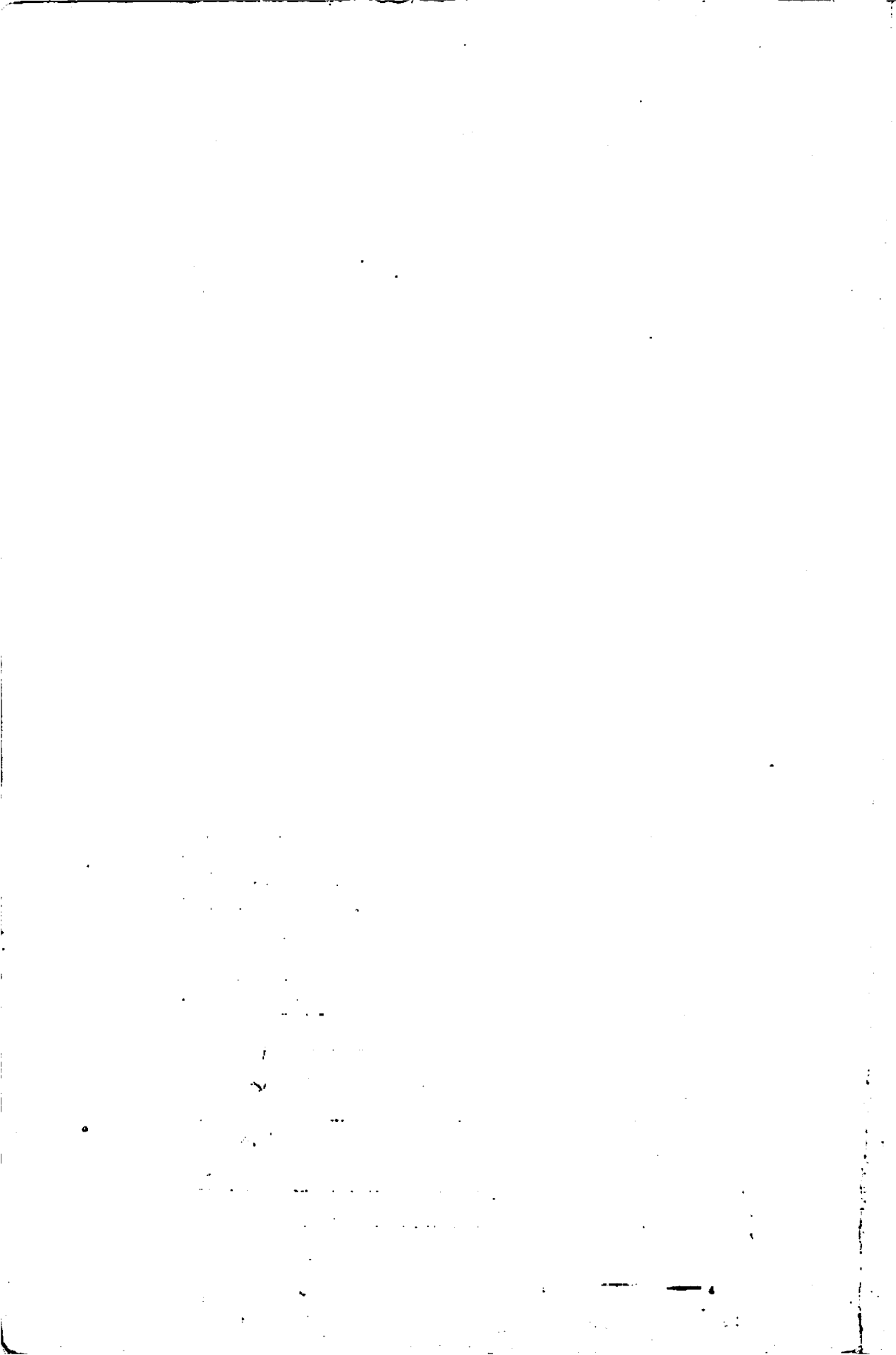


**AVALIAÇÃO DA MELHOR PREDIÇÃO LINEAR NÃO
TENDENCIOSA (BLUP) ASSOCIADA AO USO DE
MARCADORES MOLECULARES NA ANÁLISE
DIALÉLICA**

CLAUDOMIRO MOURA GOMES ANDRÉ

1999



CLAUDOMIRO MOURA GOMES ANDRÉ

**AVALIAÇÃO DA MELHOR PREDIÇÃO LINEAR NÃO
TENDENCIOSA (BLUP) ASSOCIADA AO USO DE
MARCADORES MOLECULARES NA ANÁLISE
DIALÉLICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Pós-Graduação em Agronomia, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de "Mestre".

Orientador

Prof. Dr. Eduardo Bearzoti

LAVRAS

MINAS GERAIS - BRASIL

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

André, Claudomiro Moura Gomes

Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialéctica / Claudomiro Moura Gomes André. -- Lavras : UFLA, 1999.

101 p. : il.

Orientador: Eduardo Bearzoti.

Dissertação (Mestrado) - UFLA.

Bibliografia.

1. Predição. 2. Método BLUP. 3. Cruzamento dialéctico. 4. Simulação.
 5. Marcador molecular. 6. Marcador genético. 7. Melhoramento genético.
 8. Componente de variância. 9. Modelo misto. 10. Variância genética.
- I. Universidade Federal de Lavras. II. Título.

CDD-631.53

CLAUDOMIRO MOURA GOMES ANDRÉ

**AVALIAÇÃO DA MELHOR PREDIÇÃO LINEAR NÃO
TENDENCIOSA (BLUP) ASSOCIADA AO USO DE
MARCADORES MOLECULARES NA ANÁLISE
DIALÉLICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Pós-Graduação em Agronomia, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de "Mestre".

APROVADA em 8 de janeiro de 1999

Prof. Dr. Magno Antonio Patto Ramalho

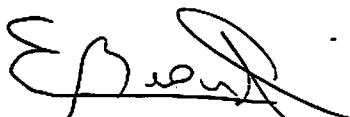
DBI - UFLA

Prof. Dr. Júlio Silvio de Sousa Bueno Filho

DEX - UFLA

Prof. Dr. Daniel Furtado Ferreira

DEX - UFLA



Prof. Dr. Eduardo Bearzoti

DEX - UFLA

(Orientador)

**LAVRAS
MINAS GERAIS - BRASIL**

SECRETARIA DE ECONOMIA FEDERAL

DEPARTAMENTO DE LICENCIAMENTO E REGISTRO DE EMPRESAS

REQUISITOS PARA O REGISTRO DE EMPRESAS

REQUISITOS PARA O REGISTRO DE EMPRESAS

REQUISITOS

1. Inscrição estadual no Livro de Registro de Empresas, nos termos do Regulamento de Registro de Empresas, aprovado pelo Conselho Administrativo de Recursos Fiscais (CARF), em 1964.

2. Inscrição municipal no Livro de Registro de Empresas, nos termos do Regulamento de Registro de Empresas, aprovado pelo Conselho Administrativo de Recursos Fiscais (CARF), em 1964.

- 100 - DUA
- 101 - DUA
- 102 - DUA
- 103 - DUA

- 104 - DUA
- 105 - DUA
- 106 - DUA
- 107 - DUA

108 - DUA

109 - DUA

110 - DUA

111 - DUA

112 - DUA

**“Dizem que sou louco, por eu ser assim...
mais louco é quem me diz, que não é feliz.”**

Hoje mais do que nunca eu sou feliz.

**Aos meus pais, Joaquim e Iolanda, pelo apoio e confiança a mim dedicados e
por suportarem o grande tempo de ausência.**

DEDICO

**Ao meu segundo pai e amigo, Ivaldo, responsável por grande parte do que
sou hoje.**

OFEREÇO

AGRADECIMENTOS

À Universidade Federal de Lavras e Departamento de Biologia, pela oportunidade de realização do curso.

Ao Conselho Nacional de Pesquisa (CNPq), pela concessão da bolsa de estudos.

Ao Eduardo Bearzoti, pela orientação segura, disponibilidade, incentivo e amizade demonstrados durante o curso.

Aos membros da banca examinadora, profs. Magno, Júlio e Daniel, pelas sugestões apresentadas.

Aos professores do curso de Genética e Melhoramento de Plantas, em especial aos professores Magno Antonio Patto Ramalho, Cesar Augusto Brasil Pereira Pinto e João Bosco dos Santos, sempre presentes e prestativos.

Ao professor Júlio Sílvio de Sousa Bueno Filho, pela ajuda valiosa e incentivo, durante a execução deste trabalho.

À minha família, que muito me ajudou durante toda a minha vida.

À Sônia, pelo carinho e dedicação neste tempo juntos.

Aos colegas de república, que estiveram ao meu lado durante a minha estada em Lavras.

Aos colegas de curso, os que já foram e os que ficam, o companheirismo e a amizade permanecem.

Aos amigos, Maguinho e Renatão, pelo incentivo.

A todos que confiaram e acreditaram em mim durante a realização deste trabalho, o meu **MUITO OBRIGADO**.

SUMÁRIO

RESUMO	i
ABSTRACT	iii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	5
2.1 Análise dialética	5
2.2 A metodologia dos quadrados mínimos	11
2.3 A metodologia BLUP	15
2.4 Componentes de variância.....	31
3 MATERIAL E MÉTODOS	39
3.1 Dados simulados.....	39
3.2 Dados experimentais	42
3.3 Análise dialética	43
3.4 Comparação entre os métodos BLUP e OLS	48
3.5 Predição de cruzamentos não avaliados	50
3.5.1 Dados simulados	50
3.5.2 Dados experimentais.....	51
4 RESULTADOS E DISCUSSÃO	53
4.1 Comparação entre os métodos BLUP e OLS	53
4.1.1 Ausência de dominância.....	53
4.1.2 Presença de dominância	60
4.2 Predição de cruzamentos não avaliados	69
4.2.1 Dados simulados	69
4.2.2 Dados experimentais.....	76
5 CONCLUSÕES	88
6 REFERÊNCIAS BIBLIOGRÁFICAS	89
ANEXO	98

RESUMO

ANDRÉ, Claudomiro Moura Gomes. Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica. Lavras: UFLA, 1999. 101p. (Dissertação - Mestrado em Genética e Melhoramento de Plantas)*

O objetivo deste trabalho foi o de avaliar a qualidade da melhor predição linear não tendenciosa (BLUP) na análise dialélica, quando existe informação disponível sobre a similaridade genética entre os genitores, obtida através de marcadores moleculares de herança dominante. Para tanto, utilizou-se dois critérios: a eficiência do BLUP em relação ao método dos quadrados mínimos ordinário (OLS), que é o procedimento padrão de análise dialélica para a predição de capacidades de combinação e efeitos de cruzamento, e a eficiência do BLUP em prever efeitos relativos a cruzamentos não observados (perdidos ou não avaliados). A utilização do primeiro critério baseou-se na simulação de 101 linhagens de uma espécie fictícia contendo 200 locos relativos a uma característica quantitativa (QTL) arbitrária e 100 locos marcadores de natureza dominante. Os efeitos dos QTL foram gerados conforme uma distribuição exponencial, e a distribuição dos alelos favoráveis ao longo das linhagens foi feita de maneira sistemática, de modo que não existissem linhagens de mesmo valor genotípico, e que a proporção de locos com alelos idênticos entre pares de linhagens variasse de 0 a 99%. Fez-se a distribuição dos alelos nos locos marcadores de tal maneira que a proporção média de alelos idênticos entre pares de linhagens fosse a mesma daquela dos QTL. A partir destas linhagens, construiu-se uma população de referência, considerando todos os 5050 cruzamentos possíveis, e a similaridade genética entre as linhagens foi quantificada a partir dos locos marcadores. Nesta população, variou-se a herdabilidade h^2 ao nível de médias de cruzamento (0,3, 0,5 e 0,7) e o grau de dominância (0, 0,5 e 1). Em seguida, construíram-se dialelos, tomando-se 12 linhagens ao acaso, considerando os métodos II e IV de Griffing. Tais dialelos foram analisados conforme os métodos OLS e BLUP, e as estimativas obtidas comparadas aos respectivos parâmetros (capacidades combinatórias e efeitos de cruzamentos). Para a avaliação da eficiência da predição de cruzamentos não realizados, dialelos utilizando 15 linhagens dessas populações de referência eram construídos, dos quais retiravam-se 30 cruzamentos aleatoriamente.

* Orientador: Prof. Dr. Eduardo Bearzoti, DEX-UFLA.

