set 2 with ten lines tested in 14 environments. Set 2 was used for validation. Seven probe genotypes described the environments by providing environmental covariates, namely differences in yield components, for further analysis of interaction in set 1 and set 2. Interaction was modelled with bi-additive factorial regressions including differences in yield components. Several rounds of models were fitted to determine the optimal number of probe genotypes (i.e. environmental covariates) to introduce. From the seven probe genotypes, all the possible combinations including one to seven genotypes were studied. Significance of the combinations was tested with critical values obtained from simulations through 1,000 random permutations. Taking into account the information available on the probe genotypes, one would think that two, three or four probe genotypes would be sufficient, otherwise the number should reach four or five genotypes. In all cases, these numbers will provide models more-parsimonious than the classical AMMI model. The important information to be known on the probe genotypes prior their first multi-location experiment is: interaction pattern, earliness, and differences in yield component. Tested for the first time, a quadruplet is better than a triplet because the probability of choosing complementary genotypes increases with their number.

Keywords Genotype-environment interaction · Bi-additive factorial regression · Probe genotypes · Random permutation · Winter wheat

Communicated by H.C. Becker

M. Brancourt-Hulmel (✉)
INRA, Unité de Génétique et d'Amélioration des Plantes,
80200 Estrées-Mons, France

C. Lecomte
INRA, Station de Génétique et d'Amélioration des Plantes,
17 rue Sully, BP 86510, 21065 Dijon Cedex, France

J.-B. Denis
INRA, Unité de Biométrie et Intelligence Artificielle,
INRA, domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

Introduction

Genotype-environment interaction is a common phenomenon in plant breeding. Differences between genotypes can increase or decrease from one environment to another, which is usually called quantitative interaction, and genotypes can even rank differently between environments in the case of qualitative interaction. Before explaining genotype-environment interaction, it is necessary to quantify it. Many efficient methods have been proposed and used for quantifying interaction. Explaining interaction is more difficult as it requires integrated approaches, which combine biometrics, agronomy, modeling, genetics and plant breeding. The analysis of genotype-environment interaction can be based on the description of the environments in terms of the most-important limiting factors responsible for genotype-environment interaction and on the assessment of the sensitivity of genotypes with respect to these limiting factors. Desclaux (1996) proposed this approach for soybean (*Glycine max* L.) and Brancourt-Hulmel et al. (1999) for wheat. As yield is too complex a trait to model directly, the analysis of its components can be more helpful. The

product between the two main components can model grain yield: the kernel number per square meter (KN) and the thousand-kernel weight (TKW). Individual environments can be characterized by the difference of the actual value of yield components like KN and TKW from an optimal value. Such optimal values can be estimated only when genotypes are experimented in natural conditions free from stress. This requires specific trials and can be made only on a small set of genotypes, called ''probe genotypes''. These genotypes are used to ''probe'', i.e. to capture, the influence of environmental constraints. Probe genotypes would produce yield components close to the optimal values in favorable environments and small values in unfavorable ones. Brancourt-Hulmel et al. (1999) have proposed such an approach for wheat. Differences or deviations observed for KN and TKW can be then introduced as environmental covariates in linear and bi-linear factorial regression models (Denis 1988, 1991; van Eeuwijk 1995) for explaining genotype-environment interaction for yield (Brancourt-Hulmel 1999; Brancourt-Hulmel et al. 2000). These variates, which split grain yield into more simple traits, point out shorter periods involved in genotype × environment interaction and may discard or reveal environmental variates with opposite effects between several periods of the formation of yield (early water deficits could be hidden by late water deficits for instance). They are also interesting for establishing a variety type based on the behavior of the probe genotypes. As the probe genotypes constitute the key to such an approach, they have to be carefully chosen and their number must be sufficient. If there are too few, the environments could be poorly characterized, while for too many their use could be too costly for plant breeding trials and the statistical models will be less parsimonious. In previous work, a procedure of selection of these environmental covariates in bi-additive factorial regressions was given (Brancourt-Hulmel et al. 2000). The environmental covariates can be reduced in number by discarding those poorly correlated to the synthetic variates provided by the bi-additive factorial regression. This is a way to reduce the number of probe genotypes after the experiment and to provide more-parsimonious models. It would be also interesting to know how to reduce their number prior to the experiment in order to limit the experimental costs. In recent work, the number of probe genotypes was chosen arbitrarily (Brancourt-Hulmel et al. 1999; Desclaux 1996) and information for choosing them in a good manner is lacking. The aim of this paper was to define an optimal number, as well as the main characteristics, of probe genotypes to include in multi-environment trials of winter wheat for explaining genotype-environment interaction. The definition of the optimal number was based on statistical aspects (part of the interaction explained and parsimony) and experimental considerations (experimental costs and the characteristics of probe genotypes).

Materials and methods

Description of the data

Seven probe genotypes were considered: Apollo (APO), Arminda (ARM), Camp-Rémy (CAR), Soissons (SOI), Récital (REC), Talent (TAL) and Thésée (THE). Two deviations of yield components were determined by comparisons to reference values for these genotypes in 18 environments: DKN, the difference in percentage of kernel number to the kernel threshold (or kernel reference), is defined by $100 \times (KNthreshold\text{-}KN)/(KNthreshold)$; and RTKW, the reduction in percentage of thousand-kernel weight from the potential thousand-kernel weight (or reference thousand-kernel weight), is defined by $\max[0;100(potentialTKW\text{-}TKW)/potentialTKW]$. DKN describes the time-period until flowering, while RTKW corresponds to the grain-filling period. Small deviations indicate that conditions were favorable for the formation of yield while high values correspond to unfavorable conditions. The corresponding reference values for TKW and KN are given in Table 1 and were determined from a long-term experiment carried out since 1987 which gathered about 500 yield components per genotype (Brancourt-Hulmel et al., 1999). Grain yield, DKN and RTKW are given in Tables 2 and 3. The seven probe genotypes differed in their pattern of yield formation since the maximal thousand-kernel weight varied from 41.5 g (Récital) to 52.0 g (Thésée). As the maximal yield observed for these genotypes was similar, Thésée showed a lower threshold for the kernel number (21,278) than Récital (25,616). They also differed for earliness at heading: Récital flowered 13 days on average before Arminda (Table 1). Regarding earliness at maturity (Table 1), the genotypes are classified into four groups: early varieties (Récital and Talent), 1/2 early (Soissons, Thésée), 1/2 late (Camp-Rémy), and late (Apollo and Arminda).

