

# SIMULATION OF MIXED MODELS IN AUGMENTED BLOCK DESIGN<sup>1</sup>

Aladir Horacio Santos<sup>2\*</sup>; Eduardo Bearzoti<sup>3</sup>; Daniel Furtado Ferreira<sup>3</sup>; João Luís da Silva Filho<sup>4</sup>

<sup>2</sup>Depto. de Ciências Exatas - Faculdade de Filosofia - FUOM, C.P. 102 - CEP: 37570-000 - Formiga, MG.

<sup>3</sup>Depto. de Ciências Exatas - UFLA, C.P. 37 - CEP: 37200-000 - Lavras, MG.

<sup>4</sup>Depto. de Biologia - UFLA, C.P. 37 - CEP: 37200-000 - Lavras, MG.

\*Corresponding author <aladir@stratus.com.br>

**ABSTRACT:** The augmented block design is widely used in breeding programs, with non-replicated treatments generally being selection units, and replicated treatments being standard cultivars. Originally, an intrablock analysis (fixed model) was proposed. Although non-replicated treatments and/or blocks can be considered of random nature, mixed linear models could be used instead. This work evaluated such an approach, using computer simulation. Populations consisted of sets of randomly generated inbred lines. Molecular marker data were also simulated to allow the estimation of the genetic covariance matrix. Different conditions were considered, varying heritability and the coefficient *b* of Smith of soil heterogeneity. For each condition 100 simulations were performed, considering four linear models, varying respectively the nature of the effects of blocks and non-replicated treatments (fixed - F, or random - R): FF, FR, RF and RR. In relation to FF, the mixed models were more efficient under low to intermediate heritability and high *b*. Mixed models could improve inference in breeding programs using the augmented block design and the choice of the model should rely on the kind of selection. If this is truncated, the RF model should be preferred; if it is not, then the RR model would be more suitable.

Key words: augmented design, mixed model, molecular marker

## SIMULAÇÃO DE MODELOS MISTOS NO DELINEAMENTO EM BLOCOS AUMENTADO

**RESUMO:** O delineamento em blocos aumentados é amplamente utilizado em programas de melhoramento, geralmente com tratamentos não-repetidos correspondentes a unidades de seleção, e tratamentos repetidos sendo cultivares comerciais. Originalmente, uma análise intrabloco (modelo fixo) foi sugerida. No entanto, se os tratamentos não-repetidos e/ou blocos puderem ser considerados de natureza aleatória, modelos lineares mistos poderiam ser utilizados. Este estudo objetivou a avaliação de tal abordagem, utilizando simulação. Linhagens endogâmicas foram geradas aleatoriamente, bem como dados de marcação molecular, para estimar a matriz de covariâncias genéticas. Variaram-se a herdabilidade e o coeficiente *b* de heterogeneidade de solo de Smith; em cada condição, 100 simulações foram feitas, considerando 4 modelos lineares, variando respectivamente a natureza dos efeitos de bloco e de tratamentos não-repetidos (fixo - F, ou aleatório - A): FF, FA, AF e AA. Em relação ao modelo FF, os modelos mistos foram mais eficientes especialmente sob herdabilidade baixa a intermediária e alto *b*. Modelos mistos podem melhorar a inferência em programas de melhoramento utilizando o delineamento, e a escolha do modelo deve se basear no tipo de seleção. Se esta for truncada, o modelo AF deveria ser preferido; se não for, então o modelo AA é mais apropriado.

Palavras-chave: delineamento aumentado, modelo misto, marcador molecular

### INTRODUCTION

The augmented designs were proposed by Federer (1956) as an alternative under low availability of genetic material for replication, and many treatments. They are a modification of straightforward designs, by adding treatments that appear only once in the experiment (herein designated simply as "treatments") to the set of replicated treatments ("checks"). Among the augmented designs, the augmented block design is perhaps the most used, and inference has traditionally been made by means of an intrablock analysis (considering treatment and block effects fixed).

In plant breeding, this approach may be unsuitable, for in many cases such design is used with

treatments being selection units sampled from a population. Besides, recovery of information on treatments among blocks (interblock approach) can potentially improve the estimates, and this is achieved by regarding block effects as random. Genetic variances can be underestimated if the intrablock analysis is used in the augmented block design (Bearzoti, 1994). Therefore, the theory of mixed models (Henderson, 1984) could be used to take into account the randomness of the effects of treatments and/or blocks (Robinson, 1991). The effects of checks could be considered fixed in plant breeding, for they are generally standard released varieties.