Além disso, dados experimentais de um dialelo com a cultura do feijão, utilizando informação de similaridade genética por meio de RAPD, foram também utilizados, simulando a retirada de cruzamentos. Os resultados permitiram verificar que o BLUP mostrou-se notoriamente superior ao método OLS, para a predição das capacidades de combinação e do efeito de cruzamentos, apresentando maiores correlações de Pearson e Spearman entre os valores preditos e os paramétricos. Embora maiores vieses tenham ocorrido com a predição utilizando o BLUP, no tocante à capacidade geral de combinação, a precisão em erro quadrático médio foi sempre marcadamente superior com o BLUP. Estas tendências se mantiveram ao longo dos diversos valores de h^2 e graus de dominância. No tocante à predição de cruzamentos não observados, houve correlações de Pearson e Spearman moderadas a elevadas, entre os valores preditos e os paramétricos, e entre aqueles e os valores observados nos dialelos amostrados (desconsiderando a perda de cruzamento em questão). Com os dados experimentais esta tendência se manteve, ainda que sob elevados níveis de desbalanceamento, com a perda de um número elevado de cruzamentos. Os resultados gerais puderam, assim, mostrar que o BLUP é uma alternativa eficiente para a análise dialélica, desde que haja informação disponível sobre a similaridade genética entre os genitores, com a vantagem adicional de possibilitar a previsão de cruzamentos que não puderam ser realizados, ou que foram perdidos.

ABSTRACT

ANDRÉ, Claudomiro Moura Gomes. **Evaluation of best linear unbiased prediction (BLUP) associated with information from molecular markers in diallel analysis.** Lavras: UFLA, 1999. 101p. (Dissertation – Master degree in Genetics and Plant Improvement)*

This work aimed at the evaluation of best linear unbiased prediction (BLUP) in diallel analysis, when information about genetic similarity among parents is available, obtained from dominant molecular markers. Two criteria for such evaluation were used: BLUP efficiency over ordinary least squares method (OLS), the standard method of diallel analysis for predicting combining abilities and cross effects, and BLUP efficiency on predicting effects of non-evaluated crosses (missing or not carried out). The former criterium was based on the simulation of 101 inbred lines of a fictitious species with 200 quantitative trait loci (QTL) and 100 dominant marker loci. QTL effects were generated from an exponential density function, and distribution of favorable alleles among lines was carried out systematically, in such a way that every line had an own genotypic value, and that proportion of loci with identic alleles among pairs of lines ranged from 0 to 99%. Distribution of alleles of marker loci was such that the average proportion values of identic alleles among pairs of lines was the same of those of QTL. Using such lines, a reference population was constructed, considering all 5050 crosses, and genetic similarities among lines were quantified using marker loci. In this simulated population, heritability h^2 values at the level of cross means were 0.3, 0.5 and 0.7, and degree of dominance was equal to 0, 0.5 and 1. Diallels were constructed by randomly taking 12 inbred lines from the whole set of 101, considering methods II and IV of Griffing. The corresponding diallel analyses were carried out using OLS and BLUP, and estimates were compared to the reference parameters (combining abilities and cross effects). The evaluation of BLUP efficiency on predicting non-observed crosses was carried out considering diallels constructed using 15 lines randomly taken from those reference populations, simulating cross losses. Experimental data were also used, from a common bean diallel, using similarity information from RAPD data, and also simulation of cross losses. Results showed that BLUP was considerably superior to OLS, for predicting

* Adviser: Prof. Dr. Eduardo Bearzoti, DEX-UFLA.

combining abilities and cross effects, with higher Pearson's and Spearman's correlations among predicted and parametric values. Higher biases were observed with BLUP, with regard to general combining ability, but mean squared error precision was considerably higher with this methodology. Such trends were observed for different h^2 and degree of dominance values. With regard to the prediction of non-observed crosses, moderate to high Pearson's and Spearman's correlations were obtained, among predicted and parametric values, and among predicted and those obtained considering the 15-line diallels individually, without cross losses. With experimental data, similar trends were obtained, even under high levels of unbalancement. In general, results suggested that BLUP is an efficient alternative for diallel analysis, given that molecular information about genetic similarities is available, with the additional advantage of allowing the prediction of crosses that could not be carried out, or missing ones.

1 INTRODUÇÃO

O melhoramento genético tem por principal finalidade o aumento da freqüência dos alelos favoráveis em uma determinada população, seja em animais ou plantas. Este aumento é obtido restringindo-se os indivíduos que serão utilizados como pais e que contribuirão com alelos para a formação de novas gerações. O processo de restrição de pais para gerações posteriores é conhecido como seleção (Falconer, 1981).

Em situações reais, a seleção é baseada no fenótipo dos indivíduos, que por sua vez é reflexo da ação conjunta do genótipo, do ambiente e da interação entre estes (Falconer, 1981). Assim, um dos pontos primordiais para se proceder a seleção com eficiência, consiste em se basear em informação que reflita o mais fielmente possível o valor genético de cada indivíduo que participa do processo.

Com este propósito, foram desenvolvidas, várias metodologias estatísticas que procuram o aproveitamento máximo das informações disponíveis. O desenvolvimento da genética quantitativa, através do emprego de técnicas estatísticas aliadas a modelos genéticos adequados, tem proporcionado métodos que propiciaram um grande impulso no melhoramento de plantas. Neste sentido, os cruzamentos dialélicos consistem em uma das técnicas que auxiliam na escolha de genitores com base nos seus valores genéticos, principalmente por considerarem a capacidade de se combinarem em híbridos, redundando em uma maior eficiência do programa de seleção (Vencovsky e Barriga, 1992).

Embora os valores genéticos sejam, por definição, efeitos aleatórios e devam ser preditos, os mesmos têm sido tratados, por conveniência dos métodos de análise, como sendo efeitos fixos (Resende, 1997). Neste caso, a predição dos valores genéticos é realizada pelo método dos quadrados mínimos ordinário (OLS), para o caso de homogeneidade de variâncias; dos quadrados mínimos ponderado (WLS), quando a pressuposição de homogeneidade não é atendida e pelo método dos quadrados mínimos generalizado (GLS), em casos de observações correlacionadas com ou sem heterogeneidade de variâncias (White e Hodge, 1989). O método dos quadrados mínimos ordinário é o comumente utilizado na análise dialélica.

Contudo, a utilização do método dos quadrados mínimos, em muitos casos, dependendo da estrutura dos dados, não apresenta as melhores estimativas dos valores genéticos (Henderson, 1974). Assim, a melhor estratégia de análise seria a de utilizar metodologias que consideram os valores genéticos como sendo efeitos aleatórios e como tal obter suas predições (White e Hodge, 1989; Resende, 1997 e Bueno Filho, 1997).

Tomando-se os valores genéticos como efeitos aleatórios, a sua predição pode ser efetuada utilizando-se uma metodologia que combina a melhor estimativa linear não tendenciosa dos efeitos fixos (BLUE), através da utilização de quadrados mínimos generalizados (GLS), com a melhor predição linear não tendenciosa dos efeitos aleatórios (BLUP) (Henderson, 1974, 1975a e 1984).

Esta técnica de análise tem a vantagem de considerar as covariâncias de caráter genético existentes entre os indivíduos, através da utilização de informações de parentesco ou similaridade genética obtidas por dados de genealogia ou por marcadores moleculares (Bernardo, 1994, 1995 e 1996a;

Euclides, 1996 e Nejati-Javaremi et al., 1997). A utilização de informações de parentesco melhora as predições dos valores genéticos, remove tendências atribuídas aos diversos grupos genéticos, bem como diminui influências de processos seletivos anteriores e pode até prever o valor genético de indivíduos não avaliados (Henderson, 1984; White e Hodge, 1989; Soh, 1994 e Bernardo, 1994). Este último aspecto adquire considerável importância na análise dialética, na predição de cruzamentos não realizados, seja pela perda aleatória ou pela economia de recursos.

A metodologia BLUP foi idealizada e empregada inicialmente em programas de melhoramento de gado de corte e leite nos Estados Unidos e Canadá, onde seu emprego é comum. Nos últimos anos, esta técnica de análise vem sendo amplamente empregada por melhoristas do mundo inteiro em diversas outras espécies de animais e também em vegetais, como por exemplo: Hill e Rosenberg (1985) em alfafa; Bouchez e Goffinet (1990) em soja; Bernardo (1994, 1995, 1996a, 1996b e 1996c) em milho; Soh (1994) em dendê; Huang et al. (1995) em gerbera; Panter e Allen (1995a e 1995b) em soja; Resende et al. (1996a e 1996b) em *Pinus* e Bueno Filho (1997) em *Eucalyptus grandis*.

Em princípio, a melhor predição linear não tendenciosa BLUP (teoria de modelos mistos) seria o método mais adequado para a predição de valores genéticos. Por outro lado, Henderson (1986) afirma que a principal restrição ao uso desta técnica é a grande demanda de recursos computacionais e que, dependendo da estrutura dos dados, com outras metodologias, inclusive a dos quadrados mínimos ordinário, pode-se obter estimativas semelhantes às dos modelos mistos. Neste contexto, deve-se procurar identificar sob quais situações o melhorista pode lançar mão de uma metodologia, em particular a

dos quadrados mínimos ordinário, em detrimento dos modelos mistos, tendo como principal vantagem a facilidade computacional.

Este trabalho teve como objetivos:

1. Comparar as metodologias de quadrados mínimos ordinário (OLS) e melhor predição linear não tendenciosa (BLUP), com informações de similaridade genética obtidas através de marcadores moleculares de natureza dominante, em modelos dialélicos com efeitos aleatórios, na predição das capacidades de combinação, usando dados simulados;
2. Avaliar a eficiência do BLUP na predição do desempenho de híbridos não realizados ou perdidos na análise dialélica.

2 REFERENCIAL TEÓRICO

2.1 Análise dialélica

O termo dialélico pode ser definido como o conjunto de p^2 possíveis combinações envolvendo p genitores, estando contidas nestas combinações os p genitores, o conjunto de $p(p-1)/2$ híbridos F_1 e o conjunto de $p(p-1)/2$ híbridos F_1 recíprocos, e à apresentação destas numa tabela, dá-se o nome de 'tabela dialélica' (Hayman, 1954a e Griffing, 1956). No entanto, este termo vem sendo usado de modo bastante geral, considerando variações em relação a este esquema.

Em cruzamentos dialélicos pode-se utilizar a tabela dialélica de quatro maneiras distintas (Vencovsky e Barriga, 1992; Ramalho, et al. 1993 e Cruz e Regazzi, 1994). A tabela completa com todos os híbridos possíveis e também os pais; a tabela completa sem a utilização dos pais; a meia tabela utilizando-se apenas os híbridos e os pais, sem os recíprocos; e por último, a meia tabela simples, com apenas os híbridos. A utilização de recíprocos é motivada pelo estudo de efeitos maternos, quando a característica sofre controle de genes extracromossômicos.

O desenvolvimento da teoria de análise dialélica teve início com a introdução dos conceitos de capacidade geral e específica de combinação por Sprague e Tatum (1942). Os autores definiram capacidade geral de combinação como sendo o comportamento médio de uma linhagem, cultivar ou material genético, em um conjunto de cruzamentos; e como capacidade

específica de combinação, o desvio observado de uma determinada combinação híbrida em relação ao que seria esperado somente pelas capacidades gerais de combinação dos respectivos genitores.

Segundo Vencovsky e Barriga (1992), os tipos de capacidade de combinação são diferenciadas de acordo com a natureza da ação gênica envolvida. A capacidade geral de combinação está associada a genes de efeitos predominantemente aditivos, além de efeitos de dominância e algumas interações epistáticas do tipo aditivo-aditivo. Já a capacidade específica de combinação é decorrente dos efeitos de dominância e de outras interações epistáticas; contudo, em geral considera-se que os efeitos epistáticos são de pequena magnitude e portanto são negligenciados para efeito de análise.

Dentre os vários métodos propostos para análise dialélica, os mais utilizados são os métodos propostos por Jinks e Hayman (1953), o desenvolvido por Griffing (1956) e o proposto por Gardner e Eberhart (1966). A primeira metodologia enfoca com maior ênfase o controle genético da característica, enquanto que as duas últimas, as capacidades de combinação.

O método proposto por Jinks e Hayman (1953), e ampliado por Jinks (1954), Hayman (1954a e 1954b), Jinks (1956) e Hayman (1958), permite que se trabalhe com gerações F_1 , F_2 e retrocruzamentos, utilizando-se a tabela dialélica completa com as p^2 combinações ou com a tabela desconsiderando os cruzamentos recíprocos. Há ainda a possibilidade de se utilizar a meia tabela, através do método desenvolvido por Jones (1965) e Ferreira (1985).

A análise baseia-se num modelo exclusivamente aditivo-dominante, sendo que parâmetros genéticos de interesse podem ser estimados. Hayman

(1954a) propôs ainda uma análise gráfica que permite a classificação dos materiais genéticos utilizados com base na proporção dos alelos dominantes em relação aos recessivos presentes nos materiais, estimar o grau médio de dominância e estabelecer limites seletivos que mostram o potencial das populações avaliadas para fins de seleção.