Two sets of genotypes were studied for modelling genotype-environment interaction. The first set, namely set 1, was constituted of 12 varieties (Apollo, Artaban, Baroudeur, Camp-Rémy, Génial, Récital, Renan, Rossini, Soissons, Talent, Thésée and Viking), which were tested in France for agronomic traits (grain yield, kernel number and thousand-kernel weight). The varieties

Table 1 Group of earliness at maturity and potential values of the seven probe genotypes. TKW = thousand-kernel weight, KN = kernel number, and Std = standard deviation

Genotype	Heading date (days from 1st of January)	Group of earliness at maturity (ITCF, 1998)	Maximal grain yield t/ha 0% moist. cont.	Maximal TKW +/– Std g 0% moist. cont.	KN threshold +/– Std /m²
Thésée	146	1/2 early	11.1	52.0 +/– 1.1	21,278 +/– 471
Apollo	152	Late	11.5	45.4 +/– 1.7	25,174 +/– 950
Talent	143	Early	10.2	45.1 +/– 3.5	22,781 +/– 1,864
Soissons	144	1/2 early	11.5	43.4 +/– 1.7	26,464 +/– 1,025
Arminda	152	Late	10.5	43.0 +/– 1.5	24,737 +/– 855
Camp-Rémy	151	1/2 late	10.2	41.8 +/– 1.4	24,414 +/– 811
Récital	141	Early	10.6	41.5 +/– 1.6	25,616 +/– 971

differed for heading date: Apollo was the latest and flowered 11 days after Récital. A second set, called set 2, was constituted of ten lines bred by several units of INRA (DI003, RE001, RE006, RE009, RE813, RE914, VM002, VM003, VM014, VM017). Regarding heading date, differences between lines were smaller than for the varieties of the previous set because there were no early lines (such as Récital for instance).

Set 1 was used for determining the optimal number of probe genotypes as well as their main characteristics. Set 2 was independent from the previous one and was used for validation.

The 18 environments were combinations of the year (1991 or 1992), the site (Mons (49°56′ N Lat., 2°56′ E Long.), La Minière (48°48′ N, 2°08′ E), data available only in 1991, Rennes (48°05′ N, 1°41′ W), Dijon (47°19′ N, 5°01′ E) and Ondes (43°36′ N, 1°26′ E) and the treatment (medium sowing date with fungicides, medium sowing date without fungicides and late sowing date). Set 1 was tested in all these environments, while set 2 was tested in 14 environments, the site Ondes was lacking for the four treatments. In the 18 environments, probe genotypes differed for grain yield, DKN and RTKW (see Tables 2 and 3).

More details, about the experiments dealing with characteristics of the 12 varieties and probe genotypes or the environments, plant sampling and measurements, are reported by Brancourt-Hulmel (1999) and Brancourt-Hulmel et al. (1999). Brancourt-Hulmel and Lecomte (1994) also gave further information about the ten lines.

Statistical developments, covariates and inference test

Preliminary analyses were carried out to describe the seven probe genotypes. Genotype-environment interaction was analyzed according to a classical multiplicative (AMMI) model (Gollob 1968; Mandel 1971; Gauch 1992) with three significant multiplicative terms. Three traits were considered: grain yield (t/ha), deviation of kernel number (percentage of the threshold value) and reduction of thousand-kernel weight (percentage of the threshold value). For each genotype, interaction was described in terms of ecovalence (Wricke 1962) and the interaction pattern with genotype scores provided by the AMMI model. Only scores of the first two terms were considered as they incorporated most of the interaction.

For the two sets of genotypes (set 1 and set 2), genotype-environment interaction for grain yield was analyzed using AMMI model and bi-additive factorial regression (Denis 1988, 1991). This model generalizes both factorial regression (Denis 1988) and AMMI. Just as for AMMI, axes or synthetical environmental variates are built, but under the restrictions of being linear combinations of environmental variates. This model was applied here introducing only environmental covariates, in the same manner as Wood (1976).

It is written here with three multiplicative terms (Denis 1991):

$$E[Y_{ge}] = \mu + \alpha_g + \beta_e + \lambda_1\gamma_{g1}\left(\sum_{h=1}^{H_B}\delta_{h1}E_{eh}\right) + \lambda_2\gamma_{g2}\left(\sum_{h=1}^{H_B}\delta_{h2}E_{eh}\right) + \lambda_3\gamma_{g3}\left(\sum_{h=1}^{H_B}\delta_{h3}E_{eh}\right),$$

where $E[Y_{ge}]$ is the expectation of performance Y_{ge} for Genotype g grown in Environment e; μ is the general mean; α_g is the Genotype main effect; β_e is the Environment main effect; each of the multiplicative terms has the same structure: λ_1 is the size, γ_{g1} is the normalized genotype vector of the genotype sensitivities, $\sum_{h=1}^{H_B}\delta_{h1}E_{eh}$ is a normalized linear combination of the H^B environmental covariates E_{eh}, assigned to the first term. The next multiplicative terms (λ_2, λ_3, γ_{g2}...) follow the same definitions. g varies from 1 to G and e from 1 to E.

Classical bi-additive models were fitted by the ''INTERA package'' (Decoux and Denis 1991) and bi-additive factorial regressions were performed with BiaReg, a set of SPlus-functions developed by Denis (1998).

Deviations of yield components were used as environmental covariates for characterizing the environments. The environmental covariates were first centered and scaled to unit variance. Several rounds of bi-additive factorial regressions were carried out to determine the optimal number of probe genotypes to introduce. From the seven probe genotypes, all possible subsets (singles, pairs, triplets, quadruplets, quintuplets, sextuplets, and septuplets) were tested as environmental covariates in bi-additive factorial regressions.

From an agronomic viewpoint, it is useful to analyze the two periods of the formation of yield: before flowering with DKN and after with RTKW. For each probe genotype, the two covariates (DKN and RTKW) were then introduced to give a full description of the formation of grain yield. The second reason of this choice is to limit experimental costs: it is better to obtain the description of the whole plant cycle from DKN and RTKW measured only on one probe genotype rather than from DKN given by one genotype and RTKW by another.

A pair of probe genotypes then corresponded to four environmental covariates: DKN and RTKW determined for each of the two probe genotypes. A triplet involved six environmental covariates and so on. Results of the different combinations are given in terms of efficiency, i.e. in the percentage of the sum of squares of the interaction explained by each model (%SSI).

The optimal number of probe genotypes was determined by an empirical approach considering statistical aspects (%SSI compared to critical values and parsimony) and experimental considerations (experimental cost and genotype characteristics). The %SSI from set 1 was compared to critical levels determined by a permutation approach. They were computed under the null hypothesis (covariates unrelated to interaction) to take into account the fact that we are considering the best combination of probe genotypes among all those possible for a given size. These critical levels were determined by simulation of 1,000 random permutations in the spirit of the permutation tests proposed by Fisher (1935) as referred by Scheffé (1959). Every random permutation was obtained as follows: within each probe genotype, DKN and RTKW values were simultaneously and randomly permuted over the environments, the permutations being independent from one environment to another. This permutation procedure enables one to detect complementary probe genotypes by breaking correlations among them, the effect of these correlations being included in the underlying test. It is important to detect if probe genotypes are not only exchangeable at the single level, indicating that {1} can be equivalent to {2} for instance, but also at the combination level, indicating, for example, that {1,4} can be equivalent to {3,6,7}. For each random permutation, all possible combinations (singles to septuplet) were performed and the maximal %SSI by size stored. For each size, the 5% critical values of %SSI were deduced from the obtained empirical distributions. Permutations were limited to 1,000 as only slight differences were observed between 500 and 1,000 permutations for the assessment of these critical values. A total of 127,000 permutations were performed.