The computational improvement that the past decade witnessed made the analysis of augmented

<sup>1</sup>Part of the MS Thesis of the first author, presented to UFLA - Lavras, MG.

designs straightforward. Scott & Milliken (1993) show an example of a SAS routine that yields the analysis of variance, hypothesis tests, intra and interblock estimation of treatment effects in augmented block designs. Boyle & Montgomery (1996) pointed out how intra and interblock approaches could be dealt using GLM and MIXED procedures of SAS software. Wolfinger et al. (1997), using such procedures, show routines to recover interblock and intervariety information when the nature of block and variety effects are actually random. Duarte (2000) presents a detailed study on the augmented block design, highlighting statistical issues and its use in plant breeding, showing many SAS routines to different alternatives of analysis of such designs, including the modeling of spatial residual dependence.

If treatment effects are regarded as random, e.g. if they are sampled genetic materials, and if intergenotypic information is recovered, the predicted treatment means tend to show a decreased dispersion around the overall mean, in comparison to the dispersion of intrablock means. This trend is known as "shrinkage" effect, and reflects a more accurate distribution of genetic effects, if the assumption of randomness is correct (see, for instance, Duarte, 2000). The accuracy on estimating such distribution could be further improved if the matrix of covariances among genetic effects was known. In plant breeding, if treatments are genetic materials, such matrix could be constructed (apart, perhaps, from a constant corresponding to the genetic variance) using relationship information from molecular data, which can account for genetic similarities among treatments. This can be considered a simple way of marker-assisted selection, and relatively low-cost and fast generating markers like RAPD (random amplified polymorphic DNA) could be used.

There are reports in the literature on the use of a mixed model approach to analyze data of augmented block designs (Duarte, 2000; Federer, 1998), but its efficiency when covariances among treatments are not null and can be estimated has not yet been studied. Computer simulation is a powerful tool in such studies, for a wide range of conditions could be covered, in situations where an analytical approach has limitations.

This work aimed at the evaluation of the use of mixed model theory in augmented block designs, when treatments are selection units in a breeding program, and genetic covariances are estimated using molecular data.

## MATERIAL AND METHODS

Augmented block designs were simulated using computational routines written with Delphi language, version 3. Treatments consisted of inbred lines of a fictitious diploid species with 200 independent genes controlling the trait of interest, and 100 marker loci were used to estimate the coefficients of genetic covariances among lines.

The frequency of the favorable allele in the population of lines of each gene was generated using a uniform distribution with parameters 0.2 and 0.8. A given line had a favorable allele with probability equal to the frequency of the corresponding gene. Genotypes regarding marker loci were simulated using the assumption that, for a given pair of lines, the number of marker loci with the same genotype had a binomial distribution with parameters  $r$  and  $s$ , where  $r$  was the number of markers (100) and  $s$  the genetic similarity in that pair of lines, considering the genes of the hypothetical trait. This assumption reflected basically the idea that similarities based on marker and trait loci are related.

The number  $n$  of lines were 50, 100 and 200. In each simulated condition, a different set of lines was used. The effects of each gene were generated as outcomes from an exponential distribution with parameter equal to 1, multiplied by -1 if the allele were not that favorable. Genotypic value of a given inbred line was calculated as the sum of the effects of all genes of that particular line. The variance among such genotypic values was the genetic variance across lines. Environmental variances were then obtained in such a way to yield heritability values (at the level of experimental plots) equal to 0.2, 0.5 and 0.8. This total environmental variance was partitioned into two components, associated to the variation within and among blocks, with magnitudes determined according to the value of the coefficient of soil heterogeneity of Smith (1938), represented by  $b$ , as suggested by Bearzoti (1997). Three values of  $b$  were evaluated: 0.1, 0.5 and 0.9.

The number of blocks was dependent on the number of lines, the former being 5% or 20% of the latter. Therefore, considering the 3 sample sizes (number of lines), the 3 values of the coefficient  $b$ , 2 amounts of blocks, and the 3 heritability values, 54 situations were analyzed. Each situation was simulated 100 times.

Every time a data set was generated, the following models were considered: 1) fixed effects of blocks and treatments (FF), corresponding to the intrablock analysis; 2) random block effects and fixed treatment effects (RF), corresponding to the analysis with recovery of interblock information; 3) fixed block effects and random treatment effects (FR), corresponding to the analysis with recovery of intergenotypic information; and 4) random effects of blocks and treatments (RR), recovering both interblock and intergenotypic information. The effects of the checks were always taken as fixed.