A metodologia proposta por Griffing (1956) baseia-se num modelo genético-estatístico em que podem ser obtidas estimativas das capacidades geral e específica de combinação, além do efeito de recíprocos dos materiais utilizados no dialelo. Esta metodologia foi desenvolvida para quatro tipos de tabelas dialélicas, correspondendo às diferentes alternativas de sua utilização mencionadas anteriormente, podendo-se considerar ainda o modelo fixo ou aleatório, dependendo de como os genótipos são determinados.

Griffing (1956) definiu como método I, o esquema de análise com a tabela completa, ou seja, incluindo os híbridos, os recíprocos e os próprios genitores. Para o método II, a análise é feita com os híbridos e os genitores, sem a presença dos recíprocos. O método III utiliza os híbridos e os recíprocos, excluindo-se os genitores. E por fim o método IV utiliza na análise apenas os híbridos, sem a presença dos pais e dos recíprocos.

Ainda segundo Griffing (1956), o modelo fixo supõe que os genitores foram escolhidos criteriosamente em relação a uma população de referência e, portanto, as estimativas obtidas da análise dialélica não poderiam ser extrapoladas para toda a população. Já o modelo aleatório supõe que os pais foram tomados ao acaso e representam a população de referência; neste caso as estimativas obtidas podem ser generalizadas.

Ramalho et al. (1993) avaliaram as implicações da escolha do modelo mais apropriado, se fixo ou aleatório, afirmando que se o modelo é fixo, o

interesse do melhorista está centrado nos componentes genéticos de médias que são estimados na análise dialélica. Para o modelo aleatório, o interesse se volta para a estimação de parâmetros populacionais, principalmente os componentes de variância associados às capacidades geral e específica de combinação. Os autores discutem ainda que a escolha dos genitores deve ser feita criteriosamente, baseando-se nos caracteres agronômicos de interesse e que, neste caso, estes não podem ser considerados uma amostra representativa de uma população, sendo assim aconselhável que os efeitos do modelo sejam considerados fixos, tendo as estimativas dos componentes de variância utilidade limitada.

Uma discussão relevante no que diz respeito a se considerar o modelo fixo ou aleatório, é apresentada por Henderson (1984). O autor define a escolha do modelo em decorrência dos objetivos do trabalho. O modelo deverá ser fixo, se o objetivo é a avaliação do desempenho de um cruzamento específico. Neste caso, a inferência deve ser feita em relação aos indivíduos do cruzamento já realizado. Entretanto, se o objetivo é avaliar o potencial produtivo de um indivíduo ou de uma linhagem quando cruzados, mesmo que estes sejam criteriosamente selecionados, deve-se considerar um modelo aleatório, dado que se tem apenas uma amostragem dos possíveis descendentes deste indivíduo, sendo as inferências pertinentes ao potencial reprodutivo do indivíduo ou da linhagem em questão.

Um terceiro enfoque de análise dialélica foi proposto por Gardner e Eberhart (1966), o qual possibilita a análise de tabelas dialélicas completas e também meias tabelas. As diferenças entre as médias da tabela são analisadas, tendo por referência um modelo de efeitos fixos. Neste método, o efeito de cruzamento é decomposto em efeitos de genitor e de heterose, com

significados próximos aos das capacidades geral e específica de combinação no modelo proposto por Griffing (1956). Entretanto, no modelo proposto por Gardner e Eberhart (1966), o efeito de heterose é ainda decomposto em heterose média, heterose de genitores e heterose específica.

Segundo Cruz e Vencovsky (1989), uma crítica comumente levantada ao método de Griffing, referente aos modelos em que são incluídos os pais, é a consideração do parâmetro ' s_{ii} ', que é a capacidade específica do cruzamento de um genitor com ele mesmo. Segundo eles, para outros autores, o parâmetro ' s_{ii} ' não teria significado genético, sendo também discutível o fato de a capacidade geral de combinação ser devida aos efeitos do desempenho dos genitores 'per se', quando em combinações híbridas. Neste caso, o método proposto por Gardner e Eberhart seria mais apropriado, já que este estima separadamente os efeitos de variedade e heterose de genitores, que seriam desdobramentos dos efeitos de capacidade geral de combinação.

No entanto, Cruz e Vencovsky (1989) estudaram os parâmetros do modelo de Griffing (1956), considerando os genitores como populações em equilíbrio de Hardy-Weinberg, e concluíram que o parâmetro ' s_{ii} ' indica a direção dos desvios de dominância, sendo negativo, quando os desvios forem predominantemente positivos; e positivo, em caso contrário; e quanto menor for o valor absoluto de ' s_{ii} ', mais próxima estará a frequência alélica do i -ésimo genitor da frequência média da população. Ainda ao comparar as duas metodologias em termos dos parâmetros genéticos estimados e da quantidade de informação obtida da análise dialélica, os autores concluíram que os dois métodos são equivalentes e também consideram o método II de Griffing igualmente eficiente em termos de informações contidas nos efeitos do

modelo, apesar de reconhecerem o modelo de Gardner e Eberhart como mais parametrizado.

O emprego dos cruzamentos dialélicos, na forma definida anteriormente, tem se limitado, na prática, a um pequeno número de pais, pois além do número elevado de cruzamentos a serem realizados, existem ainda as dificuldades das operações sob condições de campo, com nível desejável de precisão experimental. Essas limitações fizeram surgir uma série de alternativas para viabilizar sua utilização (Cruz e Regazzi, 1994). Assim, surgiu o conceito de dialelos parciais, com um número menor de cruzamentos, permitindo a avaliação de um número maior de pais com os mesmos recursos. Os sistemas dialélicos parciais se adaptam perfeitamente aos programas de melhoramento que visam a obtenção de híbridos, já que estes avaliam um número elevado de linhagens.

Kemphorne e Curnow (1961) desenvolveram uma metodologia para análise de cruzamentos dialélicos baseada em uma amostra de todos os possíveis cruzamentos de um número elevado de genitores. Esta metodologia passou a ser denominada de análise de dialelos parciais circulantes. O número de cruzamentos é reduzido em relação ao dialelo completo, mas a eficiência da metodologia para estimação de componentes de variâncias genéticas e capacidade de combinação não é afetada, desde que sejam avaliados um grande número de genitores.

Miranda Filho e Geraldi (1984) propuseram um modelo para análise de dialelos parciais que representa uma adaptação do modelo proposto por Gardner e Eberhart (1966). A metodologia baseia-se na decomposição da variação total dos cruzamentos, em efeitos associados a dois grupos de genitores distintos, ao contraste entre estes e também aos efeitos de heterose.

Adicionalmente, é realizado um estudo detalhado da heterose manifestada nos híbridos, dividindo a heterose total em heterose média, heterose de variedade dentro de cada grupo e, por fim, a heterose específica manifestada em cada combinação híbrida.

Posteriormente, Geraldi e Miranda Filho (1988) apresentaram uma metodologia para análise de dialelos parciais, segundo uma adaptação do método II de Griffing (1956). Foram desenvolvidas expressões para estimação dos efeitos de capacidades geral e específica de combinação e um contraste envolvendo os dois grupos de genitores, além do cálculo das somas de quadrados associadas a cada efeito.

A utilização de esquemas dialélicos é bastante difundida entre os melhoristas, dado que uma parcela importante do melhoramento é representada pela hibridação de cultivares ou linhagens, possibilitando a recombinação da variabilidade existente, para produzir novos materiais (cultivares ou híbridos) adaptados às mais diversas condições (Ramalho et al., 1993).

2.2 A metodologia dos quadrados mínimos

Considerando que a maioria das características quantitativas de interesse apresentam efeitos ambientais significativos, é fundamental o uso de técnicas estatísticas adequadas para a predição de efeitos genéticos. Em geral, estimadores lineares não tendenciosos são convenientes e, dentre estes, é usual buscar os melhores no sentido de apresentarem variância mínima.

Segundo Henderson (1974), a primeira descrição completa da técnica dos quadrados mínimos, como método de estimação de valores genéticos, foi feita por Yates em 1934. Desde então, esta tem sido a principal metodologia

de análise de dados em melhoramento, seja pela sua facilidade computacional, seja pelo número e tipo de pressuposições a serem consideradas no modelo em estudo. Em casos mais simples, com relação à estrutura de variâncias e covariâncias entre os efeitos, e também sobre o grau de desbalanceamento dos dados, sua eficiência é satisfatória, sendo suas estimativas ditas não tendenciosas e de variância mínima.

O método apresenta algumas restrições quanto à sua aplicação, tais como não poder estimar o valor genético de indivíduos não observados; as suas funções nem sempre são estimáveis, mas apenas algumas combinações lineares destas; e, dependendo do grau de desbalanceamento dos dados, os valores genéticos para alguns indivíduos podem ser sub ou superestimados (Henderson 1974).

O objetivo do método dos quadrados mínimos é estimar os efeitos genéticos ou funções lineares destes ajustados, para os demais fatores não genéticos, através do emprego de modelos lineares (Martins, 1995).

Considerando-se um conjunto de observações fenotípicas dispostas no vetor $y' = [y_1 \ y_2 \ \dots \ y_n]$, distribuído conjuntamente com os valores genéticos $g' = [g_1 \ g_2 \ \dots \ g_i]$, pode-se admitir o modelo linear:

$$y = X_1\beta + X_2g + e \quad (1)$$

em que,

- y : vetor de observações;
- X_1 : matriz de incidência dos efeitos fixos;
- X_2 : matriz de incidência dos efeitos genéticos, tomados como fixos;
- β : vetor de efeitos fixos a serem estimados;

g : vetor de efeitos genéticos a serem preditos e,

e : vetor de erros aleatórios associados a cada observação.

Ao se utilizar o método, os efeitos genéticos são tidos como fixos e assim pode-se admitir que a distribuição conjunta de y e e é:

$$\begin{bmatrix} y \\ e \end{bmatrix} \sim N \left(\begin{bmatrix} X_1\beta + X_2g \\ \phi \end{bmatrix}, \begin{bmatrix} R & R \\ R & R \end{bmatrix} \right)$$

em que ϕ é um vetor nulo e R é a matriz de variâncias e covariâncias das observações de y , admitida positiva definida.

O critério de estimação do método consiste em encontrar estimadores lineares de β e g , tais que as $E(\beta^\circ)$ e $E(g^\circ)$ sejam iguais a β e g respectivamente, e que a variância dos estimadores seja mínima, o que implica que $E[\hat{e} - e]^2$ seja mínimo. Isso equivale a encontrar vetores β° e g° , soluções do sistema:

$$\begin{bmatrix} X_1'R^{-1}X_1 & X_1'R^{-1}X_2 \\ X_2'R^{-1}X_1 & X_2'R^{-1}X_2 \end{bmatrix} \begin{bmatrix} \beta^\circ \\ g^\circ \end{bmatrix} = \begin{bmatrix} X_1'R^{-1}y \\ X_2'R^{-1}y \end{bmatrix} \quad (2a)$$

ou:

$$\begin{bmatrix} \beta^\circ \\ g^\circ \end{bmatrix} = \begin{bmatrix} X_1'R^{-1}X_1 & X_1'R^{-1}X_2 \\ X_2'R^{-1}X_1 & X_2'R^{-1}X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'R^{-1}y \\ X_2'R^{-1}y \end{bmatrix} \quad (2b)$$

Se os resíduos são não correlacionados, ou seja, todas as observações tomadas são independentes e admite-se variâncias iguais (homogeneidade de variâncias), V pode ser escrita da forma $V = I\sigma^2$, em que σ^2 é a variância residual; neste caso β° e g° de (2b) são ditos estimadores de quadrados mínimos ordinário (OLS) e o sistema (2a) é chamado de sistema de equações

normais. Se os resíduos são não correlacionados, mas não há homogeneidade de variâncias entre os diversos níveis dos fatores fixos, V pode ser escrita da forma $V = D\sigma^2$, em que $D = \text{Diag}(n_i)$ com n_i o número de observações em cada sub-classe. Neste caso, os estimadores são ditos de quadrados mínimos ponderado (WLS). Finalmente, para o caso em que as observações contidas em y são correlacionadas com ou sem homogeneidade de variâncias, V é uma matriz não diagonal, admitida positiva definida; tem-se então as estimativas de quadrados mínimos generalizados (GLS).

Uma restrição à utilização do método dos quadrados mínimos, como método de predição de valores genéticos é que, na maioria dos casos, as matrizes X_1 e X_2 não possuem posto coluna completo, sendo o sistema dado em (2a) consistente mas indeterminado, admitindo um número infinito de soluções para β e g . Entretanto, algumas funções lineares de interesse de β e g são perfeitamente estimáveis e invariantes (Searle, 1971 e 1987).