The comparisons of %SSI to critical levels, determined by means of the randomization tests, give ideas about the consequence of using the best subset of probe genotypes and control in some way the problem of multiplicity. Besides these statistical precautions, the multiplicity problem was also addressed with the use of a second data set to validate the best subsets found on the first set.

Results

Description of seven probe genotypes

Grain yield and deviations of yield component, DKN and RTKW, were all significant for genotype, environment and genotype-environment interaction effects. For all traits, the highest effect was observed for the environment (Tables 2 and 3). The seven probe genotypes expressed diversity for grain yield (Table 2): the highest

Table 2 Grain yield (t/ha) for the seven probe genotypes in each environment. Std = standard deviation, snk = Student-Newman-Keuls multiple range test

Environments	SOI	THE	APO	REC	ARM	TAL	CAR	Mean	Std	snk grouping at 0.05 probability level
91MININ	9.5	8.9	10.0	9.4	8.6	8.5	8.2	9.0	0.6	*
91RENIN	9.5	7.8	5.9	8.0	7.4	7.5	7.0	7.6	1.0	*
91MIN-F	8.6	6.6	9.1	6.9	6.3	6.6	7.4	7.4	1.0	**
91MONIN	7.9	7.1	8.4	7.0	7.3	6.9	6.9	7.3	0.5	**
92DIJIN	7.0	7.5	8.0	7.1	7.4	7.1	6.6	7.2	0.4	**
92DIJS2	7.3	6.9	7.3	6.8	7.3	6.9	7.0	7.1	0.2	**
91DIJIN	6.9	7.3	6.3	7.3	8.3	6.9	6.5	7.1	0.6	***
92MONIN	6.7	7.4	7.4	6.8	6.7	5.7	6.2	6.7	0.6	***
92ONDIN	6.5	7.1	6.8	7.4	6.8	6.1	6.0	6.7	0.5	***
91DIJS2	6.6	6.7	6.7	6.7	7.2	6.4	6.4	6.7	0.2	***
91ONDIN	6.4	7.1	5.8	7.4	5.8	6.9	5.6	6.4	0.7	***
92OND-F	6.5	7.2	6.1	6.8	6.1	6.6	5.5	6.4	0.5	***
92RENIN	7.5	6.9	5.7	6.2	5.9	6.2	4.6	6.1	0.8	**
91OND-F	6.3	6.4	6.0	6.0	6.0	6.1	4.9	6.1	0.5	**
91REN-F	7.0	7.2	4.2	5.9	5.3	5.9	5.7	5.9	0.9	**
92MON-F	5.8	5.8	5.9	5.2	6.3	4.3	4.7	5.4	0.7	**
91MON-F	6.0	5.3	6.4	4.1	5.7	4.8	5.0	5.3	0.7	**
92REN-F	5.9	6.1	4.3	5.4	4.9	4.9	4.1	5.1	0.7	*
Mean	7.1	6.9	6.7	6.7	6.6	6.3	6.0	6.6		
Std	1.1	0.8	1.5	1.1	1.0	1.0	1.1			
snk grouping at 0.05	*	* *	*	*	*	*	*			
Ecovalence (%SSI)	15.5	10.3	37.9	10.1	12.2	8.2	5.8			

yield was obtained by SOI (7.1 t/ha at 0% moisture content on average) and the lowest by CAR (6.0 t/ha). APO was the most interactive (ecovalence of 37.9% of the total sum of squares of interaction) while CAR was the least interactive (5.8%). SOI displayed intermediate ecovalence (15.5%) (Table 2). For DKN, TAL showed the lowest value and low ecovalence, while APO displayed the highest reduction of kernel number and a high ecovalence (Table 3). This indicated that during the formation of kernel number, genotypes showed distinct behavior: for instance, APO behaved badly in some environments in comparison to TAL. About RTKW, all genotypes reduced greatly their thousand-kernel weight, around 26.5% on average (Table 3). Thus yield was limited for most of the genotypes during grain filling. TAL and SOI were the least interactive for this trait (respectively 7.0% and 8.6% of the ecovalence) while ARM and APO were the most interactive (Table 3). In comparison to early genotypes such as TAL or SOI, ARM and APO were more subjected to stress during the grain-filling period because they are late maturing genotypes. Figure 1 gives a summary of the interactive pattern (from the multiplicative model) for all traits as well as an indication of similarities between genotypes, genotypes with overlapping ellipses behaving similarly. APO showed a high interaction, as it was the most distant from the origin for all the traits. ARM was interactive also but behaved differently than APO. Other genotypes, such as TAL and REC, stood near the origin for most of the traits which corresponded to less-interactive genotypes. Those genotypes behaved similarly as they were always associated.

Analysis of set 1 (12 varieties)

Interactive model

Genotype, environment and genotype-environment interaction observed for grain yield measured on the 12 genotypes in the 18 environments were all significant. The component with the largest effect was the environment. Some genotypes were more interactive than others for grain yield: Apollo, Viking, and Renan accounted for half of the sum of squares of interaction while Baroudeur and Camp-Rémy displayed little interaction.

Determination of the optimal number

To determine the optimal number of probe genotypes to introduce in multi-environment trials of winter wheat, genotype-environment interaction was modelled through several rounds of bi-additive factorial regressions using deviations of yield components measured on the probe genotypes. These deviations explained genotype-environment interaction from 14.7%, the poorest level using a single probe genotype, to 75.1%, the highest level using all seven probe genotypes (Fig. 2). This last value was close to the maximum (77.4%) obtained with a classical bi-additive model (AMMI) using three terms as well.

All 127 combinations were tested: seven for the group of singles, 21 for pairs, 35 for triplets, 35 for quadruplets, 21 for quintuplets, seven for sextuplets and one septuplet. Box-plots displayed the distributions of the

Table 3 Description of the sites: deviations (in %) from KNthreshold (a) and from potentialTKW (b) for the seven probe genotypes in each environment. Std = standard deviation, snk = Student-Newman-Keuls multiple range test

a

Environments	APOdkn	CARdkr	SOIdkn	THEdkr	ARMdkr	RECdkr	TALdkn	Mean	Std	snk grouping at 0.05 probability level
92REN-F	33.9	37.6	26.9	19.1	21.2	24.6	20.6	26.3	6.5	*
92OND-F	37.0	25.9	28.7	24.1	35.1	15.4	10.4	25.2	9.0	**
92MON-F	24.0	31.6	26.0	24.0	16.1	19.4	27.1	24.4	4.7	**
92ONDIN	32.0	22.6	31.3	24.5	30.8	10.2	14.4	23.7	8.0	**
91DIJS2	30.9	18.4	25.5	29.2	13.8	25.0	19.9	23.2	5.7	**
92RENIN	25.6	36.9	16.5	24.2	17.9	17.3	7.9	20.9	8.4	**
91DIJIN	37.7	18.8	25.9	26.6	–0.5	19.8	17.5	20.8	10.8	**
91OND-F	30.8	30.8	17.4	19.9	25.3	19.3	3.1	20.2	8.8	**
91MON-F	17.1	26.5	16.8	17.4	19.7	25.9	16.0	19.9	4.1	**
92DIJIN	22.0	19.2	29.1	6.9	20.9	22.5	10.8	18.8	7.0	**
92DIJS2	21.5	18.7	23.0	22.2	8.0	22.0	12.6	18.3	5.3	**
91ONDIN	29.2	14.4	22.7	13.0	29.6	9.9	–0.2	16.9	10.1	*
92MONIN	10.1	17.3	23.7	16.0	18.2	14.0	18.7	16.8	3.9	*
91REN-F	17.7	12.8	–4.0	–3.4	3.9	5.6	–3.5	4.2	7.9	*
91MIN-F	1.7	8.9	–2.6	6.1	2.9	7.7	3.0	3.9	3.6	*
91MONIN	6.7	9.8	1.0	6.6	13.1	–3.8	–9.2	3.5	7.3	*
91RENIN	20.8	4.9	–7.1	–11.0	–5.3	–4.2	–8.4	–1.5	10.2	*
91MININ	3.0	4.3	–1.0	2.8	–12.3	–1.4	–7.2	–1.7	5.6	*
Mean	22.3	20.0	16.6	14.9	14.4	13.8	8.5	15.8		
Std	10.9	9.8	12.7	11.0	12.5	9.5	10.7			
snk grouping at 0.05 probability level	*	*	*	*	*	*	*			
Ecovalence (%SSI)	20.8	10.1	10.2	9.8	26.0	9.4	13.8			

b

	TALrtkw	RECrtkw	CARrtkw	ARMkw	THErtkw	APOrtkw	SOIrtkw	Mean	Std	
91REN-F	42.0	41.0	36.4	50.7	35.2	55.3	38.7	42.7	7.0	*
91MON-F	45.0	48.3	33.4	33.3	42.1	32.3	37.6	38.9	5.9	**
92REN-F	39.4	32.8	36.2	45.3	32.2	42.8	30.0	36.9	5.3	**
92MON-F	43.0	39.6	32.6	28.1	31.4	31.1	31.6	34.2	4.9	**
91OND-F	38.6	29.6	31.3	22.3	27.5	22.8	33.2	29.2	5.3	**
91MIN-F	34.2	29.7	20.3	37.3	36.2	19.0	25.0	28.8	7.0	**
92RENIN	35.0	29.7	27.6	36.4	18.2	32.7	21.7	28.8	6.3	**
91ONDIN	32.8	22.4	36.4	22.4	26.4	27.9	27.4	28.0	4.8	*
91MONIN	32.7	34.2	25.6	20.1	30.9	21.2	30.3	27.9	5.2	*
91RENIN	26.4	25.1	28.1	31.7	29.2	35.4	17.0	27.6	5.3	*
92MONIN	31.2	25.6	27.1	24.0	20.1	28.1	23.0	25.6	3.4	*
92OND-F	28.5	24.5	27.6	9.9	13.9	15.0	20.2	19.9	6.7	*
92DIJS2	23.3	17.8	15.2	24.1	19.6	19.1	17.3	19.5	3.0	*
91DIJS2	21.8	16.3	23.3	20.6	14.8	15.1	22.7	19.2	3.4	*
92ONDIN	30.9	22.4	24.4	7.4	15.3	12.1	17.4	18.6	7.4	*
92DIJIN	22.0	14.3	20.1	13.6	26.9	9.7	14.1	17.2	5.5	*
91DIJIN	18.4	13.9	21.1	19.9	10.3	12.1	19.2	16.4	3.9	*
91MININ	16.6	11.4	16.2	16.8	17.4	10.2	16.9	15.1	2.7	*
	31.2	26.6	26.8	25.8	24.9	24.5	24.6	26.4		
Std	8.3	9.9	6.5	11.3	8.7	12.0	7.3			
snk grouping at 0.05 probability level	*	*	*	*	*	*	*			
Ecovalence (%SSI)	7.0	10.0	13.0	24.7	15.8	21.0	8.6			

combinations of a given size (Fig. 2). The %SSI increased with the number of probe genotypes of course, and the variability within combinations of a given size decreased. Singles and pairs showed the lowest medians for the %SSI explained by the model (respectively 19.6% and 45.5%) and the largest ranges (22.4% and 24.7%). From triplets to sextuplets, medians were all above 50% (respectively 57.2%, 64.7%, 70.3% and 73.2%) and ranges were all below 20% (respectively 18.3%, 16.6%, 8.9% and 2.6%). This clearly showed that probe genotypes did not play the same role. Some combinations were significant as the %SSI explained was above the critical value estimated from the distribution of the 1,000 samples created by the random permutation

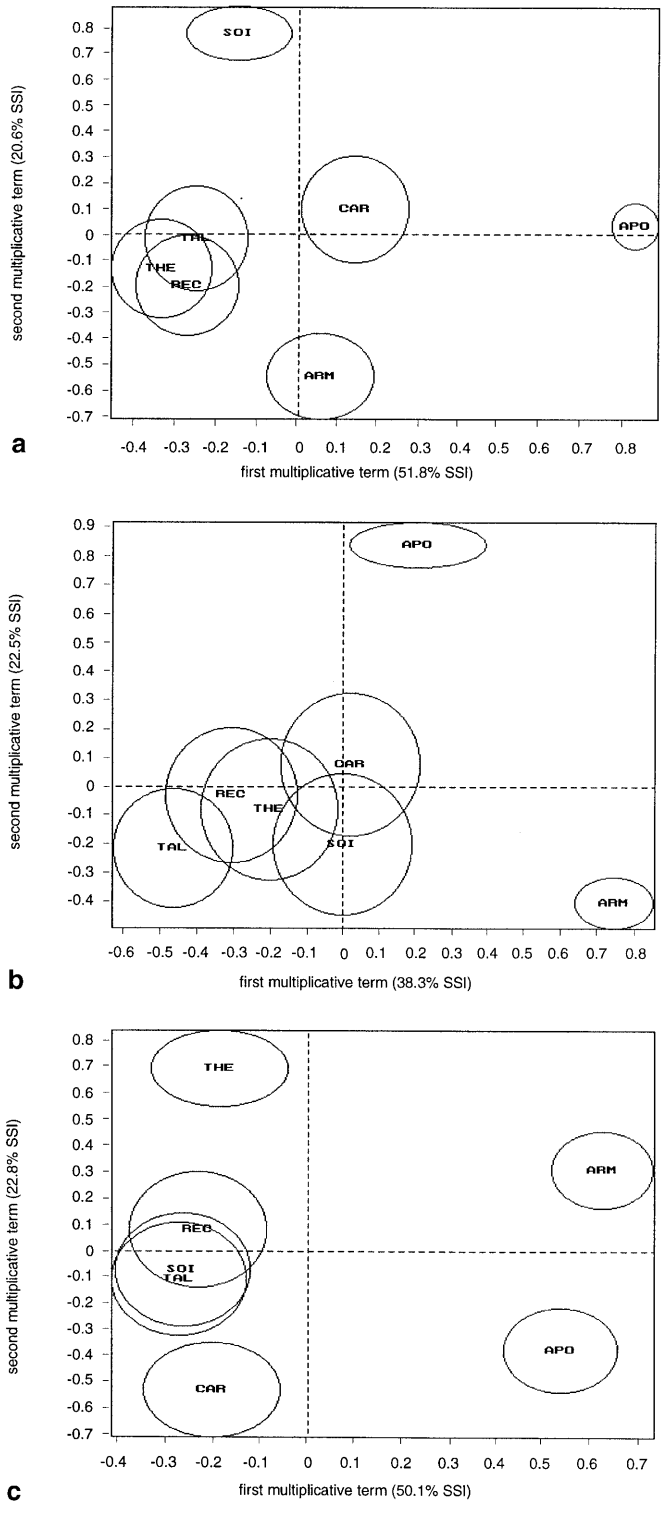

Fig. 1a–c Analysis of the seven probe genotypes. Multiplicative scores γ_{g1} and γ_{g2} with a bi-additive model of two terms. Grain yield (**a**), deviation of kernel number (**b**), and reduction of thousand-kernel weight (c). Ellipses, at the 0.05 probability level, indicate approximate confidence intervals by the span of their vertical and horizontal axes

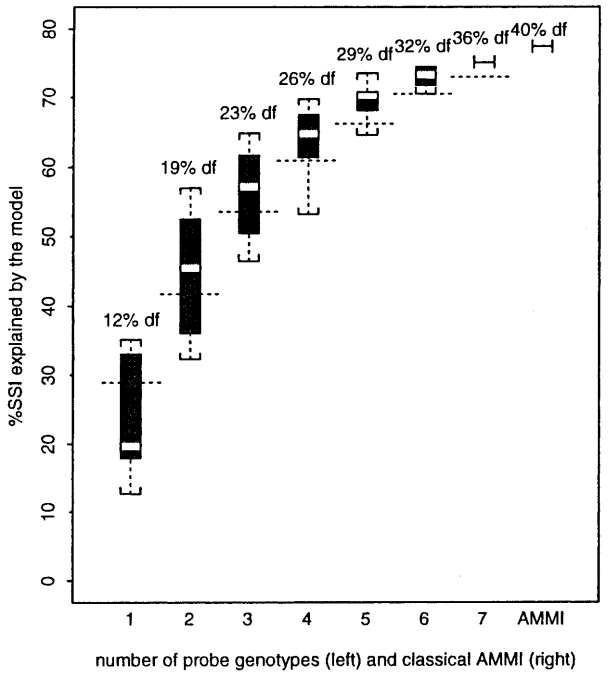

Fig. 2 Boxplots of %SSI (sum of squares of interaction) according to the size of the combination (from 1 to 7). For each boxplot, the computed critical value is symbolized with an horizontal dotted line and the degrees of freedom associated to the modelled interaction are given above. The classical AMMI model including three terms is given at the end

procedure. This critical value was symbolized for each combination with a dotted horizontal line on Fig. 2 and Fig. 3. The proportion of significant combinations increased with the number of probe genotypes: 29% for singles, 66% for pairs, 69% for triplets, 77% for quadruplets, 95% for quintuplets and nearly 100% for the remaining ones. Under four or five genotypes, the proportion of non-significant combinations could exceed 25% and thus be too high (Fig. 2). Over five genotypes, the proportion of interaction explained by the model became too small in comparison to the consumption of degrees of freedom (Fig. 2). Without additional information on the probe genotypes, it is thus difficult to choose a small number of probe genotypes because of the non-significant proportions (34% of the pairs are non-significant for instance). When no information is available on the probe genotypes, the safe number to introduce would be five because most of the combinations were significant. Such a model including five probe genotypes could be seen to be not enough parsimonious. But it is more parsimonious than the classical AMMI model with a consumption of 29% degrees of freedom against 40% for an AMMI

Fig. 3 Boxplots of %SSI (sum of squares of interaction) according to the size of the combination (from 1 to 7) and within each size, the ones containing each of the seven probe genotypes indicated on the x-line. For each size, the *dotted lines* indicate the critical values as in Fig. 2 and the corresponding degrees of freedom are given *in brackets* ▶

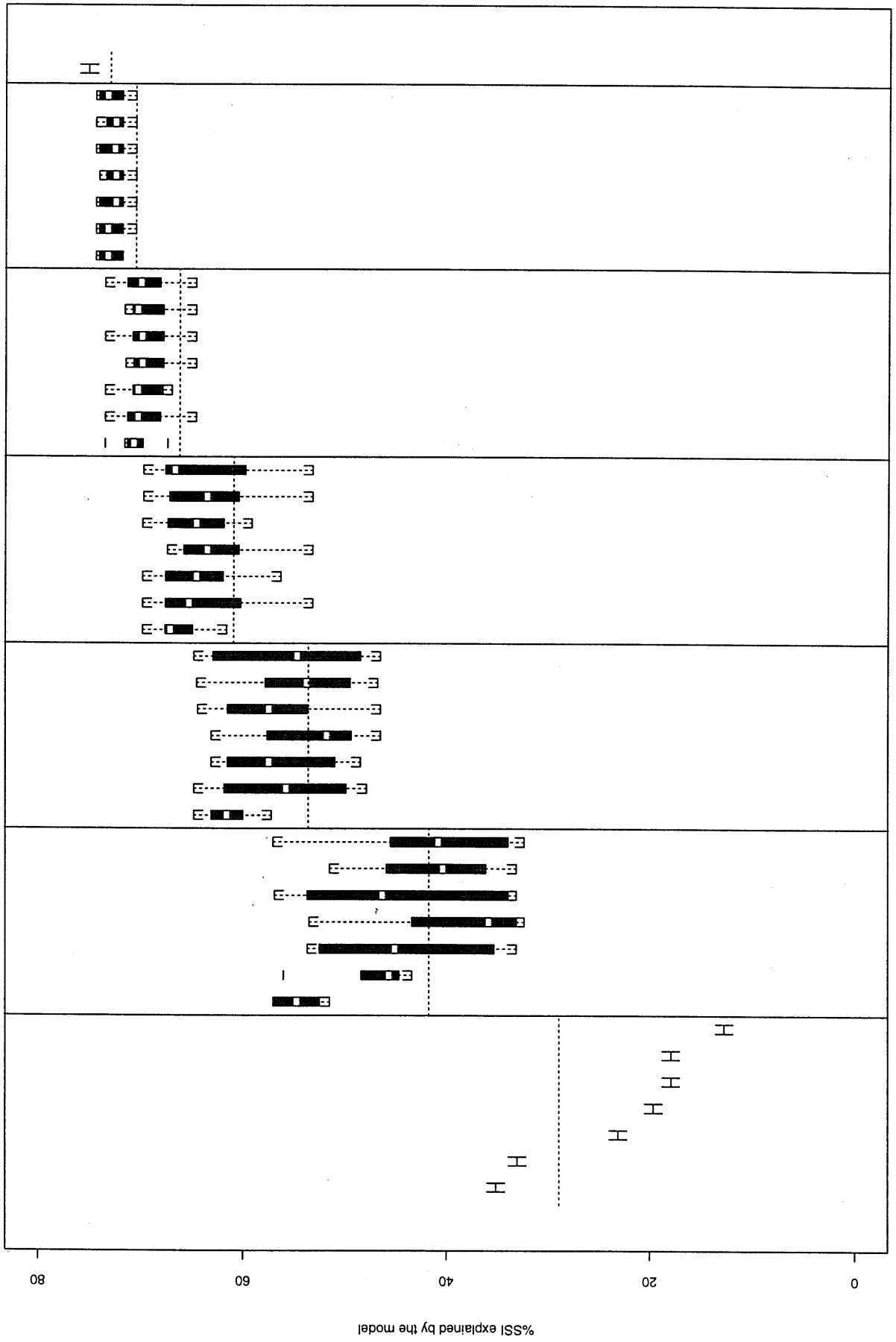
%SSI explained by the model
0
20
40
60
80
apo car soi tal the arm rec apo car soi tal the arm rec apo car soi tal the arm rec apo car soi tal the arm rec apo car soi tal the arm rec all
1 probe genotype (12% df)
2 probe genotypes (19% df)
3 probe genotypes (23% df)
4 probe genotypes (26% df)
5 probe genotypes (29% df)
6 probe genotypes (32% df)
7 (36% df)

Table 4 %SSI explained by each pair and corresponding rank for the two sets: set 1 = varieties and set 2 = lines. The degrees of freedom associated with the modelled interaction are given in brackets. From left to right, genotypes are ranked by increasing earliness

	Genotype							Set 1 (19% *df*)		Set 2 (26% *df*)	
	ARM	APO	CAR	THE	SOI	TAL	REC	%SSI	Rank	%SSI	Rank
		x					x	57.0[a]	1	58.9	4
		x		x				56.9	2	59.2	3
		x	x					56.0	3	57.4	6
				x	x			53.7	4	53.0	8
		x				x		53.5	5	57.7	5
		x			x			52.5	6	62.6	1
	x	x						51.5	7	60.3	2
			x	x				48.4	8	47.9	17
	x		x					45.9	9	50.8	13
					x		x	45.5	10	52.5	10
			x				x	45.5	11	49.3	15
			x		x			44.7	12	52.6	9
	x			x				44.3	13	51.1	12
			x			x		43.3	14	42.4	20
	x					x		36.4	15	51.2	11
	x						x	36.1	16	50.1	14
					x	x		35.3	17	48.0	16
				x			x	33.9	18	42.9	19
				x		x		33.1	19	43.0	18
	x				x			33.1	20	56.7	7
						x	x	32.3	21	41.2	21
Mean rank (set 1)	13	4	10	11	12	15	13				
Mean rank (set 2)	10	4	13	13	9	15	14				

[a] For example, 57.0 is the %SSI explained by the pair (APO, REC)

model including three terms as well (Fig. 2). A compromise is thus needed between the probability of choosing a significant combination and the number of probe genotypes. If too numerous, their experiment is too costly because specific experiments are needed for determining the potential values of TKW and KN. If too small, the risk of choosing poor combinations is too high as shown earlier. From an experimental and statistical viewpoint, the observation of five probe genotypes could be too costly and the optimal number would probably be reduced between two to four. More information is then needed on the probe genotypes to reduce their number and choose them properly. Thus, it was important to examine the prerequisites for a good choice of probe genotypes in a small number and for explaining significantly the interaction.

Determination of the main characteristics

Very often, information is available on the probe genotypes prior their first multilocation experiment and this could modify the way of choosing them. Considering the results of the different groups for each probe genotype, it is interesting to determine the genotypes involved in the significant combinations. The good performance of APO was highlighted for any kind of combination (box-plots of Fig. 3). For singles (left part of Fig. 3), results varied from 35.1% with APO to only 12.7% with REC and significant results were obtained only by APO and CAR. Among pairs, APO was the first (the best partitioning being 57.0%) while the variability between pairs containing APO was small. CAR was the second with smaller but still significant results. In contrast, most pairs with TAL provided the poorest performances, the smallest partitioning being 32.3% (Fig. 3). There was huge variability for pairs including the genotypes such as REC, THE, TAL and SOI, and ARM to a less extent. For triplets and quadruplets, the advantage of combinations containing APO was still noticeable, but to a lesser extent than previously. The advantage of combinations containing APO decreased greatly for quintuplets and sextuplets. These results showed that the efficiency of a given combination was clearly influenced by the probe genotypes involved and their complementary characteristics.

As the %SSI varied with the probe genotypes, some genotype characteristics might be involved. One important characteristic would be found in the interaction pattern (or instability). For grain yield and deviations of yield components, it is important to note that APO is the most-interactive genotype (Fig. 1). It would have been interesting to observe Viking or Renan, two other interactive genotypes, but no potential values of TKW and KN were available for them. In contrast, the probe genotypes CAR, TAL or REC were observed as less interactive for the same traits (Fig. 1).

Earliness was probably another important criterion. In Table 4, one can note that pairs containing early genotypes, with the exception of (APO, REC), usually show small values. The smaller value of partitioning is given by (REC, TAL), a pair of two early genotypes. In contrast, two other pairs containing genotypes of the same earliness (APO and ARM which are late genotypes, or SOI and THE which are both 1/2 early) obtained a good performance (respectively 51.5 and 53.7%SSI). The good performance of the pair (APO, ARM) was associat-

Table 5 %SSI explained by each triplet and corresponding rank for the two sets: set 1 = varieties and set 2 = lines. The degrees of freedom associated with the modelled interaction are given in brackets

Genotype							Set 1 (23% *df*)		Set 2 (31% *df*)	
ARM	APO	CAR	THE	SOI	TAL	REC	%SSI	Rank	%SSI	Rank
	X	X				X	64.8[a]	1	64.1	10
X	X					X	64.5	2	65.7	5
	X	X	X				64.4	3	64.1	9
	X				X	X	63.1	4	60.1	20
		X	X	X			63.1	5	61.0	17
	X		X			X	62.9	6	62.2	15
	X			X		X	62.6	7	66.4	4
X	X	X					61.8	8	67.7	2
	X	X		X			61.6	9	69.5	1
	X		X	X			61.5	10	64.4	8
X	X		X				61.0	11	64.0	11
	X		X		X		60.3	12	60.7	18
	X	X			X		60.0	13	63.8	14
			X	X		X	58.8	14	55.4	29
X	X			X			57.8	15	67.1	3
	X			X	X		57.6	16	64.8	7
			X	X	X		57.5	17	56.3	27
X	X				X		57.2	18	64.8	6
X			X	X			56.6	19	64.0	12
		X	X			X	55.9	20	53.9	31
X		X		X			55.8	21	61.5	16
		X		X		X	54.7	22	59.8	22
X		X	X				53.8	23	59.3	24
X			X		X		53.7	24	60.0	21
		X	X		X		51.8	25	51.0	34
X				X	X		51.0	26	63.9	13
X			X			X	50.5	27	53.5	32
		X		X	X		49.9	28	57.2	25
X		X			X		49.5	29	59.3	23
				X	X	X	49.4	30	53.9	30
X		X				X	49.1	31	56.7	26
X				X		X	48.5	32	60.7	19
		X			X	X	47.9	33	53.1	33
X					X	X	46.8	34	56.1	28
			X		X	X	46.5	35	45.8	35

[a] For example, 64.8 is the %SSI explained by the triplet (APO, CAR, REC)

ed with the presence of APO. For (SOI, THE), these genotypes were not especially interactive for DKN or RTKW (Table 3). The way they formed their yield could provide complementary behavior: THE produced big grains in a small number while SOI yielded small grains in higher number.

These conclusions were also supplied by examining the triplets (Table 5). Considering three genotypes, APO was very often associated with CAR, REC or THE in the first triplets. This is in agreement with the previous results where (APO, REC), (APO, THE) and (APO, CAR) constituted the best previous pairs. The best triplet used APO, REC, and CAR. The three previous criteria (interaction pattern, earliness, and pattern for the formation of yield) might be of importance once again as these characteristics were complementary for the genotypes under consideration. In addition, CAR appeared as an interesting choice due to its susceptibility to lodging, revealing conditions where lodging is a main limiting factor of yield.

In conclusion, four or five probe genotypes resulted in a convenient number without considering the genotype characteristics. The comparisons of %SSI to critical levels helped to define this number but this single criterion did not take into account the problem of parsimony as the procedure will end up with a model including too many covariates. The optimal number of probe genotypes could be reduced to two, three or four if taking into account the information available on probe genotypes, such as interaction pattern, earliness, and pattern for the formation of yield. Such numbers are more interesting because they will lead to more-parsimonious models and to less-expensive experiments.

Analysis of set 2 (ten lines)

Influence on the optimal number

A table of 12 genotypes × 18 environments was previously examined. For other tables, the probe number must be revised because it will vary according to the size of the table. The %SSI decreased to 5–7% when the size of the table increased from 117 degrees of freedom to 273 (Fig. 4). Singles do not follow this rule (Fig. 4). For all sets, a good partitioning of the interaction could be obtained with two to four genotypes.

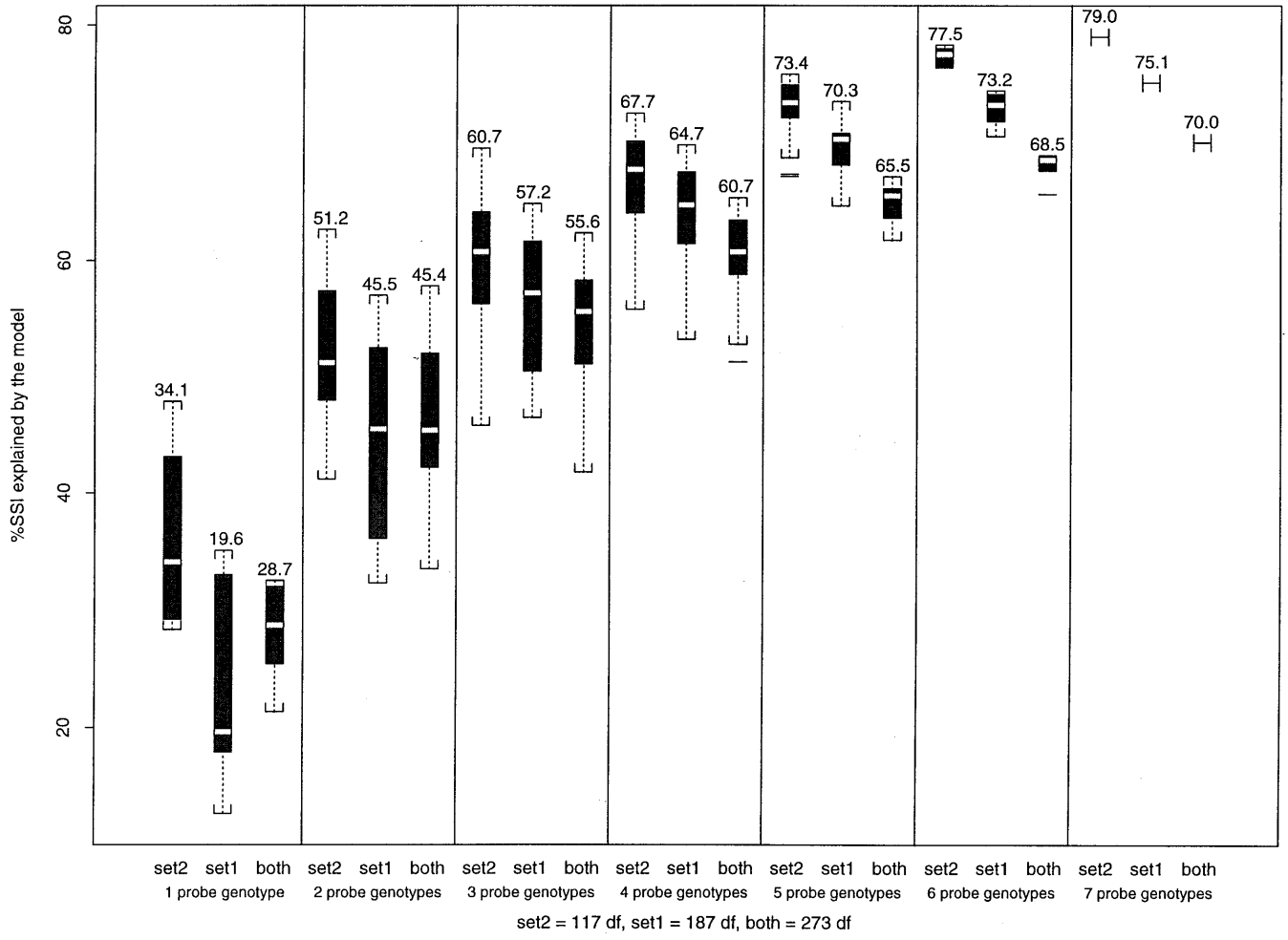

Fig. 4 Boxplots of %SSI (sum of squares of interaction) according to the size of the combination (from 1 to 7) and according to two sets (set2, set1 and both) for each size. The median of each boxplot is given on top

Influence on the characteristics

The %SSI could be compared by the correlations between the two sets (Table 6). The same combinations of environmental covariates (within pairs to quintuplets) are compared in the table. Pairs and triplets showed the highest significant correlations (respectively 0.71 and 0.69) while quadruplets and quintuplets displayed the lowest (respectively 0.53 and 0.43). Regarding the pairs, pairs with contrasting results kept their rank in the two sets: this is the case for pairs containing APO which were the first ones (average rank of 4 in both sets) and those containing REC and TAL which ranked at the end (Table 4). The four other probe genotypes, ARM, CAR, THE and SOI, obtained an intermediate %SSI in both sets and their ranking was not as stable as the ranking of APO, REC and TAL. This was particularly the case for the pair (ARM, SOI) that was the most distinct. In contrast, it can be noticed that the pair (THE, SOI) showed a similar performance between the two sets.

Table 6 Correlations for %SSI between the sets, given for the same subsets among pairs, triplets, quadruplets, and quintuplets

Item	Number of observations	Sets	Set 1 (varieties)	Set 2 (lines)
Pairs	21	Set 1	–	
		Set 2	0.71*	–
		Set 1 + set 2	0.81*	0.81*
Triplets	35	Set 1	–	
		Set 2	0.69*	–
		Set 1 + set 2	0.80*	0.74*
Quadruplets	35	Set 1	–	
		Set 2	0.53*	–
		Set 1 + set 2	0.72*	0.71*
Quintuplets	21	Set 1	–	
		Set 2	0.43*	–
		Set 1 + set 2	0.64*	0.70*

* Significant at the 0.05 probability level

In conclusion, most observations were supplied by the two sets. Interaction pattern seemed to be the most-important criterion because APO, the most-interactive probe genotype for grain yield, deviation of kernel number and reduction of thousand-kernel weight, was always among the superior combinations. This was also support-

ed by the lesser performance of ARM, which was as late as APO. Earliness was another criterion to consider. It could be assumed that the earliness of probe genotypes has to be in accordance with the earliness of the genotypes under consideration. For instance, the bad performance of early probe genotypes in set 2 could be expected as there were few early genotypes in this set of lines in comparison to set 1.

Discussion

To reduce the number of probe genotypes (for experimental and statistical costs), probe genotypes have to be complementary and, from the previous observations, it seems that the three previous criteria, namely the interaction pattern, earliness and the pattern for the formation of yield, need to be considered at the same time for a convenient choice. Thus a reduced number will depend on the information available on the probe genotypes. When the three previous criteria are available on each probe genotype (and this will be a more-frequent event than the total lack of pre-existing information), the number of probe genotypes could be reduced to three. In the present study, it is shown that a triplet, containing well-chosen genotypes, can provide a good understanding of what happened in the environments under study. When the choice is getting difficult due to the lack of pre-existing or accurate information, quadruplets are more convenient because genotypes are more complementary in such a subset than in a triplet. For newly released cultivars, more information would be needed about the interaction pattern. This criterion, i.e. earliness at heading, could be regularly obtained in official trials conducted for the registration of the cultivars, while earliness at other stages (such as the onset of stem elongation, earliness at maturity) or pattern for the formation of yield are given by trials conducted after the registration of the cultivars.

Earliness is a very interesting criterion for establishing a variety type based on the behavior of the probe genotype. Varieties can be compared to a probe genotype of the same earliness because it could be assumed that genotypes of the same earliness would be subjected to similar environmental conditions. For that reason, it is more convenient to choose probe genotypes with an earliness similar to that of the varieties under analysis.

It is important to note also that the number of probe genotypes to include in an experiment could vary with the sowing-to-harvest duration of the species and the extent of the area of cultivation. The winter wheat sowing-to-harvest duration is quite long (8–11 months in France) in comparison to spring species such as soybean or spring wheat. It is also cultivated on a wide area. For these reasons, the number of probe genotypes could be greater for winter wheat than for spring species where the optimal number of probe genotypes could be reduced. With soybean, Desclaux (1996) used growth and developmental variables of only two soybean genotypes to characterize *a posteriori* environmental conditions. Cultivars of soybean in France are cultivated in smaller areas than those for winter wheat. This could explain why fewer probe genotypes would be sufficient for a good explanation of genotype-environment interaction. For maize (*Zea mays* L.), although hybrids are also cultivated in several smaller areas in France, this number would also depend on the magnitude of the genotype-environment interactions that seem higher than for soybean (Giauffret, personal communication).

From the present results obtained on wheat, earliness and interaction pattern are the two criteria that could concern other species. As for winter wheat, earliness enables comparisons between genotypes of the same earliness. It can be assumed also that coupling interactive and non-interactive genotypes would provide a good explanation of genotype-environment interaction. The third criterion, i.e. pattern for the formation of yield, might be more specific to winter wheat. A model using yield components is not convenient for maize due to the lack of tillering. Physiological models, such as those based on the light interception, for instance could be more efficient for maize (Giauffret, personal communication).

In such an approach, the results could be influenced by the quality of the available criteria on the probe genotypes. Nevertheless, it can be assumed that the diversity of the 18 environments over the 2 years (Brancourt-Hulmel et al. 1999) allowed a good assessment of the instability and that biases were small. Deviations of yield components were also determined with enough accuracy (Brancourt-Hulmel et al. 1999). In case of a small environment effect, more than 2 years would have been needed. When similar criteria are available, this approach using experimental data and simulations could be helpful to define candidates for becoming probe genotypes or to determine the optimal number of probe genotypes in species other than winter wheat.

Acknowledgements We appreciate the careful technical assistance of Paul Bataillon, Denis Beghin, Michel Leleu and Claude Sausseau. We also express thanks to Eric Hanocq and Jacques Le Gouis for their careful reading of the manuscript and to Catherine Giauffret for her helpful suggestions. We thank the reviewers for many useful improvements to the initial version.

References

Brancourt-Hulmel M (1999) Crop diagnosis and probe genotypes for interpreting genotype-environment interaction in winter wheat trials. Theor Appl Genet 99:1018–1030

Brancourt-Hulmel M, Lecomte C (1994) Sélection et stabilité du rendement chez le blé tendre d'hiver. Agronomie 14:611–625

Brancourt-Hulmel M, Lecomte C, Meynard JM (1999) A diagnosis of yield-limiting factors on probe genotypes for characterizing environments in winter wheat trials. Crop Sci 39:1798–1808

Brancourt-Hulmel M, Denis JB, Lecomte C (2000) Determining environmental covariates which explain genotype-environment interaction in winter wheat through probe genotypes and bi-additive factorial regression. Theor Appl Genet 100:285–298

Decoux G, Denis JB (1991) Intera. Logiciels pour l'interprétation statistique de l'interaction entre deux facteurs. Laboratoire de biométrie, Inra, Route de Saint-Cyr, 78026 Versailles cedex, France

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

Denis JB (1991) Ajustement de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. Rev Stat Appliquées 34:5–24

Denis JB (1998) BIAREG Splus functions to perform bi-additive regressions. Technical report, Laboratoire de Biométrie, INRA, Route de Saint-Cyr, F-78026 Versailles Cedex

Desclaux D (1996) De l'intérêt de génotypes révélateurs de facteurs limitants dans l'analyse des interactions génotype × milieu chez le soja (*Glycine max*. L. Merill). Thèse de doctorat, Institut national polytechnique de Toulouse

Eeuwijk FA van (1995) Linear and bilinear models for the analysis of multi-environment trials: I. An inventory of models. Euphytica 84:1–7

Fisher RA (1935) The design of experiments. Oliver and Boyd, Edingburgh

Gauch HG (1992) *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Elsevier, Amsterdam

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

ITCF (1998) Choisir ses variétés. Céréales Nord Picardie. ITCF, Paris, France

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

Scheffé H (1959) The analysis of variance. Wiley, New York

Wood JT (1976) The use of environmental variables in the interpretation of genotype-environment interaction. Heredity 37: 1–7

Wricke G (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Z Planzenzücht 47:92–96