The alternatives RF, FR and RR had two variants, considering that the variance components were known, designated by RF(k), FR(k) and RR(k), and considering that they were not known. This allowed the accounting for the loss in precision due to the necessity of estimation of such components. In those situations where treatment effects were fixed, a parametric constraint was used, considering the sum of such effects equal to zero.

In order to compare estimates (or predictions) of treatment effects,  $\hat{\theta}_i$  to the corresponding genotypic values  $\theta_i$  ( $i = 1, 2, \dots, n$ ), the following statistics were calculated: mean square error (MSE); Spearman correlation, that reflects the accuracy on the ranking of lines, which is specially useful when selection is truncated; and “elite bias”, defined as the bias in the estimation of the percentage of treatments superior to the best check. The magnitude of such bias is relevant when selection is based in relation to standard cultivated varieties (non-truncated selection).

When variance components were considered unknown, they were estimated using restricted maximum likelihood (REML) and the EM algorithm (Dempster et al., 1977) associated to the mixed model equations of Henderson (1984). Convergence was considered to have been reached if, in a given iteration, the ratio(s) of the variance component of the factor (block or treatment) and residual variance component were different from that of the previous iteration in a magnitude no greater (in absolute value) than  $10^{-8}$ .

## RESULTS AND DISCUSSION

The MSE values of the estimated genetic effects, for all 54 configurations, are presented in Table 1. The RR(k) approach presented the smallest MSE values in 34 situations (53.1%). Most of the remaining 20 configurations were conditions of high heritability ( $h^2 = 0.8$ ). Among all situations with  $h^2 = 0.8$ , only one presented the lowest MSE with RR(k) model, corresponding to  $n = 50$  lines,  $b = 0.9$  and number of blocks equal to 2. Three out of the 20 remaining configurations had  $h^2 = 0.5$ ,  $b = 0.1$  and number of blocks equal to  $0.2n$ . In such cases, however, the most precise approach was RF(k), which also assumes known variance components.

For the 17 situations with  $h^2 = 0.8$  and RR(k) approach not being the most precise, it can be seen that the MSE values of this model were more distant from the lowest MSE with  $b = 0.1$ , that is, when a higher fraction of environmental variance is due to the differences among blocks. Comparing, for example, situations 43 and 47 in Table 1, which have similar conditions, except for the value of  $b$  (0.1 and 0.9, respectively), in the former the RR(k) model had an MSE about six times the smallest, while in the latter this MSE was only 1.18 times the smallest. A single case of  $b$  different from 0.1, in which the RR(k) MSE was relatively high, was that with  $h^2 = 0.8$ ,  $n = 50$ ,  $b = 0.9$  and 10 blocks. In this configuration, the MSE of RR(k) was almost two times the smallest MSE.

The reasons why RR(k) did not show the smallest MSE, with high heritability, especially under relatively higher block effects, are not evident, since this approach theoretically yields the best linear unbiased predictions of random effects. However, when one

estimates variance components, predictors are no longer linear nor unbiased. A possible reason of reduction in precision might have been that the genetic variance taken as the parametric value was that among the actual 50, 100 or 200 lines, instead of the true population variance, considering all infinite possible inbred lines. The conditions under lower residual variation (higher  $h^2$ ) may have been more sensitive to this approximation. And, finally, even in the models where variance components were known, covariance coefficients were still unknown, and estimated via molecular information. So, it is not rigorously true that the RR(k) model should be the most precise ever.

The FR(k) approach had a trend very similar to that of RR(k), with regard MSE values, being less precise, in general. In 15 out of the 17 situations under high heritability, the RF(k) model had the smallest MSE. Therefore, there was a clear tendency of those approaches with known variance components to be most precise.

In an actual breeding program, variance components are generally (if not always) unknown, and so the comparison among FF, FR, RF and RR models is particularly relevant, especially to see whether the use of any mixed model (FR, RF or RR) shows any superiority to the intrablock analysis originally proposed by Federer (1956).

Considering only those models with unknown variance components in Table 1, the RR approach had the smallest MSE (higher precision) in most of the situations (53.7%), followed by RF (38.9%), similarly to the trends observed with the counterparts with known variance components. Again, most of the situations with RR not being the most precise model had high heritability (situations 37 to 54 of Table 1). However, in 4 of them heritability was equal to 0.5 and  $b$  was equal to 0.1 (situations 19, 25, 31 and 32), similarly to what happened with those models with known variance components. These results suggest that under low residual variation (high  $h^2$  and low  $b$ ), prediction of treatment effects is less precise in models where such effects are random than those where they are fixed. There is, however, even in such circumstances, an increase in precision if interblock information is recovered.