Uma solução para o sistema indeterminado em (2a), consiste na aplicação de restrições aos parâmetros a serem estimados, de tal forma que se complete convenientemente o posto das matrizes X_1 e X_2 . As restrições podem ser nos parâmetros ou nas soluções do sistema de equações normais. Desta forma, dependendo do tipo de restrições utilizadas, pode-se obter soluções distintas para o mesmo parâmetro (Searle, 1971). Segundo Henderson (1975a e 1975b) utilizando-se o método dos quadrados mínimos em modelos mistos, com efeitos fixos e aleatórios, as estimativas encontradas podem ser tendenciosas, além de não apresentarem variância mínima.

Descrições mais detalhadas do método dos quadrados mínimos e também sobre as propriedades pertinentes às suas estimativas podem ser

encontradas em publicações sobre modelos lineares tais como: Searle (1971 e 1987), Graybill (1976), dentre outros.

2.3 A metodologia BLUP

Na utilização dos quadrados mínimos para avaliação genética, admite-se que os valores genéticos sejam fixos, quando na realidade não o são, visto que, mesmo admitindo que as observações, realizadas sejam fruto de agregados gênicos determinados, e desta forma, podendo ser considerados como de efeitos fixos, deve-se considerar a segregação alélica, tratando o indivíduo como um veículo de alelos que segregam e novamente se agregam aleatoriamente formando novos genótipos, cuja observação da expressão se realiza na descendência (Martins, 1995). Some-se a isso o fato de que, em um determinado experimento, não se avalia todos os descendentes de cada indivíduo, mas apenas uma amostra deste quando cruzado com a população. A pressuposição errônea de que o valor genético seja fixo induz a distorções nas avaliações e influi nas estimativas dos valores genéticos (Henderson, 1974, 1975a e 1984).

Uma metodologia que consiste na avaliação dos valores genéticos tratados como sendo de efeito aleatório, corrigindo-os para os demais efeitos fixos contidos no modelo, sendo também apropriada para graus elevados de desbalanceamento, foi proposta por Henderson em 1949. Esta técnica foi denominada de Metodologia de Modelos Mistos (MMM), em que pode-se obter o melhor predição linear não tendenciosa, "Best Linear Unbiased Prediction - BLUP", dos valores genéticos de cada indivíduo, além da melhor estimativa linear não tendenciosa, "Best Linear Unbiased Estimation - BLUE", dos efeitos fixos (Henderson, 1974 e 1975a).

Um modelo misto pode ser expresso na forma geral:

$$y = X\beta + Zg + e \quad (3)$$

em que:

y : o vetor de observações;

X : a matriz de incidência dos efeitos fixos;

Z : a matriz de incidência dos efeitos genéticos, tomados como aleatórios;

β : o vetor de efeitos fixos a serem estimados;

g : o vetor de efeitos genéticos a serem preditos, e,

e : o vetor de erros aleatórios associados a cada observação.

Considerando-se que a média e a distribuição conjunta de y e g não são conhecidas, mas que se conhecem as variâncias e covariâncias de y e g , pode-se obter o vetor de predições \hat{w} , o BLUP de w (Henderson 1974, 1975a e 1984), que por sua vez é um vetor de variáveis aleatórias dados por:

$$w = K'\beta + L'g$$

em que K e L são matrizes conhecidas de contrastes estimáveis, ou seja, pertencentes aos espaços gerados pelas colunas de X e Z , respectivamente.

Segundo Henderson (1974), um preditor linear de w pode ser obtido por:

$$\hat{w} = \alpha + By$$

em que α e B são escolhidos de forma que seja minimizado o erro quadrático médio de predição dado por $E[\hat{w} - w]^2$. Isto equivale a minimizar a expressão:

$$E[(\alpha + By - K'\beta - L'g)' A(\alpha + By - K'\beta - L'g)] \quad (4)$$

em que A é uma matriz simétrica positiva definida sujeita à restrição de não tendenciosidade, ou seja, $E(\hat{w}) = E(w)$.

Sendo a distribuição conjunta de y e w tal que:

$$\begin{bmatrix} y \\ w \end{bmatrix} \sim N \left(\begin{bmatrix} X\beta \\ K'\beta \end{bmatrix}, \begin{bmatrix} V & LC \\ L'C' & L'GL \end{bmatrix} \right),$$

em que C é a matriz de covariâncias entre y e w ; G a matriz de variâncias e covariâncias entre os valores genéticos a serem preditos.

O preditor produzido pela minimização de (4), segundo Henderson, (1984) e Searle et al. (1992) é:

$$\hat{w} = K'\beta^\circ + L'C'V^{-1}(y - X\beta^\circ)$$

em que β° é a solução GLS para os efeitos fixos, num modelo $y = X\beta + Zg + e$, com $y \sim N(X\beta, V)$, dada por:

$$\beta^\circ = (X'V^{-1}X)^{-1}X'V^{-1}y \quad (5)$$

$K'\beta^\circ$ é o melhor estimador linear não tendencioso (BLUE) de $K'\beta$, enquanto que:

$$\hat{g} = C'V^{-1}(y - X\beta^\circ) \quad (6)$$

é o melhor preditor linear não tendencioso (BLUP) de g .

O modelo utilizado para estimação de β^o em (5) não ignora a presença dos efeitos genéticos, porque estes são coerentemente tomados como aleatórios e, desta forma considerados no modelo, e sua matriz de variâncias e covariâncias está contida em V , cuja inversa é utilizada na estimação de β^o , o que não acontece para solução GLS em (2b), quando os efeitos genéticos são considerados fixos. É preciso evidenciar que, em (2b), R contém apenas covariâncias residuais, enquanto que V de (5) apresenta também covariâncias genéticas. São portanto, matrizes diferentes que levam a soluções distintas para β .

A obtenção de BLUP, na forma como descrito em (5) e (6),

$$\beta^o = (X'V^{-1}X)^{-1} X'V^{-1}y$$

$$\hat{g} = C'V^{-1}(y - X\beta^o)$$

envolve frequentemente grande esforço computacional, uma vez que é necessária a obtenção da inversa de V , a matriz de variâncias e covariâncias de y , a qual é frequentemente de grande dimensão. Contudo, Henderson et al. (1959) já haviam derivado uma forma alternativa para obtenção de BLUP, através da maximização da função densidade de probabilidade conjunta de y e g , em relação aos efeitos fixos e valores genéticos.

Reconsiderando o modelo (3), e admitindo-se que:

$$\begin{bmatrix} y \\ g \\ e \end{bmatrix} \sim N \left(\begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} ZGZ' + R & ZG & R \\ & GZ' & G & \phi \\ & & R & \phi & R \end{bmatrix} \right)$$

em que,

0 é um vetor nulo;

ϕ é uma matriz nula;

$G = A\sigma_g^2$, sendo A a matriz que indica o grau de associação genética entre os indivíduos (matriz de parentesco genético aditivo) e σ_g^2 a variância genética;

R é a matriz de variância e covariâncias residual.

Dessa forma, a solução conjunta para os efeitos fixos e aleatórios, segundo Henderson (1984) é dado pelo sistema de equações:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^o \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad (7)$$

Essas são as chamadas Equações de Modelos Mistos (EMM), que permitem obter os melhores estimadores lineares não tendenciosos para os efeitos fixos ($X\beta^o$), através de quadrados mínimos generalizados, bem como os melhores preditores lineares não tendenciosos para os efeitos aleatórios (\hat{g}).

Da absorção de \hat{g} na primeira linha de (7), é possível a obtenção da solução para β , dada por:

$$\beta^o = (X'R^{-1}X)^{-1}(X'R^{-1}y - X'R^{-1}Z\hat{g}) \quad (8)$$

Substituindo-se (8) na segunda linha de (7) têm-se:

$$\begin{aligned} Z'R^{-1}X(X'R^{-1}X)^{-1}(X'R^{-1}y - X'R^{-1}Z\hat{g}) \\ + (Z'R^{-1}Z + G^{-1})\hat{g} = Z'R^{-1}y \end{aligned}$$

1. as soluções para os efeitos fixos e aleatórios são obtidas em seqüência por (9) e (10) e de forma independente, o que reduz a dimensão das matrizes a serem invertidas;

computacional, pelas razões a seguir:
 obtenção das soluções de EMM reduz, em geral, sensivelmente o esforço Conforme comentado por Martins (1995), essa forma alternativa de

$$\beta = \{X'R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}\}X' \\ X'R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1} \quad (10)$$

Da absorção de (9) em (8) resulta:

$$\hat{\beta} = (Z'SZ + G^{-1})^{-1}Z'Sy \quad (9)$$

ou seja:

$$(Z'SZ + G^{-1})\hat{\beta} = Z'Sy$$

será:

Fazendo-se $S = R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}$, o sistema resultante

$$\begin{aligned} & \{Z'R^{-1} - R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}\}Z + G^{-1}\} \hat{\beta} \\ & = Z'R^{-1}y - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}y \\ & [Z'R^{-1}Z - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}Z + G^{-1}]\hat{\beta} \\ & = Z'R^{-1}y - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}y \\ & - Z'R^{-1}X(X'R^{-1}X)^{-1}X'R^{-1}Z\hat{\beta} \\ & + (Z'R^{-1}Z + G^{-1})\hat{\beta} = Z'R^{-1}y \end{aligned}$$

2. A ordem das matrizes a serem invertidas é em geral menor que a dimensão de V .
3. a matriz R é em geral tomada como uma matriz diagonal ou bloco diagonal, o que torna sua inversão facilitada;
4. as soluções para os efeitos fixos em (10) são soluções de quadrados mínimos generalizados (GLS) de β , já que se pode mostrar que $V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}$ conforme mostrado em Lopes et al., (1993);
5. apesar de a matriz $(Z'SZ + G^{-1})$ em (9) ser de grande dimensão e não diagonal, a sua inversão pode ser obtida de forma eficiente através de procedimentos iterativos, como mostrado em Lopes et al. (1993) e Verneque (1994).

As propriedades pertinentes ao BLUP foram descritas de maneira detalhada por Henderson (1984), e evidenciam a superioridade teórica do BLUP como metodologia de predição de valores reprodutivos. Esta superioridade tem sido confirmada pelas comparações, a partir de dados de campo e, principalmente, através do uso de simulações como, por exemplo, em Sorensen e Kennedy (1984); Panter e Allen (1995a e 1995b); Jeyaruban et al. (1995) e Bueno Filho (1997).

White e Hodge (1989) apresentam várias propriedades do BLUP, referentes ao melhoramento vegetal, dentre as quais se destacam:

1. estimação e predição não-viesadas num procedimento linear único;
2. pode levar em conta efeitos de seleção e de endogamia ao longo das gerações, desde que o grau de relacionamento genético (parentesco) entre os indivíduos avaliados seja conhecido;

3. a correlação entre os valores genéticos verdadeiros e os preditos é máxima dentre todas as classes de preditores lineares não tendenciosos;
4. pode-se prever o valor genético de indivíduos observados ou não;
5. as variâncias e os erros das predições BLUP são menores em relação a outros métodos;
6. para dados multivariados com distribuição normal, a probabilidade de se selecionar o melhor indivíduo é máxima entre as classes de estimadores lineares não viesados com média zero.

Sorensen e Kennedy (1984), usando simulação de populações com características semelhantes às de suínos, compararam a eficiência de dois métodos de seleção, o método de quadrados mínimos ordinário e o BLUP. Foram avaliados os ganhos obtidos com a seleção em cada um dos métodos durante três gerações consecutivas. Os autores observaram que as estimativas de ganho obtidas com o BLUP foram sempre superiores às obtidas pelo método dos quadrados mínimos ordinário. Além disso, os ganhos estimados pelo método BLUP foram mais próximos das estimativas de ganho quando a seleção era feita utilizando os verdadeiros valores genéticos.

Panter e Allen (1995a) estudaram a eficiência dos métodos de quadrados mínimos ordinário (OLS) e o BLUP, na identificação de linhagens superiores em cruzamentos de variedades de soja. Os autores simularam situações com igual número de dados em cada sub-classe (balanceamento dos dados), e também situações de desbalanceamento. Em todas as situações estudadas, as predições BLUP apresentaram menor erro padrão, maiores valores de correlação classificatória entre os valores preditos e o desempenho

dos cruzamentos avaliados em campo, bem como maior porcentagem de identificação de cruzamentos superiores dentre os realizados.

Panter e Allen (1995b), ao comparar dois modelos de BLUP (com e sem a inclusão da informação de parentesco genético das linhagens utilizadas nos cruzamentos) com o método de quadrados mínimos ordinário (OLS) para a predição de cruzamentos em soja, observaram que os erros padrões das estimativas obtidas pelo método OLS foram em média 6% maiores que as obtidas com os modelos de BLUP. Entretanto, não foram observadas diferenças marcantes entre os dois modelos de BLUP, sendo o modelo com informações de parentesco ligeiramente superior. Segundo os autores, estimativas com menores erros associados permitem ao pesquisador uma maior segurança na discriminação dos genótipos, e esta característica do método de análise pode ser traduzida em maiores ganhos com a seleção.

Jeyaruban et al. (1995), usando simulação de populações durante 10 gerações de seleção, encontraram uma maior eficiência do BLUP como método de seleção, quando comparado com outros cinco índices de seleção. Entretanto, a eficiência diminuiu à medida que o valor da herdabilidade da característica estudada aumentou. Um fator ressaltado pelos autores para a maior eficiência do BLUP, quando aplicado a características de baixa herdabilidade, é o fato de esta metodologia levar em consideração informações da covariância genética entre os genitores que participam do processo seletivo, informações estas desprezadas pelas outras metodologias utilizados no estudo. Contudo, em algumas situações de herdabilidade elevada ($h^2 > 0,6$), alguns índices apresentaram respostas semelhantes ao BLUP, sendo a sua aplicabilidade vantajosa devido ao fato de o dispêndio computacional envolvido ser menor, ou seja, em algumas situações a perda de

informações seria compensada pela facilidade e agilidade dos cálculos envolvidos.

Zhu e Weir (1996) compararam diferentes métodos de predição, como o BLUP com componentes de variâncias conhecidas, com suas estimativas, e um modelo denominado como "Adjusted Unbiased Prediction" (AUP) em modelos dialélicos em presença de efeitos maternos. Foram utilizados dados simulados e dados reais. Os autores encontraram boa adequação com o BLUP; entretanto, as predições BLUP com estimativas simultâneas dos componentes de variância mostraram-se menos eficientes que o BLUP com variâncias conhecidas, sendo que o método AUP foi considerado o de melhor adequação ao fenômeno estudado.

Embora a teoria BLUP tenha sido inteiramente desenvolvida e aplicada com sucesso em programas de melhoramento animal, sua extensão para o melhoramento vegetal também tem sido proposta, como por exemplo em White e Hodge, 1989; Resende, 1997 e Bueno Filho, 1997. A predição de valores genéticos se daria em modelos mistos onde todos os parâmetros ambientais fossem tidos como fixos e os valores genéticos tomados como aleatórios.

Uma das principais características do BLUP é a de utilizar informações de parentesco genético aditivo entre todos os pares de indivíduos. Os indivíduos aparentados têm uma proporção maior de alelos em comum relacionada ao grau de parentesco (ex: o parentesco genético aditivo entre pai e filho é de 50%, entre meios irmãos 25%, entre irmãos germanos 50%, etc.), sendo este parentesco levado em consideração na análise.

O termo definido como "parentesco genético aditivo" refere-se à proporção de alelos idênticos por ^a ascendência entre dois indivíduos e é

equivalente ao dobro do valor do coeficiente de parentesco de Mallecot (Van Vleck, 1993b).

Dado que indivíduos aparentados têm alelos em comum, o uso de suas observações nas avaliações incrementa a precisão das predições. Adicionalmente, a inclusão de parentes permite maior conexão entre os diferentes experimentos, sem a necessidade da inclusão de testemunhas adicionais. De forma similar, a utilização de informações de parentesco também possibilita a comparação de indivíduos pertencentes a gerações distintas de seleção, o que geralmente não é possível na prática (Bueno Filho, 1997).

O BLUP possibilita, além da predição mais precisa do mérito genético, a obtenção do coeficiente de endogamia para todos os indivíduos avaliados, se toda a informação de parentesco é fornecida (Henderson, 1974 e 1984). Quando se dispõe de pouca ou nenhuma informação de um indivíduo, informações de parentes podem contribuir para a predição do seu valor genético. A magnitude desta contribuição é determinada pelo grau de parentesco entre os indivíduos envolvidos na análise, sendo que, quanto maior for o grau de relacionamento genético, maior será esta contribuição (Wood et al., 1991; Van Vleck, 1993a; Panter e Allen, 1995a).

Bueno Filho (1997) afirma que, para fins de seleção, a utilização do BLUP seria uma excelente estratégia, desde que boas estimativas dos componentes de variância estejam disponíveis. Entretanto, na prática do melhoramento vegetal, ainda é dada preferência a métodos de seleção fazendo uso de médias fenotípicas gerais, derivadas de um modelo fixo (OLS), pois estes fazem pressuposições (em geral falsas) sobre a dependência entre os valores genéticos dos indivíduos sob seleção.

Em casos de balanceamento de dados (número igual de observações em cada classe ou subclasse), e ausência de parentesco entre os genótipos avaliados, as predições BLUP e de quadrados mínimos ordinário apresentam a mesma ordem de classificação, sendo as predições BLUP, as que apresentam menores estimativas de erro padrão de predição (White e Hodge, 1989 e Van Vleck, 1993b).

Até recentemente, a utilização de dados de parentesco pelo método BLUP restringia-se apenas aos casos em que se dispunha de informações confiáveis da genealogia dos indivíduos avaliados, definida como identidade alélica por ^oascendência, ou seja, a semelhança entre os indivíduos só é levada em consideração para os alelos idênticos advindos de um ancestral comum (identidade por ^oascendência definida por Mallecot, 1948). Entretanto, com o avanço das técnicas de amostragem molecular, alguns autores propõem a substituição da matriz de parentesco genético aditivo, obtido através de informações de genealogia por valores de similaridade alélica em estado, como por exemplo: Bernardo (1994, 1995 e 1996a); Euclides (1996) e Nejati-Javaremi et al. (1997).

Segundo Nejati-Javaremi et al. (1997), a vantagem da utilização de dados de similaridade alélica em estado seria a possibilidade de levar em consideração, para efeito de análise, toda a semelhança alélica existente entre os indivíduos avaliados, sendo até mesmo possível diferenciar graus de similaridade entre diferentes pares de indivíduos irmãos germanos, o que não ocorre com dados de genealogia.

Bernardo (1994) comparou os resultados das predições BLUP obtidos por meio de informações de parentesco através de dados de genealogia de dois grupos de linhagens de milho, com o uso de dados de marcadores

moleculares do tipo RFLP, sendo que o principal objetivo do trabalho consistiu em avaliar a eficiência do BLUP em prever cruzamentos não realizados. Não foram verificadas alterações significativas nos resultados obtidos pelos dois procedimentos; desta forma, a utilização da informação de marcadores moleculares seria uma alternativa viável, desde que não se disponha de informações confiáveis da genealogia das linhagens avaliadas. Uma restrição à utilização desta abordagem seria a elevação dos custos experimentais decorrente do processo de amostragem molecular.

Euclides (1996), em estudo de seleção por vários procedimentos estatísticos, entre eles, duas abordagens do método BLUP; sendo a primeira, com informações de parentesco genético aditivo decorrente de genealogia; e a segunda, com a matriz de parentesco através de marcadores moleculares, simulou 100 ciclos seletivos para situações de alta e baixa herdabilidades, com cinco repetições cada. Ao longo das 100 gerações de seleção, não foram observadas diferenças elevadas entre as estimativas de ganho para as duas abordagens do método BLUP, sendo que em algumas situações o BLUP, com uso de marcadores moleculares, apresentou uma pequena superioridade.

Nejati-Javaremi et al. (1997) avaliaram a possibilidade de substituição da matriz de parentesco genético aditivo, por dados de similaridade alélica, substituindo os valores de parentesco genético aditivo por aqueles. Foram simuladas populações com herdabilidades 0,1 e 0,3, e a seleção foi aplicada a famílias de irmãos germanos e meios irmãos por quatro gerações, com 10 repetições cada. Em todas as situações estudadas, o BLUP, com informações de similaridade alélica, apresentou maiores estimativas de correlação entre o valor predito e o verdadeiro valor genético. Além disso, foi o método que propiciou maiores estimativas de ganho genético, ao longo das

gerações estudadas. Os autores concluem que a inclusão de informações de identidade alélica dos indivíduos avaliados é uma importante fonte de controle do erro de predição, e ainda que este tipo de informação aumenta consideravelmente os ganhos com a seleção.

Uma possível utilização do BLUP em melhoramento vegetal, citada por Bernardo (1994 e 1995), consiste na predição do desempenho genético de híbridos que não estejam presentes ou que porventura tenham sido perdidos. Assim, existindo uma estrutura de covariâncias genéticas entre as linhagens a serem cruzadas, pode-se prever o desempenho genotípico de alguns dos seus cruzamentos, sem que estes sejam avaliados.

A vantagem desta técnica seria a de prever o desempenho de uma quantidade considerável de combinações híbridas, com um número reduzido de cruzamentos avaliados em campo. Apesar de a correlação entre o desempenho genotípico do cruzamento predito e o seu valor fenotípico ser em muitos casos baixa, no máximo igual à raiz quadrada da herdabilidade da característica (Bernardo, 1994), o autor considera relevante o uso de técnicas preditivas como uma alternativa viável para diminuir os custos e também o tempo com a avaliação de híbridos em um programa de melhoramento de grande porte, onde são avaliados milhares de combinações híbridas anualmente e apenas uma pequena quantidade (<1%) é aproveitada como híbridos comerciais.

A utilização de métodos precisos de predição do desempenho de cruzamentos ainda não realizados pode aumentar consideravelmente a eficiência de programas de melhoramento. Contudo, esforços neste sentido não têm obtido êxito (Bernardo, 1994). A utilização do BLUP, como método de predição de cruzamentos não realizados, tem mostrado eficiência

considerável, conforme relatado por Bernardo (1994, 1995, 1996a e 1996b). A habilidade da metodologia BLUP em prever o desempenho de cruzamentos não realizados, bem como o número de preditores (cruzamentos a serem realizados) necessários para obter previsões suficientemente confiáveis, depende do grau de parentesco dos pais utilizados como preditores (Bernardo, 1995).

Bernardo (1994), usando a metodologia de modelos mistos, com estimativas de parentesco obtidas através de marcadores moleculares do tipo RFLP, para prever o desempenho de alguns cruzamentos não realizados em um grupo de 54 híbridos de milho, provenientes de dois grupos de linhagens avaliadas em um dialelo parcial, obteve estimativas de correlação entre o valor genotípico predito e o desempenho do cruzamento em campo para a produção de grãos, variando de 0,654 a 0,800. Apesar de o autor considerar os valores de correlação entre o valor observado e o predito não muito elevados, estes representam cerca de 70 a 90% do valor máximo possível (raiz quadrada da herdabilidade estimada para a característica). A eficiência na previsão de cruzamentos não realizados aumenta à medida que se utiliza um maior número de cruzamentos como preditores, o que mostra a viabilidade da utilização desta metodologia, como forma de identificação de cruzamentos promissores antes de sua realização. Contudo, o número de híbridos utilizados no trabalho era bastante restrito, de apenas 54.

Bernardo (1995) observou resultados semelhantes aos anteriores, ao testar alguns modelos genéticos de análise e graus elevados de desbalanceamento do dialelo. Os valores de correlação entre o desempenho observado e predito para produção de grãos, variaram entre 0,58 a 0,74, quando foi utilizado um grupo de 67 híbridos. Os valores das estimativas de

correlação foram considerados elevados, mesmo em situações em que o número de preditores utilizados era menor que 20% do número de híbridos com desempenho a ser predito (maiores graus de desbalanceamento do dialelo). Estes resultados são indícios de que o BLUP pode ser eficiente na predição do valor genético de um número elevado de híbridos, avaliando-se um número bem menor de cruzamentos que o utilizado geralmente.

Em soja, a utilização da metodologia BLUP, como método de predição do desempenho genotípico de cruzamentos ainda não realizados, comparado com o valor médio dos pais (método '*Mid Parent*') estimado por quadrados mínimos ordinário, foi investigada em várias situações por Panter e Allen (1995a e 1995b). As predições obtidas pela metodologia BLUP foram sempre mais próximas do desempenho fenotípico dos cruzamentos avaliados em campo; de todos os possíveis cruzamentos, o método BLUP foi o que apresentou a maior proporção de cruzamentos superiores identificados, bem como maior coincidência na classificação dos cruzamentos, mesmo quando o número de observações era restrito.

Bernardo (1996a e 1996b), utilizando um conjunto de 4099 híbridos de 16 combinações entre nove grupos heteróticos de milho avaliados em até 15 locais dos EUA, Canadá, Itália e França, encontrou estimativas de correlação entre o valor predito e o desempenho fenotípico para a produção de grãos de híbridos de milho variando entre 0,49 a 0,76, com uma tendência de estimativas mais elevadas para os grupos com maior número de híbridos avaliados. Segundo o autor, estes valores podem ser considerados moderadamente elevados, e sugerem que o BLUP seja empregado de forma eficiente para identificação de cruzamentos não realizados com desempenho superior. Entretanto, sua utilização em larga escala, como método preditivo

de cruzamentos não realizados, necessita ser investigada, principalmente no tocante à quantidade e ao tipo de cruzamentos a serem utilizados como preditores, bem como aos modelos genéticos mais adequados.

2.4 Componentes de variância

O uso da metodologia de modelos mistos, para a predição de valores genéticos, requer o conhecimento prévio dos componentes de variâncias e covariâncias envolvidos. Desta forma, uma determinação precisa das estimativas dos componentes da variância melhora tanto o processo de estimação, como o de predição, podendo resultar em alterações nas estimativas dos efeitos fixos, bem como nas predições dos efeitos aleatórios.

Diversas metodologias têm sido sugeridas para a estimação desses componentes. Dentre elas, citam-se os métodos I, II e III de Henderson ou métodos da análise de variância (ANAVA) ou ainda, método dos momentos (Henderson, 1953); Método dos Estimadores Não Viesados de Norma Mínima, ou MINQUE (Rao, 1970 e Rao, 1971a); Método de Estimação Não Viesada de Mínima Variância Quadrática, ou MIVQUE (Rao, 1971b); Máxima Verossimilhança, ML (Hartley e Rao, 1967); Máxima Verossimilhança Restrita, REML (Paterson e Thompson, 1971) e o da Máxima Verossimilhança Integrada, VEIL (Gianola e Foulley, 1990).


O método I de Henderson aplica-se a procedimentos de análise de variância para dados balanceados ou desbalanceados, em modelos completamente aleatórios, consistindo em igualar os quadrados médios às suas esperanças matemáticas e resolver o sistema de equações formado. Esse método fornece estimativas não viciadas, com variância mínima, quando os dados são balanceados e os efeitos são não correlacionados.

Tendo em vista a limitação do método I de Henderson, a de exigir modelos completamente aleatórios, o autor desenvolveu os métodos II e III, os quais podem ser aplicados a modelos mistos.

O método II consiste em estimar os efeitos fixos do modelo, ignorando os efeitos aleatórios, ajustar os dados para os efeitos estimados e, posteriormente, aplicar o método I, considerando-se o modelo completamente aleatório. Esse método fornece estimadores únicos para os componentes de variância e covariância. Além disso, pode ser utilizado para qualquer modelo misto, desde que não haja interações entre os efeitos fixos e aleatórios ou aninhamento de fatores aleatórios dentro de efeitos fixos (Henderson, 1984).

O método III de Henderson é capaz de estimar componentes de variância para qualquer tipo de modelo misto, utilizando o método de ajustamento de constantes de Yates, em que a soma de quadrados de cada fator é calculada pela diferença entre as somas de quadrados obtidas pelo ajuste de um modelo completo e de um submodelo que exclui o fator em questão (reduções de formas quadráticas). Após a obtenção das somas de quadrados, são calculados os quadrados médios que, igualados às suas respectivas esperanças matemáticas, fornecem as estimativas dos componentes de variância.

Este método fornece estimadores não viciados e invariantes à translação, mas tais estimadores não são únicos, uma vez que se podem definir ou obter mais reduções do que as necessárias para estimar os componentes de variância; em alguns casos, o método fornece estimativas mais precisas do que as dos métodos I e II, mesmo nos casos em que estes possam ser plenamente aplicados (Verneque, 1994). Em populações sob seleção, o caso mais comum em melhoramento, este método produz



estimativas dos componentes de variância viesados (Sorensen e Kennedy, 1988).

Os métodos dos Estimadores Não Viesados de Norma Mínima (MINQUE) e dos Estimadores Não Viesados de Mínima Variância Quadrática (MIVQUE), são baseados diretamente na estimação de funções quadráticas dos componentes de variância, por meio de formas quadráticas do vetor de observações.

A estimação de tais funções é sujeita às restrições de não tendenciosidade e de invariância à translação. Os dois métodos diferem apenas na forma usada para estabelecer a precisão dos estimadores, sendo que o primeiro assume que a norma Euclidiana da matriz núcleo seja mínima, e o segundo impõe-se a restrição de que a variância dos estimadores seja mínima. Estes critérios não asseguram que estes sejam os estimadores mais adequados e também dependem do conhecimento prévio dos valores dos componentes a serem estimados (Henderson, 1984).

Brown (1976) propôs a utilização do procedimento MINQUE como um processo iterativo denominado I-MINQUE (Iterative-MINQUE), em que a solução MINQUE de um determinado passo serve de aproximação dos verdadeiros valores e estes geram uma nova solução, o processo se repetindo até uma convergência estabelecida. As soluções I-MINQUE apresentam propriedades estatísticas mais vantajosas, quando comparadas às do MINQUE original.

Sob normalidade, os estimadores obtidos pelos dois métodos são equivalentes. Assim, o método MIVQUE é mais utilizado e conhecido, tendo em vista que na maioria dos trabalhos científicos, na área biológica, assume-se que os dados seguem distribuição normal (Verneque, 1994).

[REDACTED]

O método da Máxima Verossimilhança para estimação de componentes de variância foi desenvolvido por Hartley e Rao (1967). O método consiste em maximizar a função densidade de probabilidade das observações, em relação aos efeitos e aos componentes de variância dos efeitos aleatórios considerados no modelo, pressupondo normalidade. Esse método é iterativo e fornece estimativas não negativas para os componentes de variância, bem como elimina o viés atribuído às mudanças nas frequências gênicas, decorrente da seleção, caso o parentesco entre os indivíduos for considerado.

Embora os estimadores de máxima verossimilhança apresentem propriedades estatísticas ótimas, estes apresentam viés, por não considerar a perda de graus de liberdade, resultante da estimação dos efeitos fixos. A perda dos graus de liberdade faz com que os estimadores dos componentes variância obtidos, pelo método, em geral, não coincidam com aqueles estimados pelo método do tipo ANAVA, mesmo quando os dados são balanceados.

O viés presente com o método da máxima verossimilhança pode ser removido, considerando-se apenas a parte da função de verossimilhança que independe dos efeitos fixos. Este procedimento, pelo qual cada observação é dividida em duas partes independentes, uma referente aos efeitos fixos e outra aos efeitos aleatórios, foi sugerido por Thompson (1962). Mas foram Patterson e Thompson (1971) que desenvolveram o método para modelos mistos balanceados ou não, o qual foi denominado de Máxima Verossimilhança Restrita (REML). Para dados balanceados os componentes de variância são os mesmos encontrados pelos métodos do tipo ANAVA.

Para a derivação dos estimadores REML, o procedimento requer que o vetor de observações y tenha distribuição normal, e é utilizada a função de verossimilhança de $L'y$ (função de verossimilhança restrita), um conjunto de contrastes de erros com esperança nula, que representa efetivamente as observações ajustadas para os efeitos fixos.

Nos últimos anos, principalmente devido aos avanços da computação, tem-se dado preferência ao método REML, quando se trata de dados desbalanceados (Verneque, 1994). As razões para esta preferência baseiam-se nas propriedades estatísticas desejáveis que os estimadores REML apresentam, a maioria das quais, comuns ao método ML. Algumas dessas propriedades são:

1. os estimadores são funções estatísticas suficientes, consistentes e assintoticamente normais e eficientes;
2. as aproximações são sempre bem definidas;
3. o uso da restrição de não negatividade dos componentes de variâncias ou de outras restrições no espaço dos parâmetros, não causam dificuldades conceituais na aplicação do método;
4. as soluções pertencem quase sempre ao espaço dos parâmetros;
5. variâncias amostrais e erro quadrático médio menores do que aqueles obtidos por outros métodos que produzem estimadores não viciados;
6. o método REML reduz ou elimina o viés de pequenas amostras e/ou decorrente da seleção.

A desvantagem do método REML, que é também uma desvantagem adicional do método ML, é o fato de o procedimento de procura do ponto de máximo da função de verossimilhança em geral requerer esquemas iterativos,

e a convergência, além de ser lenta, não é garantida. Todavia, o avanço considerável na área da informática, bem como a adaptação e implementação de eficientes algoritmos computacionais, têm permitido um uso generalizado deste método na área de melhoramento.

Durante a última década, o método REML tem-se tornado o método preferido para estimação de componentes de variâncias e covariâncias, principalmente em melhoramento animal, onde os dados são, via de regra, desbalanceados. Esta escolha deve-se, não apenas às propriedades estatísticas desejáveis que o método apresenta, mas também à grande evolução dos recursos computacionais disponíveis, ao desenvolvimento e adaptação de algoritmos especializados, explorando características específicas da estrutura dos dados ou do modelo de análise, e à utilização de uma variedade de técnicas numéricas (Verneque, 1994).

A utilização de componentes de variância pelo método REML, para dados não balanceados, geralmente requer esquemas iterativos (Searle et al., 1992). A análise numérica geralmente classifica os procedimentos para encontrar o ponto de ótimo (máximo ou mínimo) de uma função, conforme a quantidade de informação requerida de suas derivadas (Verneque, 1994). Há basicamente três grupos: os métodos denominados de Newton, que utilizam primeira e segunda derivadas, métodos que utilizam apenas a primeira derivada e, finalmente, os métodos livres de derivadas, que envolvem estratégias de procura direta ou aproximação numérica de derivadas.

Até recentemente, os principais algoritmos empregados na área de melhoramento faziam uso das duas primeiras categorias. Os algoritmos principais delas são, respectivamente, o método do Escore de Fisher e o de Newton-Raphson, que requerem os valores esperados da primeira e segunda

derivadas do logaritmo da função de verossimilhança. Os métodos de Newton apresentam a vantagem de convergência regular, com um menor número de iterações; entretanto, exigem a inversão de matrizes de dimensões elevadas, o que, em alguns casos, torna a sua utilização restrita (Verneque, 1994).

Dentre os algoritmos que utilizam apenas informações da primeira derivada, destaca-se o algoritmo Maximização da Esperança (EM-REML) descrito por Dempster et al. (1977). Pela facilidade de se impor restrições às soluções e também pelo menor requerimento computacional frente aos procedimentos de Newton, quando se utiliza o algoritmo EM, o uso deste é defendido por vários autores, entre eles Henderson, (1984 e 1986) e Laird et al. (1987).

Por outro lado, a grande crítica que se faz ao uso do EM-REML, é que ele pode apresentar convergência lenta, mesmo nas situações em que outros algoritmos convergem rapidamente (Verneque, 1994). A razão para isto é que o algoritmo EM é um método de substituições sucessivas de primeira ordem e exibe convergência linear ao término das iterações (Laird et al., 1987). Isto pode ocorrer especialmente quando os componentes de variância são próximos de zero, quando a matriz de variâncias e covariâncias " V " tende a tornar-se não positiva definida. Igualmente em modelos com poucas observações e muitos fatores aleatórios, podem ocorrer problemas de convergência. No entanto, a convergência normalmente ocorrerá, se o número de observações for grande, e esta corresponderá ao máximo da função de verossimilhança (Verneque, 1994).

A máxima verossimilhança integrada VEIL, proposto por Gianola e Foulley (1990), baseia-se em métodos bayesianos. Através deste método, os

componentes de variâncias e covariâncias são estimados a partir de verossimilhanças integradas.

Pelo método da máxima verossimilhança ML, as fórmulas são derivadas a partir da função de verossimilhança de y . No método REML, as fórmulas são derivadas a partir da função de verossimilhança de um vetor de contrastes livres dos efeitos fixos. Na terminologia bayesiana, a máxima verossimilhança é derivada da maximização da função densidade conjunta a posteriori dos efeitos fixos e dos componentes de variância e covariâncias de y . Se a função de densidade conjunta for integrada em relação aos efeitos fixos e em seguida a função marginal for maximizada em relação aos componentes de variância e covariâncias, os estimadores resultantes serão os estimadores REML.

O método VEIL utiliza o mesmo conceito, um passo a frente, integrando a verossimilhança marginal em relação a cada componente da matriz V , assumindo que outros valores destes componentes de variância tenham sido estimados. As fórmulas resultantes são similares às obtidas pelos métodos ML e REML; a diferença básica é que o método VEIL considera os graus de liberdade utilizados para estimar os efeitos fixos e cada um dos componentes de variância. Além disso, o uso de procedimentos bayesianos permite ainda que informações passadas sobre os componentes de variância sejam levadas em conta na análise.

3 MATERIAL E MÉTODOS

A utilização da melhor predição linear não tendenciosa (BLUP) foi avaliada mediante dois critérios: pela comparação com o método dos quadrados mínimos ordinário (OLS) na predição das capacidades de combinação, e pela eficiência do BLUP em prever a capacidade específica de combinação e o valor genotípico de cruzamentos não realizados. Em ambos os critérios, dados simulados foram utilizados, sendo que no segundo também foram utilizados dados reais.

3.1 Dados simulados

Uma população de 101 linhagens totalmente endogâmicas foi simulada, considerando 200 locos quantitativos (QTL) e 100 locos marcadores de natureza dominante, todos de distribuição independente. A finalidade dos marcadores foi a de simular um processo de quantificação de similaridades genéticas entre as linhagens.

Os efeitos dos QTL (diferença entre os dois homozigotos) foram gerados a partir de uma função densidade probabilidade exponencial, com média igual a 1. Esta função caracteriza-se por apresentar frequências mais elevada de valores com menor efeito. A distribuição de frequência dos efeitos gerados dos 200 QTL pode ser visualizada na Figura 1.

A distribuição dos alelos para composição dos genótipos das linhagens foi feita de forma sistemática, de tal forma que foi atribuído para a primeira linhagem os 100 primeiros locos (os de maior valor), com alelos

favoráveis fixados, sendo os demais desfavoráveis; para a segunda linhagem, os locos com alelo favorável foram também em número de 100, a partir do segundo de maior efeito; e assim por diante, sendo que a última linhagem apresentava alelos favoráveis fixados para os 100 locos com menor efeito. Este critério de distribuição foi adotado objetivando que:

1. cada linhagem apresentasse 50% dos locos com o alelo favorável fixado;
2. não existissem linhagens com o mesmo valor genotípico ou reprodutivo;
3. a proporção de locos com alelos idênticos entre pares de linhagens variasse de 0 a 99%.

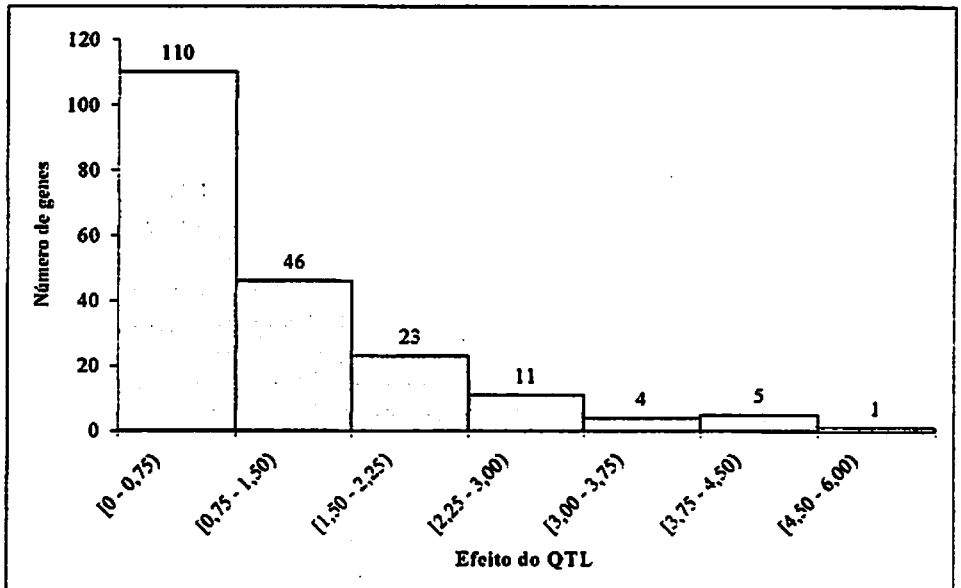


FIGURA 1. Distribuição de frequência absoluta dos valores atribuídos aos 200 genes, gerados pela função densidade de probabilidade exponencial.

A distribuição dos alelos marcadores foi feita tomando-se como referência a similaridade alélica existente no tocante aos QTL. Para isso, foi calculada a similaridade genética (proporção de locos com alelos idênticos) para todos os pares de linhagens. A composição dos genótipos marcadores das linhagens foi feita como descrito a seguir.

Considerando as linhagens ordenadas como descrito na distribuição dos QTL, para cada loco marcador da primeira linhagem, foi sorteado um alelo podendo assumir os valores 0 (ausência) ou 1 (presença) do alelo marcador com probabilidade $\frac{1}{2}$ para cada. Para este mesmo loco, nas demais linhagens, o alelo foi definido com base na similaridade genética entre as linhagens, sendo que a probabilidade de o alelo ser idêntico ao sorteado para a primeira linhagem foi igual ao valor da similaridade entre esta linhagem e a outra em questão.

Após definidos todos os genótipos marcadores, foi calculada a similaridade genética com base nos marcadores, simulando o que seria feito em uma situação real. Para isso, utilizou-se o coeficiente de similaridade de Sorensen-Dice, o qual, segundo Duarte (1998), é apropriado para marcadores de natureza dominante, como simulado no presente estudo:

$$S_{ij} = \frac{2a}{(2a + b + c)} \quad (11)$$

sendo S_{ij} o valor de similaridade alélica entre as linhagens i e j , respectivamente; a o número de presenças simultâneas do alelo marcador dominante nas linhagens i e j ; b e c o número de presenças do alelo marcador dominante em apenas uma das linhagens consideradas.

Após definido o genótipo das 101 linhagens, estas foram cruzadas em esquema dialélico, considerando três tipos de interações alélicas no tocante

aos QTL: ausência de dominância, grau de dominância 0,5 para todos os locos e dominância completa. Para cada uma das situações de interação alélica (graus de dominância) consideradas, foi obtida uma população de referência contendo todos os 5050 cruzamentos e as 101 linhagens.

Conforme definido originalmente por Sprague e Tatum (1942), o efeito da capacidade geral de combinação (CGC) foi considerado como a diferença entre a média dos cruzamentos de um indivíduo e a média da população. A capacidade específica de combinação (CEC) correspondeu ao desvio do valor de cada cruzamento, em relação ao esperado, baseando-se nas capacidades gerais de combinação de cada genitor. Este valores de CGC e CEC foram calculados para as três populações de referência, sendo tomados como valores paramétricos, uma vez que foram calculados sem erro.

3.2 Dados experimentais

Para avaliação da eficiência do BLUP na predição de cruzamentos não realizados, foram utilizados dados experimentais, referentes à produtividade (kg/ha), de um dialelo, envolvendo 22 linhagens de feijão comum, conduzido por Abreu (1997). Devido a problemas de incompatibilidade entre as diversas linhagens, foram realizados apenas 120 dos 253 cruzamentos possíveis. Os cruzamentos foram avaliados em 3 locais do Estado de Minas Gerais, em látices simples. Detalhes sobre as características das linhagens e também da condução dos experimentos foram descritos em Abreu (1997).

Os dados de similaridade genética entre as 22 linhagens utilizadas neste estudo foram obtidos por Duarte (1998), através de marcadores moleculares do tipo RAPD. Foram utilizados 33 iniciadores (primers), que

geraram 265 produtos de amplificação (bandas), das quais, 137 apresentaram polimorfismo. Os valores de similaridade genética foram calculados a partir do coeficiente de Sorensen-Dice, conforme recomendado pelo autor.

3.3 Análise dialélica

Dois enfoques de análise dialélica foram empregados, utilizando o método dos quadrados mínimos ordinário (OLS) e a melhor predição linear não tendenciosa (BLUP).

Com o método dos quadrados mínimos ordinário, as estimativas dos efeitos fixos e dos efeitos das capacidades de combinação, foram obtidas a partir das expressões propostas por Griffing (1956), e apresentadas por Cruz e Regazzi (1994).

Com o método BLUP, as estimativas dos efeitos fixos e as predições das capacidades de combinação foram obtidas de maneira semelhante à apresentada por Bernardo (1995), em cruzamentos dialélicos parciais.

O modelo linear considerado foi dado por:

$$Y = X\beta + Z_1g + Z_2s + e$$

sendo

Y : o vetor de observações;

X : a matriz de incidência dos efeitos fixos;

β : o vetor de efeitos fixos (efeitos de bloco, local e média geral);

Z_1 : a matriz de incidência dos efeitos da capacidade geral de combinação;

g : o vetor de efeitos da capacidade geral de combinação;

Z_2 : a matriz de incidência dos efeitos da capacidade específica de combinação;

s : o vetor de efeitos da capacidade específica de combinação;

e : o vetor de resíduos.

Admitiu-se que:

$$\begin{bmatrix} Y \\ g \\ s \\ e \end{bmatrix} \sim N \left(\begin{bmatrix} X\beta \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_{i=1}^2 Z_i G_i Z_i' + R & Z_1 G_1 & Z_2 G_2 & R \\ G_1 Z_1' & G_1 & \phi & \phi \\ G_2 Z_2' & \phi & G_2 & \phi \\ R & \phi & \phi & R \end{bmatrix} \right)$$

em que

0 é um vetor nulo;

ϕ é uma matriz nula;

$R = I\sigma^2$ a matriz de variâncias e covariâncias residual, com I uma matriz identidade e σ^2 a variância residual.

$G_1 = A_1\sigma_g^2$, a matriz de variâncias e covariâncias das capacidades gerais de combinação das linhagens, com A_1 a matriz de similaridade genética aditiva entre as linhagens e σ_g^2 o componente de variância da capacidade geral de combinação. Os elementos da matriz A_1 na diagonal são iguais à unidade e fora da diagonal igual aos valores do coeficiente de similaridade de Sorense-Dice, calculados conforme (11);

$G_2 = A_2\sigma_s^2$, a matriz de variâncias e covariâncias das capacidades específicas de combinação das linhagens, com A_2 a matriz de similaridade

genética de dominância entre os cruzamentos e σ_s^2 o componente de variância da capacidade específica de combinação. Os elementos da matriz A_2 na diagonal são iguais à unidade e fora da diagonal iguais ao valor definido como coeficiente de dominância ($k_{(ij) \times (lm)}$) de acordo com Henderson (1984) e Van Vleck (1993b), como descrito a seguir.

Henderson (1984) e Van Vleck (1993b) definiram o coeficiente de dominância como sendo a proporção de genótipos idênticos ao longo de todos os locos, entre dois indivíduos relacionados geneticamente. Considerando dois híbridos H_{ij} e H_{lm} , resultantes dos cruzamentos das linhagens ($L_i \times L_j$) e ($L_l \times L_m$), respectivamente, se essas quatro linhagens são relacionadas, então o coeficiente de dominância $k_{(ij) \times (lm)}$ entre os híbridos H_{ij} e H_{lm} é dado por:

$$k_{(ij) \times (lm)} = \frac{1}{4} [S_{ii} \times S_{jm} + S_{im} \times S_{jl}] \quad (12)$$

A solução conjunta para os efeitos fixos e aleatórios, foi obtida pelo sistema de equações (Henderson, 1984):

$$\begin{bmatrix} \beta^0 \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \quad (13)$$

ou

$$\begin{bmatrix} \beta^0 \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix}$$

$$\text{com } \gamma_1 = \frac{\hat{\sigma}^2}{\hat{\sigma}_g^2} \text{ e } \gamma_2 = \frac{\hat{\sigma}^2}{\hat{\sigma}_s^2}.$$

As estimativas dos componentes de variância foram obtidas pelo método da máxima verossimilhança restrita (REML), utilizando o algoritmo de maximização da esperança (EM-REML), proposto por Dempster et al. (1977) e apresentado por Henderson (1984).

As estimativas dos componentes de variância em uma determinada iteração foram obtidas por:

$$\hat{\sigma}^2 = \frac{Y'Y - \beta'X'y - \hat{g}'Z_1'y - \hat{s}'Z_2'y}{N - r(X)} \quad (14a)$$

$$\hat{\sigma}_g^2 = \frac{\hat{g}'A_1^{-1}\hat{g} + tr(A_1^{-1}C_{11})\hat{\sigma}^2}{q_1} \quad (14b)$$

$$\hat{\sigma}_s^2 = \frac{\hat{s}'A_2^{-1}\hat{s} + tr(A_2^{-1}C_{22})\hat{\sigma}^2}{q_2} \quad (14c)$$

sendo N o número de observações, $r(X)$ o posto da matriz X , $tr(\cdot)$ o operador traço matricial, q_1 o número de linhagens e q_2 o número de híbridos avaliados.

Deve-se ressaltar que não existem soluções explícitas para o sistema apresentado em (13), dado que as estimativas dos efeitos fixos e as predições das capacidade de combinação dependem dos valores atribuídos aos componentes de variância, e estes, por sua vez, como observado em (14), dependem das estimativas dos efeitos fixos e das predições dos efeitos aleatórios.

A solução para o sistema consistiu em atribuir valores iniciais para os componentes de variâncias e obter uma solução para os efeitos fixos e aleatórios. Com esta solução, eram recalculados os componentes de variância, que eram utilizados em (13), obtendo uma nova solução para os efeitos fixos e aleatórios. O processo era assim continuado até que não ocorressem mudanças significativas nas estimativas dos componentes de variância.

As estimativas dos efeitos do modelo linear, bem como dos componentes de variância, permitiram a predição de cruzamentos não realizados, quando este fosse o caso.

Sendo \hat{Y}_A o vetor das predições observadas para conjunto de híbridos avaliados em um determinado experimento, tem-se que \hat{Y}_A pode ser representado pelo seguinte modelo linear:

$$\hat{Y}_A = \hat{\mu} + Z_{1A}\hat{g} + Z_{2A}\hat{s}_A$$

sendo $\hat{\mu}$ a estimativa da média geral, Z_{1A} a matriz de incidência dos efeitos da CGC, \hat{g} o vetor de predições da CGC, Z_{2A} a matriz de incidência dos efeitos da CEC, e \hat{s}_A o vetor das predições da CEC.

A predição do desempenho de um conjunto de híbridos não avaliados (\hat{Y}_P), foi feita por:

$$\hat{Y}_P = C_{1AP}G_{AA}^{-1}\hat{Y}_A \quad (15)$$

na qual C_{1AP} é a matriz de covariâncias genéticas entre os híbridos avaliados e os híbridos com desempenho a ser predito, G_{AA} é a matriz de variâncias e

covariâncias genéticas dos híbridos avaliados e \hat{Y}_A o vetor das predições dos híbridos avaliados.

Os valores dos elementos de C_{1AP} e G_{AA} foram calculados por:

$$C_{1AP} = Z_{1P}G_{1P}Z'_{1A} + Z_{2P}G_{2P}Z'_{2A} \text{ e,}$$

$$G_{AA} = Z_{1A}G_1Z'_{1A} + Z_{2A}G_2Z'_{2A}$$

com Z_{1P} e Z_{2P} as matrizes de incidência das capacidades geral e específica de combinação dos híbridos com desempenho a ser predito, respectivamente, G_{1P} e G_{2P} as matriz de covariâncias genéticas das capacidades gerais e específicas de combinação entre os híbridos avaliados e os com desempenho a ser predito, calculados de acordo com (11) e (12), respectivamente, G_1 e G_2 definidos anteriormente, tomando-se as estimativas dos componentes de variâncias no lugar dos parâmetros.

De maneira análoga a (15), os valores preditos das capacidades específicas de combinação dos híbridos não avaliados foram obtidos mediante:

$$\hat{s}_A = C_{2AP}G_2^{-1}\hat{s}_P$$

em que C_{2AP} é a matriz de covariâncias genéticas das capacidades específicas de combinação entre os híbridos avaliados e os com desempenho a ser predito, calculados de acordo com (12).

3.4 Comparação entre os métodos BLUP e OLS

A partir das populações de referência descritas anteriormente, foram tomados sub-dialelos, considerando os métodos II e IV de Griffing, a partir de 12 linhagens amostradas ao acaso. A todos os cruzamentos dos sub-

dialelos, tomados em três repetições, foi adicionado um efeito aleatório residual com distribuição normal de média 0 (zero) e variância σ^2 . Os valores de σ^2 foram gerados de tal forma que fossem simuladas situações de herdabilidade ao nível de médias igual a 0,3, 0,5 e 0,7.

Ao todo foram simuladas 18 situações distintas, envolvendo três graus de dominância (ausência de dominância, dominância 0,5 e dominância completa), dois métodos dialélicos (Griffing II e IV) e três níveis de herdabilidade (0,3, 0,5 e 0,7). Para cada situação simulada, foram gerados 300 dialelos, que foram analisados segundo os métodos BLUP e OLS.

A eficiência das metodologias em questão foi avaliada segundo os seguintes critérios:

Viés médio (VM). A estimativa do viés médio foi tomada como a média, em módulo, dos desvios observados entre o valor paramétrico e a média das predições para o mesmo parâmetro, dado por:

$$VM = \frac{1}{I} \sum_{i=1}^I \left| \frac{1}{J_i} \sum_{j=1}^{J_i} (\hat{\theta}_{ij} - \theta_i) \right| \quad (i=1, 2, \dots, I \text{ e } j=1, 2, \dots, J_i)$$

sendo θ_i o parâmetro i (efeito de CEC, CGC ou de cruzamento); I o número de parâmetros a serem preditos; J_i o número de vezes (dialelos) em que o parâmetro i foi predito; e $\hat{\theta}_{ij}$ a j -ésima predição de θ_i .

Erro médio (EM). A estimativa do erro médio foi tomada como a média dos módulos dos desvios observados entre o valor paramétrico e a média das predições para o mesmo parâmetro, dado por:

$$EM = \frac{1}{I} \sum_{i=1}^I \frac{1}{J_i} \sum_{j=1}^{J_i} |\hat{\theta}_{ij} - \theta_i|$$

Erro quadrático médio (EQM). A estimativa do erro quadrático médio foi tomada como a média dos quadrados dos desvios observados entre o valor paramétrico e o predito do mesmo parâmetro, dado por:

$$EQM = \frac{1}{I} \sum_{i=1}^I \frac{1}{J_i} \sum_{j=1}^{J_i} (\hat{\theta}_{ij} - \theta_i)^2$$

Correlações. A partir das predições de CGC, CEC e valor de cruzamentos obtidas pelas duas metodologias, em todos os dialelos amostrados, foram estimados os valores dos coeficientes de correlação simples (Pearson) e classificatória (Spearman), entre os valores preditos e os paramétricos (Steel e Torrie, 1981), para verificar semelhanças entre tais valores, tanto em magnitude como no ordenamento.

3.5 Predição de cruzamentos não avaliados

A eficiência da melhor predição linear não tendenciosa (BLUP) na predição de cruzamentos não realizados foi avaliada considerando os dados simulados no item 3.1, e os dados experimentais descritos no item 3.2

3.5.1 Dados simulados

A partir das populações de referência descritas no item 3.1, foram tomados sub-dialelos envolvendo 15 linhagens tomadas ao acaso, considerando o método IV de Griffing. Em todos os cruzamentos dos dialelos, tomados em três repetições, foi adicionado um efeito aleatório residual com distribuição normal de média 0 (zero) e variância σ^2 . Os valores de σ^2 foram gerados de tal forma que fossem simuladas situações de herdabilidade ao nível de médias igual a 0,3, 0,5 e 0,7.

Da mesma maneira, como feito na comparação entre BLUP e OLS, ao todo foram simuladas 6 situações distintas, envolvendo três graus de dominância (dominância 0,5 e dominância completa), e três níveis de herdabilidade (0,3, 0,5 e 0,7). Para cada situação simulada, foram gerados 100 dialelos, que foram analisados segundo o método BLUP.

Após a análise de cada sub-dialelo completo, foram eliminados 30 cruzamentos, simulando o arranjo de um dialelo parcial circulante, em que cada linhagem era cruzada com outras 10. As informações das 75 combinações híbridas serviram de base para a predição dos valores da capacidade específica de combinação e do valor dos 30 híbridos retirados.

Em seguida, foram estimadas as correlações simples e classificatória, entre o valor predito, quando o híbrido não foi avaliado, com os valores paramétricos e também com os obtidos da análise pelo BLUP com o sub-dialelo completo, sem perda de cruzamentos.

3.5.2 Dados experimentais

A eficiência da melhor predição linear não tendenciosa (BLUP) também foi avaliada, considerando os dados experimentais descritos no item 3.2. Para tanto, a predição dos efeitos da CEC e do valor genotípico dos cruzamentos foi feita considerando duas situações. Na primeira, foi retirado um dos híbridos e, com a informação obtida da análise envolvendo os 119 restantes, foi predito o valor da CEC e do cruzamento eliminado. Este processo foi repetido em todas as 120 combinações híbridas. Na segunda situação, foram retirados grupos de 10, 20, 30, 40, 50 e 60 híbridos e, com o restante realizou-se as predições do valor da CEC e do cruzamento referentes

aos híbridos retirados da análise. Este processo foi repetido 100 vezes, para cada grupo considerado.

A eficiência do BLUP na predição de valores de CEC e dos valores genotípicos dos cruzamentos foi avaliada através do cálculo das correlações de Pearson e de Spearman, entre os valores preditos considerando a retirada de híbridos e aqueles não considerando tal retirada.

Finalmente, foi ainda realizada a predição do desempenho e da capacidade específica de combinação dos 133 híbridos que, originalmente, não puderam ser obtidos por Abreu (1997), devido a problemas de incompatibilidade.

4 RESULTADOS E DISCUSSÃO

4.1 Comparação entre os métodos BLUP e OLS

4.1.1 *Ausência de dominância*

Na Tabela 1 são apresentados os valores das estimativas de erro quadrático médio, erro médio e viés médio, referentes às predições das capacidades geral e específica de combinação e ao valor dos cruzamentos de dialelos ao longo de 300 simulações, obtidas pelas metodologias de análise BLUP e OLS, considerando os métodos dialélicos Griffing II e IV, três níveis de herdabilidade e ausência de efeitos de dominância. Também são apresentados na Tabela 1, as estimativas daqueles mesmos parâmetros, no tocante às capacidades de combinação, obtidas sem efeito ambiental, ou seja, refletindo as variações resultantes do processo de amostragem dos genitores, na simulação de dialelos.

Pode-se observar a não existência de diferenças marcantes quando se compara os dois métodos de Griffing. Entretanto, há uma pequena tendência de o método II apresentar uma maior precisão em suas estimativas, à exceção das estimativas de capacidade específica, com os quais os valores de viés médio, erro médio e erro quadrático médio foram inferiores com o método IV. Estes resultados sugerem que não haja diferenças de eficiência substanciais, no processo de predição, ao se incluir ou não os genitores na análise dialélica, ao menos considerando modelos aleatórios, como no presente estudo (amostrando genitores de uma população).

TABELA 1. Erro quadrático médio (EQM), erro médio e viés médio das predições de efeitos em 300 dialelos simulados com 12 pais, tomados de uma população de 101 genitores, desconsiderando efeitos ambientais (variações apenas de amostragem), e considerando-os em associação com os métodos BLUP e de quadrados mínimos ordinário (OLS) em ausência de dominância, e utilizando os métodos II e IV de Griffing.

	EQM		Erro Médio		Viés Médio	
	Método II	Método IV	Método II	Método IV	Método II	Método IV
Amostragem ($h^2 = 1,0$)						
CEC	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CGC	86.5705	89.3042	7.4380	7.5923	2.3405	2.3916
$h^2 = 0,3$						
Capacidade Específica de Combinação						
BLUP	0,0001	0,0001	0,0089	0,0086	0,0059	0,0056
OLS	5035,6200	4725,0000	56,6583	54,7299	31,0710	30,2433
Capacidade Geral de Combinação						
BLUP	241,2830	282,8090	12,2987	13,6060	5,4989	6,5341
OLS	473,0670	598,9440	17,3506	19,7747	3,4072	3,9447
Cruzamento						
BLUP	443,0510	515,1780	16,6626	17,9772	10,7132	12,1686
OLS	5834,3300	5720,4600	61,0191	60,1258	33,2835	33,2764
$h^2 = 0,5$						
Capacidade Específica de Combinação						
BLUP	0,0003	0,0002	0,0127	0,0125	0,0080	0,0081
OLS	2097,0700	2009,0000	36,5708	35,8249	19,7983	19,7840
Capacidade Geral de Combinação						
BLUP	158,2630	178,9580	9,9184	10,6122	3,6940	4,5394
OLS	253,4720	314,3690	12,6890	14,1042	2,7833	3,3688
Cruzamento						
BLUP	240,4580	289,4620	12,2899	13,4919	7,5415	8,7102
OLS	2431,6200	2470,8300	39,3415	39,7224	21,3107	22,0499
$h^2 = 0,7$						
Capacidade Específica de Combinação						
BLUP	0,0006	0,0005	0,0187	0,0184	0,0113	0,0112
OLS	908,4280	872,5480	24,0522	23,5661	13,0408	13,1056
Capacidade Geral de Combinação						
BLUP	109,7070	124,6330	8,3235	8,9235	3,1335	3,1830
OLS	146,9320	185,0650	9,6885	10,8495	2,4068	2,5816
Cruzamento						
BLUP	119,3160	149,0700	8,7073	9,7101	5,1179	5,9960
OLS	1047,5600	1065,6900	25,7066	25,9830	13,9840	14,4963

Na Tabela 1, pode-se observar que o erro quadrático médio, o erro médio e o viés médio das capacidades de combinação foram sempre menores, quando não foram considerados efeitos ambientais, em relação ao uso dos métodos OLS e BLUP, em que tais efeitos foram considerados. Aqueles valores refletem essencialmente o efeito da amostragem de genitores, para compor os dialelos simulados, e podem ser tomados como valores mínimos. Na ausência de dominância e de efeitos ambientais, a CEC é nula, e isso resultou nos valores também nulos de amostragem.

O método BLUP apresentou um desempenho marcadamente superior em todas as situações simuladas, na predição da capacidade específica (CEC) e do valor do cruzamento, sendo que as estimativas do viés médio observadas, para a capacidade geral de combinação (CGC) das linhagens, foram superiores às observadas pelo método OLS. Aparentemente, houve uma tendência de subestimação dos valores das predições da CGC, ao se utilizar o BLUP (dados não apresentados), resultando neste viés médio maior. Deve-se enfatizar, contudo, que foram as menores diferenças em relação aos valores de amostragem, levando em conta o erro quadrático médio e o erro médio. