Theor Appl Genet (1999) 98: 234–242

L. Moreau · H. Monod · A. Charcosset · A. Gallais

Marker-assisted selection with spatial analysis of unreplicated field trials

Received: 7 January 1998 / Accepted: 20 July 1998

Abstract Many studies have shown that molecular markers can improve the efficiency of the selection of quantitative traits in plant breeding provided that large population sizes are used. As a way to limit experimental costs it appears that the use of unreplicated trials may be more valuable than the use of replicated plots in one trial. In this particular context of unreplicated large trials, spatial heterogeneity within the field may reduce the efficiency of the selection. The problem of controlling spatial heterogeneity was seldom considered in the case of marker-assisted selection (MAS). Here, we propose an integrated method to predict genetic values considering simultaneously marker information and possible spatial heterogeneity. This method was applied to a population of 300 F_3 lines of maize evaluated in 11 unreplicated trials for grain yield. We show that when spatial field heterogeneity is considered through appropriate statistical models the accuracy of genetic value predictions is improved and the same genetic gain can be achieved with a reduced number of trials.

Key words Marker-assisted selection · Spatial analysis

Introduction

The use of markers to improve the prediction of genetic values for quantitative traits has received extensive interest in the recent past. Lande and Thompson (1990) proposed to select individuals on an index that included both phenotype and 'molecular score' (obtained by the multiple regression of phenotype on marker type). Since 1990, the efficiency of this marker-assisted selection (MAS) relative to purely phenotypic selection has been widely studied, in the case of populations derived from the cross of two inbred lines, through analytical approaches (Lande and Thompson 1990, Luo et al. 1997; Moreau et al. 1998) and simulations (Zhang and Smith 1992, 1993; Gimelfarb and Lande 1994, 1995; Whittaker et al. 1995; Hospital et al. 1997). One of the main results of these studies is that MAS has been determined to be efficient only for large population sizes (the population size required depends on the heritability of the trait, but is generally above 200). For small population sizes, the power of detection of the associations between markers and QTLs (quantitative trait loci) is small, and QTL effects associated with the markers are poorly estimated. This result implies that both phenotypes and genotypes at marker loci must be evaluated for numerous lines (individuals or progenies). Knapp and Bridges (1990) showed that for any given experimental means, the resolution of QTL mapping experiments is improved by increasing the number of lines and evaluating them only once, rather than using a smaller population size and replicated evaluations. Moreover, because of possible genotype × environment interactions, phenotypes must be evaluated at different locations. In this case, for a given number of plots, the optimum is one replication per location (Weber 1980). All these elements suggest that it may be more valuable to evaluate the lines in several unreplicated field trials than to replicate them in a smaller number of locations.

When a large number of entries is tested in a field trial, growing conditions may vary throughout the trial area, leading to a decreasing accuracy of the performance estimation and therefore to a reduced genetic gain. It is therefore important to control spatial heterogeneity. Methods adapted to unreplicated trials have

Communicated by G. Wenzel

L. Moreau (✉) · A. Chorcosset · A. Gallais
INRA-UPS-INA-PG, Station de Genétique végétale,
Ferme di Moulon, F-91190 G.F-sur-Yvette, France
E-mail: moreau@moulin.inra.fr

H. Monod
INRA Laboratoire de biométrie, Route de St-Cyr,
F-78026 Versailles cedex, France

been developed [see Kempton (1984) for a review]. In these models, genotypes can be compared with local checks or with neighbouring genotypes. More recently, methods derived from time-series analyses have been developed and can combine these two types of approaches (Cullis et al. 1989, Kempton and Gleeson 1997).

While a definite interest in controlling spatial heterogeneity to increase genetic gains has been fully demonstrated (see Cullis et al. 1992) such heterogeneity is very often neglected in QTL detection experiments. For instance, computer programmes for QTL detection based on 'interval mapping' do not take into account local field variability. In the case of spatial heterogeneity, using these programs after adjusting means by a spatial analysis can bring a significant improvement. However, this may not be optimal since the information coming from the field trial is not fully used when estimating effects associated with markers. An integrated one-step approach seems better than proceeding in two steps.

In this paper, we present such an integrated approach to predict the genetic values of genotypes derived from a cross between two homozygous lines. The method is based on a mixed model incorporating local field variations and information on markers. It has been applied to 11 unreplicated trials sown in several locations.

Materials and methods

Plant material and evaluation of grain yield

The plant material has been derived from a cross between two homozygous maize lines: F2, an early European flint line, and F252, an early dent line of USA origin. From the hybrid, 300 F_3 plants were obtained by single seed descent (SSD). Each F_3 was self-fertilized to obtain an F_4 family.

The F_4 families were evaluated with a progeny test. Each family was crossed with a third homozygous line, MBS847, a dent line of USA origin but unrelated to F252, used as a tester. In 1995, the progenies were sown in 14 trials in 11 different locations in the north of France as indicated in Table 1. This experimental network was adapted to the earliness of the population. Among the 14 trials, arbitrarily denoted 1 to 14 from West to East, 3 were conducted under stressed conditions: low nitrate input (trial 10), early sowing to induce low temperature stress during germination (trial 11) and drought stress (trial 3). In each location where a stress was applied, another trial was conducted under standard conditions. The other trials were conducted following standard agricultural practices adapted to the location (with irrigation for trials 2, 4, 5, 6, 7, 8 and without for the other trials). As seed stocks were not sufficient to test all the families in all the trials, only a subset of 280 was evaluated in a given trial. In the whole experimental design, each of the 300 genotypes was evaluated in at least 3 trials, and 243 genotypes were grown in all the trials.

Elementary plots consisted of 2 seeded rows, spaced 0.8 m apart and, depending on the location, 5–6 m long. Each trial was divided into 17 blocks of 20 plots (19 blocks of 18 plots in trial 5). Each block consisted of two rows of 10 plots (2 rows of 9 plots in trial 5). Among the 20 (18) plots within each block, 2 were sown with the single-cross hybrids F2 × MBS847 and F252 × MBS847 (parental lines evaluated in test) used as checks. Thus, each check was replicated 17 (or 19) times per trial, and the replications were evenly spread throughout the field. Among the 280 families evaluated in a given trial, a random sample of 26 (24 in trial 5) was replicated twice. These replicated genotypes were allocated to the blocks to form an efficient incomplete block design. The other 254 (256 in trial 5) genotypes were not replicated and were allocated at random throughout the trial. Thus, a total of 340 (or 342 in trial 5) plots was sown in each trial. As far as possible, the field design was composed of 10 rows of 30 plots and 2 rows of 20 plots. However, the shape of the trials was different in several trials. This was due to specific constraints such as the system of irrigation which limited the maximum number of plots in each row.

Each plot was harvested in bulk to evaluate the grain yield. Among the 14 trials, trial 3, submitted to drought conditions, was so severely stressed that only a few plants yielded grain. Because of numerous problems during germination or early growth, two 'unstressed' trials (trials 2 and 13) were discarded. Consequently, our study was then limited to only 11 trials.

Restriction fragment length polymorphic (RFLP) markers

Based on a map previously developed by Causse et al. (1996) from four maize populations, including a population obtained from the

Table 1 Locations of the 14 trials sown in 1995

Number of trial	Location (department)	Specific conditions	Discarded trials
1	Betton (35)	Standard conditions	
2	St Martin de la place (49)	Standard conditions	Discarded
3	Lusignan (86)	No irrigation	Discarded
4	Lusignan (86)	Standard conditions	
5	Blois (41)	Standard conditions	
6	Chataudun (28)	Standard conditions	
7	Senneville (28)	Standard conditions	
8	Fresnay-le-gilmet (28)	Standard conditions	
9	Gif-sur-Yvette (91)	Standard conditions	
10	Gif-sur-Yvette (91)	Low nitrate input	
11	Mons-en-Chaussée (80)	Early sowing	
12	Mons-en-Chaussée (80)	Standard conditions	
13	Provins (77)	Standard conditions	Discarded
14	Geudertheim (68)	Standard conditions	

hybridization of F2 and F252, RFLP probes were chosen to give polymorphic molecular markers evenly spread on the chromosomes. In 1995, the RFLP analyses were conducted by Linkage Genetics Inc. F_4 families were sown, and genomic DNA was extracted from the leaves of approximately 20 plants per family. The selected probes were used to detect polymorphism following classical procedures [see Causse et al. (1996) for more details] and generated 84 markers. At each codominant marker, genotypes were scored $+1$ for lines that were homozygous F2; -1 for lines that were homozygous F252; and 0, for heterozygous lines. At dominant markers, if the band was carried by F252, genotypes were scored $+1$ when the band was absent and $-2/3$ when present. When the band was carried by F2, genotypes were scored -1 when the band was absent and $+2/3$ when present. These scores correspond to the expected number of F2 alleles, given the genotype at the marker, minus 1 (the expected number of F2 alleles in the whole population).

A genetic map was constructed from the marker data set but will not be presented. Map distances were only used to replace the occasional missing marker types (about 2% of the data) by their expectation, conditional to marker types at flanking markers (as is advocated by Martinez and Curnow 1994). When predicting the genetic values, we considered all markers to be potential covariates without taking into account that markers located on the same chromosome were not independent.

Model for genetic value prediction

The usual model used for predicting the genetic values of the genotypes of a population is:

$$y_i = \mu + g_j + \varepsilon_i \tag{1}$$

where y_j is the yield performance in the plot i, μ is the average yield over the trial, g_j is the random genetic effect of genotype j and ε_i is the effect of plot i assumed to be a random error term. The effect g_j and ε_i are all assumed to be independent.

The proposed model follows the same structure as model 1, but with a more refined modelling of the genetic and plot effects.

$$\begin{cases} y_i = \mu + \tau_j + \xi_i \\ \qquad \text{if } j \text{ is a check,} \\ y = \mu + \sum_m a_m \theta_{mj} + g_j^* + \xi_i \\ \qquad \text{if } j \text{ is a genotype of the population} \end{cases} \tag{2}$$

For the checks, the genetic effect τ_j is considered as fixed (see Cullis et al. 1989). For the genotypes of the population, the genetic effect is subdivided into: the fixed effect $M_j = \sum_m a_m \theta_{mj}$, called the 'molecular score', plus g_j^* which is a random genetic effect not taken into account by the markers. In the molecular score, θ_{mj} is the score of the genotype j at marker locus m, and a_m is the additive effect of the F2 allele associated with the marker m. Ideally, only those markers linked to a QTL should be included in M_j. The term ξ_i is the effect of the plot i. It includes experimental error and local trend in the case of spatial field heterogeneity.

The choice of markers and of trend model is discussed below. Once it is made, the full model 2 is a mixed model and can be analysed by the restricted maximum likelihood (REML) method (Cullis et al. 1989; Stroup and Mulitze 1991). The programme TWOD (Gilmour 1992) was used through the TwoD procedure of the GENSTAT software package (Gilmour et al. 1995).

This general framework allows three different genetic values to be estimated for each genotype in each trial:

1) Phenotypic (P) method: if no marker is included in model 2 ($m = 0$), the genetic value is predicted by the best linear unbiased predictor (BLUP) of g_j^*, which corresponds to g_j in this case.

2) Markers-only (M) method: molecular score $\hat{M}_j$ is used as a predictor of the genetic value.

3) Combined (C) method: including markers in model 2, the genetic value is predicted with the sum of the estimated molecular score $\hat{M}_j$ and the BLUP of g_j^*. This corresponds to MAS as defined by Lande and Thompson (1990).

The aim of this paper is to evaluate whether the combined method (C) can be improved by taking into account the spatial trend heterogeneity. The P and M methods are considered as references.

Choice of markers and of the model for plot effects

The main problem of the proposed approach is how to select the trend model and the markers for each trial. For the sake of simplicity and because there was no statistical programme adapted to our case, the choice of the model was made in two steps. First, we chose the model for plot effects ξ_i using model 2 with no marker covariate ($m = 0$). Then, we kept this trend model and selected the markers to be included in the full model. Even if not optimal this procedure is valid since genotypes were randomly located in the field without taking into account their genotypes at marker loci.

Spatial method for modelling plot effects

The spatial method denotes here a procedure whereby the model for plot effects is selected from a set of candidate models. We restricted ourselves to seven models:

- In the null model, ξ_i is an error term assumed to be independent and normally distributed with zero mean and variance σ_ξ^2.
- In the random block model, ξ_i is equal to $\beta_b(i) + e_i$, where $\beta_b(i)$ is the random effect of the block b with zero mean and variance σ_β^2 and e_i is a random error term, assumed to be independent, normally distributed with zero mean and variance σ_e^2.
- in the random column model, ξ_i is equal to $c_{x(i)} + e'_i$ where $c_{x(i)}$ is the random effect of the column x with zero mean and variance σ_c^2. This effect corresponds to the longest direction of plots and can be relevant because it is also the direction of sowing.
- The fourth model is the row and column random model with ξ_i equal to $c_{x(i)} + r_{y(i)} + e''_i$ where $r_{y(i)}$ is the random effect of the row y with zero mean and variance σ_r^2.
- The last three models are neighbour models as described by Gleeson and Cullis (1987). Plot effects are assumed to be a realization of a random autoregressive integrated moving average (ARIMA) process. Following Cullis and Gleeson (1991) we applied these models in two directions assuming separability of the random process in the row and column directions. Three models were considered: AR1 × AR1 (first-order autoregressive), AR2 × AR2 (second-order autoregressive) and ARMA × ARMA (first-order autoregressive moving average) models.

For choosing the best model among the seven possible ones for a given trial, we used the Akaike's Information Criterion, AIC (Akaike 1973):

$$-2\,lod + 2p,$$

where *lod* is the loglikelihood, and p is the number of variance parameters in the model. Thus, the spatial method consists of choosing the model with the smallest AIC. It must be noted that with the REML estimation method, the maximization is made on the random terms, and comparisons of AICs are only possible for models which have the same fixed effects. Thus, this criterion can be used neither to select simultaneously the best trend model and the markers to be included in the genetic model nor to select markers once the trend model is chosen.

With no marker in the model, the difference between the loglikelihoods of two nested models follows a χ^2 with q degrees of freedom, where q is the difference between the number of parameters

estimated in each model. We used this test (further noted LR test) to compare the significance of a given trend model relative to the 'no trend' one.

Selection of markers

Once the trend model was chosen, we used it during the marker selection process. Since the programme we used (1) did not allow an automatic stepwise approach and (2) could not take into account all the covariates simultaneously, we used the following backward procedure. For each trial, classical stepwise regressions on marker scores were first performed on the phenotype means and on the BLUPs of the genetic values obtained with each of the seven trend models. A subset of preselected markers was made of all markers which were retained at the 5% significant level by at least one of the eight stepwise regressions. This subset was then introduced in the complete model (2). The marker effects were tested with a Wald test and, step by step, the less significant marker was removed from the model until all the remaining markers were significant at the 5% level. This significance level of 5% was chosen according to conclusions of Moreau et al. (1998).

Comparison of the efficiencies of the methods for modelling plot effects

The spatial method was compared with two simpler methods of modelling local trend to predict P, M and C genetic values. These two methods assumed the same model on plot effects for all the trials, respectively: (1) the null model, and (2) the random block effect model. Thus, for P, M and C, three methods of modelling plot effects will be compared in the results: 'no trend', 'block' and 'spatial'.

It is well-known that model selection can lead to biased estimation of the precision of a given model if the same data are used in the estimation and validation steps (see, for example, Hjorth 1994). To avoid this bias, Hill and Rosenberger (1985) did a cross-validation based on the error of prediction of observed values in one trial when the other trials were used for the prediction. We also did such a cross-validation (results not shown in this paper), but we prefer to present another approach based on correlation between trials, as proposed by Clarke et al. (1994). These authors have shown that the expected correlation between performances (adjusted or not by a trend model) of two trials t and t' is, assuming the same error variance in both trials,

$$\rho(t,t') = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gl}^2 + \sigma_e^2}$$

where σ_g^2 is the genetic variance common to the trials, σ_{gl}^2 is the genotype × environment interaction variance and σ_e^2 the error variance.

This criterion provides an estimation of the efficiency of the model since, if this model is well-adapted to a given trial, it must reduce the error variance for this trial and then increase correlation with other trials. It can also be noted that the square root of $\rho(t,t')$ is proportional to the genetic gain that can be expected in one site when genotypes are selected from the performances in another site. In fact, $\rho(t,t')$ is the associated heritability for such a selection. This criterion of validation is based on predicted values obtained independently in two different trials. As genotypes were randomized among the plots independently in each trial, errors and spatial heterogeneities are independent in each trial. Thus, correlations between trials are not biased by model selection.

To compare the three methods of modelling local trend, we used one global criterion based on such correlations: the empirical average correlation between estimated genetic values over all the different pairs of trials (i.e. $\bar{\rho}(t,t') = 1/55 \sum_{t=1}^{11} \sum_{t>t'}^{11} \hat{\rho}(t,t')$). Nevertheless, the interest in modelling local trend depends on the existence of a spatial heterogeneity in the trial. Therefore, we also calculated the average correlation between each given trial, analysed using the 'no trend', 'block' or 'spatial' methods, and the other trials, analysed with the 'spatial' method. In this way, differences in average correlation are only due to the change in method of modelling plot effects for this trial; the other trials always being analysed with the 'spatial' method. This provides a second criterion by which to compare the methods.

Selection of the best genotypes can be based on the average of the predicted genetic values over all the trials. In this context, the best method of modelling plot effects should provide a given accuracy with fewer trials. To evaluate this, it is necessary to compare the average predicted values over a given number of trials with an appropriate reference. The average yield performance of genotype over all the trials is a simple reference which does not spuriously favour the proposed method of modelling plot or genetic effects. Consequently, we calculated the correlations between this reference and the average predicted values obtained with the C method on subsets of l trials analysed with the 'no trend', 'block' or 'spatial' method.

We limited all these analyses of correlations to the set of 243 genotypes that were common to the 11 trials.

Results

Identification of the 'best' trend model for each trial

For each trial, the model identified by the 'spatial' method is indicated in Table 2. The 'block' model was never found to be the best. The 'row × column' or 'AR2 × AR2' models were chosen in all but 3 trials. Table 2 also indicates the AIC differences between 'block' and 'no trend' and between 'spatial' and 'no trend'. For trials 7, 8 and 11, all the AIC differences are low; LR tests show that 'block' and 'spatial' are not significantly better than the 'no trend' method at the 0.1% level. In these cases, one cannot expect that taking account of spatial heterogeneity will greatly change the predicted genetic values. In trials 5 and 9 the

Table 2 "Spatial" model chosen for each trial and comparison of the AIC of models "block" or "spatial" with "no trend"

Trials	"Spatial" model	AIC of a given model minus AIC of the "no trend" model[a]	
		"Block"	"Spatial"
1	AR2 × AR2	− 85***	− 98***
4	Row × column	− 84***	− 118***
5	Column	+ 2	− 34***
6	AR2 × AR2	− 104***	− 106***
7	AR1 × AR1	+ 2	− 8**
8	Row × column	+ 2	− 8**
9	Row × column	− 9***	− 17***
10	AR2 × AR2	− 27***	− 206***
11	AR2 × AR2	− 4*	− 4*
12	AR2 × AR2	− 64***	− 94***
14	ARMA × ARMA	− 8**	− 118***

[a] The level of significance of the loglikelihood ratio test of comparison between the "block" or "spatial" and "no trend" model is indicated: * = 5%, ** = 1%; *** = 0.1%

differences are significant but not very great. In the other trials (1, 4, 6, 10, 12 and 14), the differences are much greater, so including a spatial effect in the model should modify the predicted values. Note that in most cases when the differences in AIC values are important, a spatial heterogeneity was clearly observed within the field trial during field visits. The differences between 'block' and 'spatial' AICs also vary between trials. For trials 6 and 9, 'block' and 'spatial' AICs are close and both are significantly better than the 'no trend' model. On the other hand, for trials 5 and 14, 'block' and 'no trend' AICs are close and 'spatial' appears to be much better than the others. For trial 10, even if 'block' is already significantly better than 'no trend', the AIC of 'spatial' indicates that this model is much better.

Influence of the trend model on the prediction of the genetic values

The correlations between the genetic values predicted with the three methods for modelling plot effects are presented in Table 3 for the P, M and C predicted values. The correlations are consistent with the AIC differences: models with AIC values close to each other yield highly correlated genetic values (trials, 7, 8 and 11). When AIC values differ, the correlations are often smaller for the markers only (M) prediction method than for the phenotypic (P) method (trials 1, 4, 10 or 14). Therefore, markers seem to accentuate the differences between methods for modelling plot effects. However, in trial 6 (and trial 12 to a lesser extent), the correlation between methods is increased.

The number of markers selected in the genetic model is similar from one method to the other, even if slightly more markers are selected with the 'spatial' method (10.5 on average) than with the 'no trend' (9.8) or 'block' (9.5) methods. The different methods do not select the same subset of markers, but differences are often slight and selected markers are generally located in the same chromosomal areas. The percentage of phenotypic variance associated with markers can be expressed by the ratio of the variance of the observed $\hat{M}_j$ on the phenotypic variance (without any adjustment by plot effects). One can expect that this percentage increases when the model controls the trend variability better. In fact, this was not observed on average of the 11 trials (it varied from 0.26 for 'no trend' and 'block' to 0.24 for 'spatial'). Differences in percentage of phenotypic variance explained by the markers, estimated using this approach, may not be a good criterion to compare the methods. It is known that the effects associated with markers are generally overestimated (as mentioned by Lande and Thompson 1990). As these effects are expected to be better estimated using the 'spatial' method, they may be overestimated to a lesser extent. This may explain why the percentage of phenotypic variance does not increase when the 'spatial' method is used.

Comparison of the efficiency of the three methods for modelling plot effects

The average correlation between all the trials for the different methods is given in Table 4. Generally speaking, the correlations between trials are low, especially for P genetic values. It should be noted first, that we considered unreplicated trials, the accuracy of which is low, and second, that the 11 trials were grown in quite

Table 4 Average correlation between predicted genetic values (P, M or C) of all pairs of trials for the three methods of modelling plot effect

Genetic value	"No trend"	"Block"	"Spatial"
P	0.240	0.286	0.319
M	0.400	0.428	0.464
C	0.391	0.411	0.428

Table 3 Correlations between genetic values predicted with the "no trend", "block" and "spatial" methods of modelling plot effects

Trial	P			M			C		
	"No trend"-"block"	"No trend"-"spatial"	"Block"-"spatial"	"No trend"-"block"	"No trend"-"spatial"	"Block"-"spatial"	"No trend"-"block"	"No trend"-"spatial"	"Block"-"spatial"
1	0.87	0.80	0.89	0.79	0.68	0.85	0.79	0.73	0.82
4	0.87	0.80	0.92	0.76	0.62	0.69	0.76	0.62	0.69
5	1	0.94	0.94	0.99	0.85	0.87	0.99	0.85	0.87
6	0.85	0.76	0.91	0.92	0.93	0.94	0.90	0.83	0.92
7	1	0.97	0.96	1	0.98	0.98	1	0.98	0.98
8	1	0.98	0.98	0.98	0.99	0.97	0.99	0.99	0.98
9	0.98	0.96	0.98	0.95	0.88	0.91	0.96	0.92	0.96
10	0.95	0.71	0.76	0.66	0.66	0.82	0.77	0.72	0.80
11	0.98	0.97	0.98	0.97	0.97	1	0.98	0.97	0.98
12	0.89	0.79	0.89	0.88	0.89	0.96	0.90	0.77	0.84
14	0.98	0.80	0.81	0.94	0.68	0.71	0.96	0.75	0.76

different conditions, some of them being irrigated and others not, which generated strong genotype × environment interactions. Limiting the study to more homogeneous groups of trials (irrigated or not) increases the correlations between pairs of trials (result not presented).

For P, M and C, the average correlation between two trials increases when a trend effect is added to the model. 'Spatial' gives the best average correlations and thus appears to be more efficient than the other two methods. The differences between the 'no trend' and 'spatial' methods are +0.079 for P, +0.064 for M and +0.037 for C, which give, respectively, an average relative increase of the correlations of about 33%, 16% and 9.5%. Modelling field trial heterogeneity seems to be more efficient to increase the correlations when there is no marker in the full model. This result is not surprising because markers increase the accuracy of predicted breeding values (that is the reason why MAS is more efficient than phenotypic selection). With markers, the correlations are already high with the 'no trend' method (the average correlation is 0.391 for C but only 0.24 for P), and so they are less easily improved. The gain due to the 'spatial' method for the C method is then not negligible.

As the 'spatial' method appears to be the best, we compared the average correlation between a given trial analysed with 'no trend', 'block' or 'spatial' and each of the other trials analysed with the 'spatial' method, the latter considered as a reference. Important discrepancies appear between trials (cf. Table 5 and Fig. 1 which is the graphical representation of Table 5). As expected from the AIC values, only slight differences were observed between the correlations of the different methods for trials 7, 8 and 11. The largest differences are for trials 1, 4 and 10, for which the model chosen with the 'spatial' method has an AIC much lower than the others. The increase in correlation between 'spatial' and 'no trend' is equal to +0.092, +0.085, +0.076 for P, M and C in trial 10, which corresponds to relative increases of 56%, 29% and 28%, respectively. For trial 4, the increases are +0.083, +0.106 and +0.09 for P, M and C, which correspond to relative increases of +34%, +27% and +24%. For trial 1, the increases are slightly lower, respectively +0.062, +0.054 and +0.066, which give relative increases of 30%, 15% and 21%. One can observe that for trials 4 and 10, the 'block' method does not give such increases. For trials 6, 9, 12 and 14, modelling spatial heterogeneity ('block' or 'spatial' methods) improves the correlations. Nevertheless, for those trials, it should be noted that the 'spatial' method does not seem to be better than 'block' for the C value. For trials 6 and 9 the AICs of 'block' and 'spatial' were not really different, which could explain this result. For trials 12 and 14, this result is more surprising, since the AIC of 'spatial' was much lower than the AIC of 'block'. For trial 5 the model chosen with the 'spatial' method is much better with respect to the AIC criterion, but in fact it does not give better correlations than 'block' or 'no trend'. Generally speaking, the increase in correlation obtained by using the 'spatial' method is all the greater as the trial was less accurate (less correlated with the others when the 'no trend' method is used): the difference between the correlations obtained by the 'spatial' and the 'no trend' methods is significantly and negatively correlated with the correlations obtained by the 'no trend' method (−0.91 for P, −0.67 for M and −0.72 for C). This is also illustrated by Fig. 1.

As previously mentioned, the efficiency of a model for MAS can also be evaluated through the correlation between the average of the values estimated by the C method on a subset of l trials and the average yield performance on all the trials, the latter taken as reference. This reference provides a rather accurate predicted genetic value since the estimated heritability of

Table 5 Average correlation between predicted genetic values of a given trial analysed with the "no trend" or "block" or "spatial" method and the other trials analysed with the "spatial" method

Trials	P			M			C		
	"No trend'	"Block"	"Spatial"	"No trend"	"Block"	"Spatial"	"No trend"	"Block"	"Spatial"
1	0.209	0.251	0.271	0.357	0.388	0.411	0.322	0.351	0.388
4	0.241	0.301	0.324	0.400	0.396	0.506	0.378	0.381	0.468
5	0.384	0.384	0.370	0.517	0.521	0.499	0.484	0.491	0.494
6	0.222	0.293	0.320	0.541	0.554	0.573	0.496	0.498	0.478
7	0.344	0.344	0.322	0.447	0.447	0.444	0.409	0.411	0.409
8	0.349	0.350	0.353	0.492	0.493	0.499	0.471	0.469	0.471
9	0.345	0.362	0.368	0.515	0.521	0.518	0.487	0.488	0.479
10	0.163	0.185	0.255	0.294	0.289	0.379	0.269	0.280	0.345
11	0.336	0.344	0.343	0.426	0.435	0.435	0.415	0.422	0.422
12	0.229	0.260	0.279	0.390	0.433	0.436	0.369	0.409	0.353
14	0.249	0.264	0.306	0.380	0.428	0.402	0.373	0.408	0.403
Average	0.279	0.303	0.319	0.433	0.446	0.464	0.407	0.419	0.428

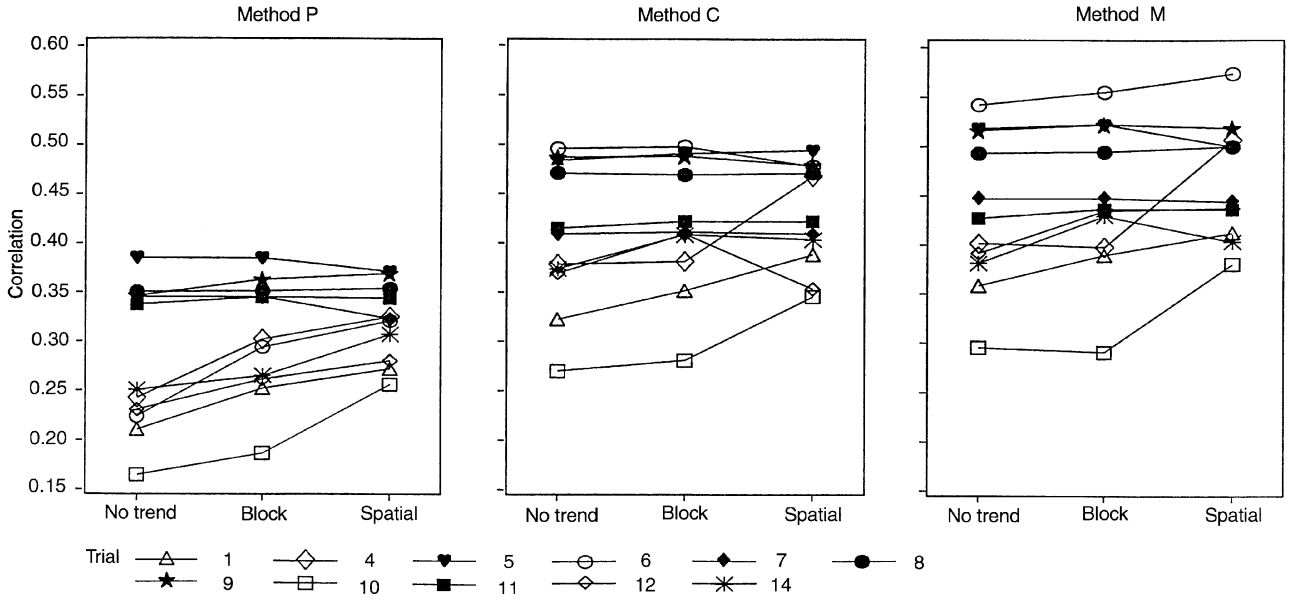

Fig. 1 Average correlations between one trial analysed with 'no trend', 'block' or 'spatial' method and the others analysed with 'spatial' method for the three methods of selection considered: *P* Phenotype; *M* Markers; *C* Combined, phenotype + markers. The *solid* symbols correspond to the trials (5, 7, 8, 9, 11) that have the highest average correlations when no spatial model is assumed ('no trend') and no marker included in the model (*P* method). The other trials are represented with *open* symbols

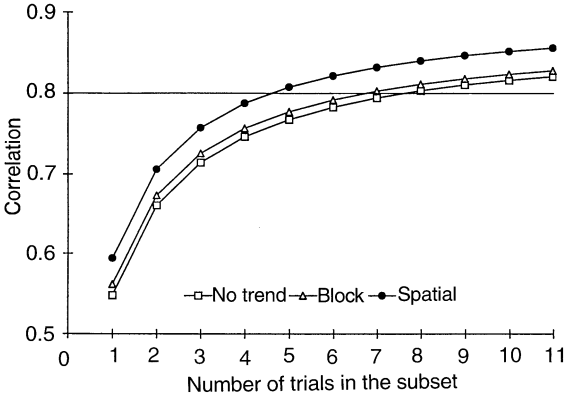

Fig. 2 Effect of the number of trials on the accuracy of the average genetic values predicted with the combined method according to the spatial method considered. Each point of this graph represents the average correlation over all the subsets of a given size between the mean rough performances over the 11 trials, taken as a reference, and the average predicted genetic values in the subset using the combined method

the average yield performance is 0.7. The change in the correlation with this reference when the size of the subset varies from 1 to 11 is represented in Fig. 2. Each point corresponds to the average of the correlations obtained from all the different subsets of l trials that can be drawn from the 11 trials. According to Fig. 2, one can observe that, whatever the number of trials considered, the 'spatial' method always yields the highest correlation and the 'no trend' method the worst. This is an additional argument in favour of the 'spatial' method. One can also observe that the correlation of 0.8 which is achieved with 5 trials for the 'spatial' method is only achieved with 7 trials for the 'block' method and 8 trials for the 'no trend' method. From an economic point of view this provides a strong argument in favour of the 'spatial' method.

Discussion

In the method proposed the genetic values are divided into two parts: one which can be explained through marker effects and the other unexplained by markers. Lande and Thompson (1990) proposed to select the markers and estimate the molecular score in a first step and then to combine the estimated molecular score and the performance in a linear index. Weights given to each term take into account the heritability and the percentage of variance explained by markers. In our approach, both parts (explained and unexplained by markers) are estimated in the same step. When spatial heterogeneity is not taken into account, the two approaches are asymptotically equivalent. However our method allows one to consider spatial heterogeneity and is more adapted to unreplicated trials, since in this case, it is difficult to estimate the heritability and then to determine the weights that should be given to the molecular score and to the phenotypic performances. This is of particular interest if the aim is to predict the genetic value of individuals in each trial. If the main interest is to select for general adaptation, then trials

can be seen as replications, and Lande and Thompson's approach can be applied on the average performances. The relative efficiency of our method compared to this one should be all the higher as the number of trials is low and the importance of spatial heterogeneity is high.

Concerning the choice of the markers to be included in the model, several methods have been investigated in the literature. Zhang and Smith (1992, 1993) and Gimelfarb and Lande (1994, 1995) included in their simulations a fixed number of markers selected in one or two stages (first on each chromosome and then among the pre-selected markers). Recently, Whittaker et al. (1995), Moreau et al. (1998) and Hospital et al. (1997) showed that it is more efficient to adapt the number of markers to the experimental power of detection. Whittaker et al. (1995) advised using the adjusted Mallow's Cp and performing a two-stage regression. Hospital et al. (1997) showed that a stepwise regression based on type-I error risk also gives good results that are not much different from those obtained with Mallow's Cp (F. Hospital, personal communication). Moreover, Mallow's Cp can only be carried out in a fixed model context. As there is no satisfying statistical tool for mixed models, we based our marker selection on the type-I error risk, testing marker effects with a Wald test.

Concerning the choice of the trend model, the AIC criterion may not be the best-adapted one, but we used it partly for the sake of simplicity. Our results suggest that in most cases the plot effect model identified as the best provides more accurate genetic values. Another criterion using the replicated genotypes to control whether the trend model correctly adjusts the spatial heterogeneity is under study. One tool advocated by Cullis and Gleeson (1991) is based on the observation of the spatial correlations matrix of the whitened residuals. Cullis et al. (1998) and Gilmour and Cullis (1997) gave additional advice, including visual inspection of two-dimensional variograms.

A further improvement of the genetic value prediction process would be to analyse simultaneously all the trials, as proposed by Cullis et al. (1998), and to adapt this method to the specific case of MAS. The information from the different trials could improve the choice of the trend model for each trial and also the power of QTL detection and then the choice of the markers to be included in the model. This deserves further investigation.

In spite of the empirical way in which models were chosen, the results are encouraging since the use of the 'spatial' method increases the correlations between trials. Using the 'spatial' method instead of the 'no trend' one increases the mean correlation between pairs of trials from 0.24 to 0.319. Including markers when using the 'spatial' method further increases the correlation to 0.428. In terms of selection, on the average of 11 trials, the relative efficiency of the C 'spatial' method compared to the P 'no trend' method can be defined as: $(\sqrt{\bar{\rho}_{C\ 'spatial'}} - \sqrt{\bar{\rho}_{P\ 'no\ trend'}})/\sqrt{\bar{\rho}_{P\ 'no\ trend'}}$, where $\bar{\rho}_{C\ 'spatial'}$ and $\bar{\rho}_{P\ 'no\ trend'}$ are the average correlations between trials obtained with the C 'spatial' method and the P 'no trend' method, respectively. This relative efficiency is close to 34%. If the 'spatial' model is not used, the inclusion of markers leads to a relative efficiency of 28%.

When several trials are considered, the interest of using the 'spatial' method decreases as the number of trials increases, as shown by the correlation of 0.97 between the average genetic values on 11 trials estimated with the 'no trend' C and the 'spatial' C methods. However, the main advantage of using the 'spatial' model in MAS is that increasing the accuracy of breeding values allows one to use fewer trials to reach the same genetic gain. Five trials using the 'spatial' C method appear to be as efficient as 8 trials using the 'no trend' C method. Moreover, it provides a more accurate genetic prediction in each trial, allowing one to select genotypes for their adaptation to specific conditions.

It should be noted that the M method leads to higher correlations between trials than C. However, the correlations between the average genetic values estimated with M and P methods are only about 0.7, whereas the correlations between P and C vary between 0.82 and 0.9 depending on the trend method considered. Thus, even if the genetic values estimated from markers are highly consistent from one trial to another, it seems that markers do not explain all the genetic variability. This can be related to other experimental results (Stuber et al. 1992, Koester et al. 1993) which suggest that the QTLs detected are those whose effect is high and less sensitive to environmental conditions than the others.

As previously announced in the Introduction, we do not provide a theoretical demonstration of the efficiency of this approach. Nevertheless, our results show that it can be efficient to take into account the field trial heterogeneity, even in the context of MAS, where the accuracy is already improved by the use of markers. This seems to be especially interesting when genotypes are grown under stressed conditions. When genotypes are grown in only a few trials, using a 'spatial' method reduces the risk of achieving poor genetic gain if a spatial heterogeneity occurs within the trials. Even if it was not formally investigated here, taking into account spatial heterogeneity should also be important in QTL detection experiments since it improves the heritability of the trait.

Acknowledgements We are grateful to all the people who participated in the experimental work at INRA and more specifically to Catherine Giauffret, Cyril Bauland, Xavier Charrier, Christine Epinat, Philippe Jamin, Daniel Jolivot, Christine Lariagon, Valerie Combes, Fabrice Dumas and Marielle Merlino. We thank Bruno Goffinet and Caude Chevalet for their useful copmments on early steps of this work. This study was supported by grants from the French Ministère de l'Agriculture et de la Pêche, and the PROMAIS

association members: Advanta, Cargill Genetique Europe, Caussade Semences, Novartis Semences – Ciba Geigy, Limagrain Genetics, Maisadour, Nordsaat France, Pioneer Genetique, Pau Semences, R.A.G.T., Rustica Prograin Genetique et S.D.M.E./KWS France. We are grateful to these companies for expert field evaluation and to their scientists for helpful discussions.

References

Akaike, H. (1973) Information theory and an extension of the maximum principle. In: Petrov BN, Czaki F (eds) Proc 2nd Int. Symp. Information Theory. Akademia Kiado, Budapest, pp 267–281

Causse M, Santoni S, Damerval C, Maurice A, Charcosset A, Deatrick J, de Vienne D (1996) A composite map of expressed sequences in maize. Genome 39 : 418–432

Clarke FR, Baker RJ, DePauw RM (1994) Moving mean and least squares smoothing for analysis of grain yield data. Crop Sci 34 : 1479–1483

Cullis BR, Gleeson AC (1991) Spatial analysis of field experiments – an extension to two dimensions. Biometrics 47 : 1449–1460

Cullis BR, Lill WJ, Fisher JA, Read BJ, Gleeson AC (1989) A new procedure for the analysis of early generation variety trials. Appl. Stat. 38 : 361–375

Cullis BR, Gleeson AC, Thomson FM (1992) The response to selection of different procedures for the analysis of early generation variety trials. J Agric Sci 118 : 141–148.

Cullis BR, Gogel BJ, Verbyla A, Thompson R (1998) Spatial analysis of multi-environment early generation variety trials. Biometrics 54 : 1–19

Gilmour A (1992) TWOD: a program to fit a mixed linear model with two dimensional spatial adjustment for local trend. NSW Agriculture, Orange NSW 2800, Australia

Gilmour A, Cullis BR (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2 : 1–25

Gilmour A, Welham S, Harding S (1995) Efficient analysis of field experiments using two-dimensional spatial models. Genstat Newsl 32 : 31–39

Gimelfarb A, Lande R (1994) Simulation of marker-assisted selection in hybrid populations. Genet Res 63 : 39–47

Gimelfarb A, Lande R (1995) Marker-assisted selection and marker-QTL associations in hybrid population. Theor Appl Genet 91 : 522–528

Gleeson AC, Cullis BR (1987) Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. Biometrics 43 : 277–288

Hill RR, Rosenberger JL (1992) Methods for combining data from germplasm evaluation trials. Crop Sci 25 : 467–470

Hjorth U (1994) Computer intensive statistical methods. Validation model selection and bootstrap. Chapman & Hall, New York

Hospital F, Moreau L, Charcosset A, Gallais A (1997) More on the efficiency of marker assisted selection. Theor Appl Genet 95 : 1181–1189

Kempton RA (1984) The design and analysis of unreplicated field trials. Vortr Pfanzenzuecht 7 : 219–242

Kempton RA, Gleeson AC (1997) Unreplicated trials. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman & Hall, London, pp 86–100

Knapp SJ, Bridges WC (1990) Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. Genetics 126 : 769–777

Koester RP, Sisco PH, Stuber CW (1993) Identification of Quantitative Trait Loci controlling days of flowering and plant height in two near isogenic lines of maize. Crop Sci 33 : 1209–1216

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124 : 743–756

Luo ZW, Thompson R, Woolliams JA (1997) A population genetics model of marker-assisted selection. Genetics 146 : 1173–1183

Martinez O, Curnow RN (1994) Missing markers when estimating quantitative trait loci using regression mapping. Heredity 73 : 198–206

Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-assisted selection efficiency in populations of finite size. Genetics 148 : 1353–1365

Stroup WW, Mulitze DK (1991) Nearest neighbor-adjusted best linear unbiased prediction. Am Stat 45 : 194–200

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander S (1992) Identification of genetic factors contributing to heterosis in a hybrid from elite maize inbred lines using molecular markers. Genetics 132 : 823–839

Weber WE (1980) Quantitativ genetische Untersuchungsmethoden und ihre Bedeutung für die Selecktion in der Selbstberfruchterzüchtung. Habilitations-schrift, University of Hanover

Whittaker JC, Curnow RN, Haley CS, Thompson R (1995) Using marker-maps in maker-assisted selection. Genet Res 66 : 255–265

Zhang W, Smith C (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor Appl Genet 83 : 813–820

Zhang W, Smith C (1993) Simulation of marker-assisted selection utilizing linkage disequilibrium: the effect of several additional factors. Theor Appl Genet 86 : 492–496