The values of the Spearman correlation among predictions with the alternative models and the parametric genetic effects are listed in Table 2. The rankings of the models considering random effects of blocks and treatments were expected to be the most accurate, since these effects were actually randomly generated. Indeed, the RR(k) approach had the highest Spearman correlation in 92.6% of the cases (50 situations), many of them even under high heritability. Therefore, these results suggest that, for truncated selection, if the variance components were known, the RR(k) approach would be the most suitable.

Table 1 - Average values of the Mean Square Error (MSE) of the predictions of genetic effects of inbred lines across 100 simulations of augmented block designs with  $n = 50, 100$  or  $200$  lines, heritability  $h^2$  equal to  $0.5, 0.5$  or  $0.8$ , coefficient of Smith  $b$  equal to  $0.1, 0.5$  or  $0.9$ , and number of blocks given through  $0.05n$  or  $0.2n$ . Each column refers to a different model, varying the nature of the effect of blocks (1st letter) and of treatments (2nd letter): fixed (F) or random (R)<sup>1</sup>.

h <sup>2</sup>	n	b	B	Block and treatment effects						
				FF	FR(k)	RF(k)	RR(k)	FR	RF	RR
1		0.1	10	442.91	238.21	435.14	237.43	300.56	436.88	297.56
2			2	541.55	263.17	533.07	262.89	338.61	550.34	335.81
3		0.5	10	1485.66	317.27	1325.30	316.03	351.54	1361.72	351.76
4	50		2	1496.51	309.32	1346.60	309.08	350.16	1411.18	347.39
5		0.9	10	2009.04	324.52	1567.63	319.75	370.36	1626.19	364.88
6			2	1868.18	370.39	1560.70	368.97	414.66	1658.49	406.90
7		0.1	20	461.23	265.82	449.90	265.02	440.77	452.11	354.44
8			5	601.199	260.90	582.40	260.68	392.28	598.17	345.47
9	100	0.5	20	1597.54	352.76	1387.57	351.09	420.04	1409.48	386.61
10	0.2		5	1609.29	325.43	1439.97	325.03	386.05	1471.42	360.37
11		0.9	20	1979.19	375.03	1569.38	368.60	431.03	1595.66	399.87
12			5	2032.83	353.82	1566.06	352.40	413.46	1677.57	383.54
13		0.1	40	470.80	253.13	463.32	252.95	404.41	464.03	345.80
14			10	609.20	240.80	591.83	240.66	367.44	601.08	328.50
15	200	0.5	40	1590.90	322.10	1395.86	319.44	387.70	1405.08	363.04
16			10	1734.79	344.89	1481.04	344.30	416.77	1529.82	396.22
17		0.9	40	2060.14	336.47	1589.24	330.42	395.16	1609.04	369.56
18			10	2087.43	344.11	1644.55	342.80	402.40	1694.74	382.11
19		0.1	10	126.04	128.74	122.70	127.61	237.74	123.18	212.52
20			2	151.90	126.83	149.553	126.35	263.27	154.06	253.95
21	50	0.5	10	381.11	238.36	339.82	231.89	309.71	349.99	283.25
22			2	408.86	242.40	369.20	240.44	335.50	388.08	323.06
23		0.9	10	519.47	229.83	400.89	218.69	287.38	416.68	275.41
24			2	479.71	235.83	404.15	232.40	298.76	430.68	294.62
25		0.1	20	126.81	137.19	124.89	136.45	375.74	125.67	272.20
26			5	169.31	143.86	165.67	143.37	392.72	171.59	316.99
27	0.5 100	0.5	20	388.27	229.60	347.05	224.60	393.22	353.32	308.56
28			5	433.92	227.82	380.62	226.77	379.87	396.61	324.80
29		0.9	20	507.20	260.75	400.18	248.09	412.410	405.69	321.81
30			5	531.92	240.81	416.59	236.64	383.96	439.07	327.20
31		0.1	40	116.90	141.15	115.28	140.38	382.99	115.40	297.84
32			10	164.94	154.49	160.92	154.21	384.80	162.67	320.20
33	200	0.5	40	412.41	252.66	365.21	247.58	417.71	368.82	351.28
34			10	431.49	219.66	374.64	218.69	358.10	385.70	317.12
35		0.9	40	519.16	251.31	408.22	240.67	384.62	411.35	326.80
36			10	544.66	256.17	435.60	252.20	401.88	445.78	359.34
37		0.1	10	40.89	58.57	40.34	58.15	210.88	40.54	174.56
38			2	57.84	72.47	57.18	72.07	295.34	57.98	269.00
39		0.5	10	114.66	126.97	105.08	120.84	239.19	106.90	206.09
40	50		2	116.83	112.89	106.58	109.93	283.68	110.53	255.33
41		0.9	10	146.83	264.89	126.58	234.93	149.68	118.53	135.33
42			2	137.14	119.15	114.27	113.21	270.38	122.39	255.94
43	0.8	0.1	20	36.19	341.11	35.60	215.80	54.40	35.56	53.78
44			5	58.68	70.21	57.54	69.55	371.27	58.48	284.29
45	100	0.5	20	113.68	135.38	102.27	129.05	417.97	104.37	288.33
46			5	121.43	118.42	108.18	116.25	389.58	111.14	295.05
47		0.9	20	132.33	137.25	105.17	124.46	364.75	106.41	251.42
48			5	149.63	132.58	123.82	127.15	368.66	129.30	293.54
49		0.1	40	38.099	93.29	37.48	93.30	358.03	37.54	268.76
50			10	54.30	93.06	53.64	92.45	359.63	54.00	288.65
51	200	0.5	40	108.65	138.40	96.84	134.20	358.44	98.00	278.46
52			10	126.45	134.14	108.38	131.68	362.31	111.83	303.89
53		0.9	40	135.07	151.16	106.39	139.51	378.38	107.17	289.64
54			10	146.30	150.49	118.29	144.75	390.42	121.24	325.32

<sup>1</sup>(k): Analysis with known variance components.

Table 2 - Average values of the Spearman correlation among predictions and actual genetic effects of inbred lines across 100 simulations of augmented block designs with  $n = 50, 100$  or  $200$  lines, heritability  $h^2$  equal to  $0.5, 0.5$  or  $0.8$ , coefficient of Smith  $b$  equal to  $0.1, 0.5$  or  $0.9$ , and number of blocks given through  $0.05n$  or  $0.2n$ . Each column refers to a different model, varying the nature of the effect of blocks (1st letter) and of treatments (2nd letter): fixed (F) or random (R)<sup>1</sup>.

h <sup>2</sup>	n	b	B	Block and treatment effects						
				FF	FR(k)	RF(k)	RR(k)	FR	RF	RR
1		0.1	10	0.509	0.556	0.513	0.557	0.469	0.514	0.471
2			2	0.476	0.511	0.480	0.511	0.428	0.477	0.429
3		0.5	10	0.332	0.352	0.340	0.356	0.291	0.342	0.294
4	50		2	0.337	0.357	0.350	0.358	0.290	0.343	0.294
5		0.9	10	0.307	0.341	0.348	0.360	0.289	0.340	0.308
6			2	0.287	0.321	0.313	0.328	0.265	0.309	0.272
7		0.1	20	0.512	0.560	0.513	0.561	0.409	0.512	0.420
8			5	0.462	0.514	0.466	0.514	0.378	0.465	0.383
9	100	0.5	20	0.324	0.370	0.342	0.375	0.276	0.340	0.280
10	0.2		5	0.311	0.364	0.323	0.365	0.278	0.321	0.281
11		0.9	20	0.297	0.344	0.322	0.360	0.267	0.322	0.282
12			5	0.285	0.340	0.315	0.342	0.269	0.306	0.274
13		0.1	40	0.498	0.570	0.502	0.571	0.398	0.502	0.393
14			10	0.466	0.558	0.471	0.559	0.386	0.468	0.384
15	200	0.5	40	0.328	0.405	0.343	0.411	0.279	0.342	0.281
16			10	0.324	0.417	0.347	0.418	0.287	0.341	0.287
17		0.9	40	0.301	0.376	0.334	0.395	0.261	0.332	0.274
18			10	0.290	0.369	0.322	0.372	0.246	0.318	0.249
19		0.1	10	0.710	0.725	0.713	0.727	0.606	0.712	0.611
20			2	0.662	0.684	0.664	0.685	0.566	0.658	0.567
21	50	0.5	10	0.538	0.571	0.560	0.580	0.471	0.557	0.484
22			2	0.542	0.573	0.561	0.575	0.468	0.551	0.474
23		0.9	10	0.478	0.520	0.523	0.539	0.441	0.516	0.459
24			2	0.470	0.513	0.494	0.520	0.428	0.485	0.437
25		0.1	20	0.710	0.727	0.713	0.729	0.500	0.711	0.519
26			5	0.678	0.714	0.684	0.715	0.492	0.678	0.504
27	0.5 100	0.5	20	0.522	0.567	0.538	0.573	0.413	0.535	0.427
28			5	0.510	0.568	0.530	0.569	0.422	0.523	0.428
29		0.9	20	0.490	0.542	0.530	0.567	0.409	0.527	0.430
30			5	0.475	0.542	0.519	0.552	0.401	0.511	0.415
31		0.1	40	0.710	0.741	0.712	0.742	0.491	0.712	0.488
32			10	0.678	0.728	0.682	0.729	0.488	0.681	0.487
33	200	0.5	40	0.536	0.600	0.553	0.610	0.406	0.552	0.406
34			10	0.515	0.597	0.538	0.599	0.407	0.534	0.406
35		0.9	40	0.488	0.567	0.529	0.587	0.386	0.527	0.394
36			10	0.493	0.583	0.528	0.588	0.397	0.525	0.398
37		0.1	10	0.855	0.858	0.857	0.860	0.680	0.856	0.694
38			2	0.838	0.843	0.840	0.846	0.644	0.838	0.652
39		0.5	10	0.732	0.741	0.744	0.749	0.619	0.742	0.637
40	50		2	0.738	0.753	0.752	0.759	0.600	0.746	0.609
41		0.9	10	0.698	0.753	0.722	0.659	0.700	0.726	0.709
42			2	0.689	0.712	0.714	0.723	0.585	0.707	0.593
43	0.8	0.1	20	0.863	0.857	0.865	0.865	0.593	0.843	0.822
44			5	0.835	0.843	0.838	0.844	0.560	0.835	0.576
45	100	0.5	20	0.724	0.747	0.740	0.758	0.512	0.737	0.542
46			5	0.723	0.756	0.738	0.759	0.534	0.734	0.549
47		0.9	20	0.691	0.718	0.723	0.741	0.506	0.721	0.542
48			5	0.704	0.735	0.732	0.745	0.526	0.726	0.544
49		0.1	40	0.856	0.838	0.858	0.840	0.526	0.858	0.526
50			10	0.843	0.827	0.845	0.828	0.523	0.843	0.522
51	200	0.5	40	0.727	0.748	0.744	0.760	0.495	0.742	0.498
52			10	0.717	0.755	0.741	0.760	0.504	0.736	0.502
53		0.9	40	0.691	0.725	0.724	0.746	0.481	0.723	0.496
54			10	0.697	0.738	0.730	0.751	0.486	0.727	0.491

<sup>1</sup>(k): Analysis with known variance components.

Table 3 - Average values of "elite bias" across 100 simulations of augmented block designs with n = 50, 100 or 200 lines, heritability  $h^2$  equal to 0.5, 0.5 or 0.8, coefficient of Smith b equal to 0.1, 0.5 or 0.9, and number of blocks given through 0.05n or 0.2n. Each column refers to a different model, varying the nature of the effect of blocks (1st letter) and of treatments (2nd letter): fixed (F) or random (R)<sup>1</sup>.

h <sup>2</sup>	n	b	B	Block and treatment effects							
				FF	FR(k)	RF(k)	RR(k)	FR	RF	RR	
1		0.1	10	0.104	0.037	0.100	0.035	-0.075	0.099	-0.066	
2			2	0.165	-0.041	0.162	-0.040	-0.129	0.168	-0.126	
3		0.5	10	0.178	-0.026	0.177	-0.025	-0.078	0.176	-0.084	
4	50		2	0.231	-0.141	0.229	-0.141	-0.106	0.232	-0.104	
5		0.9	10	0.203	-0.067	0.197	-0.061	-0.083	0.201	-0.067	
6			2	0.265	-0.118	0.249	-0.117	-0.080	0.253	-0.079	
7		0.1	20	0.056	0.039	0.056	0.036	-0.407	0.056	-0.024	
8			5	0.174	0.023	0.175	0.022	-0.248	0.177	-0.151	
9	100	0.5	20	0.101	-0.016	0.100	-0.011	-0.358	0.102	-0.123	
10	0.2		5	0.224	-0.113	0.223	-0.114	-0.196	0.224	-0.151	
11		0.9	20	0.109	-0.032	0.113	-0.024	-0.356	0.110	-0.123	
12			5	0.234	-0.137	0.227	-0.135	-0.193	0.229	-0.124	
13		0.1	40	0.046	0.020	0.047	0.020	-0.403	0.047	0.005	
14			10	0.122	0.061	0.124	0.061	-0.284	0.125	-0.083	
15	200	0.5	40	0.104	0.045	0.103	0.046	-0.340	0.104	-0.030	
16			10	0.170	-0.032	0.166	-0.032	-0.259	0.166	-0.165	
17		0.9	40	0.065	-0.004	0.067	-0.003	-0.392	0.068	-0.059	
18			10	0.178	-0.030	0.169	-0.029	-0.276	0.169	-0.138	
19		0.1	10	0.076	0.071	0.078	0.071	-0.023	0.077	0.041	
20			2	0.085	0.025	0.083	0.028	-0.175	0.085	-0.157	
21	50	0.5	10	0.109	0.036	0.100	0.034	-0.084	0.103	-0.049	
22			2	0.187	0.026	0.182	0.026	-0.100	0.183	-0.089	
23		0.9	10	0.099	0.036	0.092	0.039	-0.080	0.096	-0.055	
24			2	0.178	-0.019	0.178	-0.018	-0.117	0.176	-0.110	
25		0.1	20	0.049	0.043	0.047	0.043	-0.361	0.048	0.032	
26			5	0.083	0.039	0.085	0.040	-0.331	0.085	-0.118	
27	0.5	100	0.5	20	0.058	0.047	0.059	0.051	-0.394	0.058	0.003
28			5	0.126	0.024	0.128	0.025	-0.269	0.126	-0.134	
29		0.9	20	0.068	0.041	0.062	0.038	-0.372	0.063	-0.002	
30			5	0.148	0.022	0.146	0.024	-0.267	0.146	-0.138	
31		0.1	40	0.033	0.035	0.033	0.035	-0.370	0.034	0.021	
32			10	0.065	0.068	0.064	0.068	-0.301	0.067	0.003	
33	200	0.5	40	0.062	0.038	0.062	0.042	-0.371	0.062	0.022	
34			10	0.106	0.054	0.105	0.053	-0.271	0.107	-0.063	
35		0.9	40	0.038	0.014	0.038	0.018	-0.418	0.037	-0.004	
36			10	0.107	0.047	0.104	0.046	-0.296	0.103	-0.095	
37		0.1	10	0.020	0.001	0.019	-0.000	-0.066	0.019	0.016	
38			2	0.069	0.076	0.069	0.076	-0.090	0.069	-0.041	
39		0.5	10	0.043	0.025	0.044	0.029	-0.028	0.044	0.010	
40	50		2	0.095	0.069	0.092	0.070	-0.131	0.089	-0.066	
41		0.9	10	0.096	0.079	0.082	0.060	-0.141	0.087	-0.056	
42			2	0.120	0.050	0.115	0.055	-0.130	0.114	-0.110	
43	0.8	0.1	20	0.028	0.069	0.028	0.070	-0.339	0.027	0.017	
44			5	0.048	0.048	0.048	0.047	-0.326	0.048	-0.037	
45	100	0.5	20	0.072	0.068	0.075	0.064	-0.343	0.073	0.043	
46			5	0.118	0.091	0.119	0.092	-0.291	0.119	-0.036	
47		0.9	20	0.019	0.025	0.028	0.033	-0.398	0.026	0.017	
48			5	0.109	0.095	0.104	0.092	-0.282	0.105	-0.057	
49		0.1	40	0.028	0.048	0.028	0.048	-0.350	0.029	0.034	
50			10	0.027	0.069	0.028	0.070	-0.339	0.027	0.017	
51	200	0.5	40	0.066	0.053	0.065	0.053	-0.341	0.065	0.049	
52			10	0.059	0.069	0.055	0.071	-0.293	0.056	0.002	
53		0.9	40	0.040	0.024	0.041	0.027	-0.386	0.041	0.024	
54			10	0.103	0.092	0.098	0.100	-0.267	0.10	0.027	

<sup>1</sup>(k): Analysis with known variance components.

Focusing only on the approaches with unknown variance components, the RF model was generally superior to the others, with regard Spearman correlation (Table 2), having the highest values in 90.7% of the times (49 situations). In the remaining situations, the RF Spearman correlation was always close to the highest value observed (situations 20, 26, 38, 43 and 44). These results suggest that, for a wide range of conditions, if selection is truncated, interblock information should be recovered, the precision being at least comparable to the intrablock analysis. However, taking the randomness of the inbred lines into account seems unjustified, as well as the generation of molecular data, if one has to estimate variance components with the augmented block design.

In order to evidence those situations where recovery of interblock information is more noticeable, especially with truncated selection, one could take only those situations in Table 2 where the value of Spearman correlation of RF model was at least (say) 0.02 superior to that of FF model. There were 18 such cases, almost all with  $b$  equal to 0.9. Two exceptions were observed, with  $b$  equal to 0.5 (situations 9 and 16 of Table 2). This agrees with the fact that recovery of interblock information is more relevant when differences among blocks are small (high  $b$ ), as pointed out by Duarte (2000).

If, however, in a breeding program, the selection is made in relation to a standard cultivated variety (non-truncated selection), then the "elite bias" is of particular interest (Table 3). Contrary to the trend observed with Spearman correlation, the RR(k) model was superior (smallest elite bias in absolute values) in a relatively low frequency (29.6%, or 31 situations). However, it must be noticed that such model showed an elite bias almost always quite close to the smallest observed. In the cases where this was not observed (situations 29, 32 and 35 of Table 3), the smallest value was observed with RR model, which is different from RR(k) only in the sense that variance components are estimated.

The RR approach presented the smallest elite bias (in absolute values) in most of the cases (85.2%). These results suggest that if selection is non-truncated, then recovery of both interblock and intergenotypic information is suitable, especially if genetic similarities are accounted for using molecular data. The efficiency of the recovery of intergenotypic information when selection is non-truncated is related to the shrinkage effect, typical to be observed with the random effects of mixed models, yielding a more accurate distribution of such effects, in relation to the ordinary least squares means (fixed model). Results in Tables 2 and 3 suggest that the choice of the model should be based on the type of selection that is to be practiced.

The superiority of RR model, in relation to elite bias, is more pronounced under intermediate to high values of  $b$ . For instance, confronting situations 2, 4 and 6 of Table 3, under the same conditions but the value of  $b$ , equal to 0.1, 0.5 and 0.9, it can be seen that the difference of the RR elite bias from that of RF increases from 0.04, in situation 2, to 0.13 and 0.17 in situations 4 and 6. Similar comparisons can be made, for example, taking situations 8, 10 and 12. With high heritability, such differences tend to diminish, but the same trend is observed varying the value of  $b$  (see, for example, elite bias of situations 50, 52 and 54, of RR and RF models). Anyway, this suggests that with high precision (high heritability and low  $b$ ), non-truncated selection would also be effective considering treatment effects as fixed. The FR model was considerably poorer (Table 3), showing high (in absolute values) and negative elite biases, which would mean a high discard of elite lines.

The EM algorithm used in this study was extremely slow to converge, especially in the RR model, in which three variance components were to be estimated (residual, block and treatment). This suggests that this algorithm should be optimized to fit mixed models such as those of this study, or alternative numerical procedures should be investigated.

## REFERENCES

- BEARZOTI, E. Comparação entre métodos estatísticos na avaliação de clones de batata em um programa de melhoramento. Lavras, 1994. 128p. Dissertação (Mestrado) - Escola Superior de Agricultura de Lavras.
- BEARZOTI, E. Simulação recorrente assistida por marcadores moleculares em espécies autogamas. Piracicaba, 1997. 230p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- BOYLE, C.R.; MONTGOMERY, R.D. An application of the augmented randomized complete block design to poultry research. *Poultry Science*, v.75, p.601-607, 1996.
- DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society. Series B*, v.39, p.1-38, 1977.
- DUARTE, J.B. Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal. Piracicaba, 2000. 293p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- FEDERER, W.T. Augmented (or hoonuiaku) designs. *Hawaiian Planter's Record*, v.55, p.191-208, 1956.
- FEDERER, W.T. Recovery of interblock, intergradient, and intervarietal information in incomplete block and lattice rectangle designed experiments. *Biometrics*, v.54, p.471-481, 1998.
- HENDERSON, C.R. *Applications of linear models in animal breeding*. Ontario: Guelph, 1984.
- ROBINSON, G.K. That BLUP is good thing: the estimation of random effects. *Statistical Science*, v.6, p.15-51, 1991.
- SCOTT, R.A.; MILLIKEN, G.A. A SAS program for analyzing augmented randomized complete-block designs. *Crop Science*, v.33, p.865-867, 1993.
- SMITH, H.F. An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, v.28, p.1-23, 1938.
- WOLFINGER, R.D.; FEDERER, W.T.; CORDERO-BRANA, O. Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. *Agronomy Journal*, v.89, p.856-859, 1997.

Received May 08,2001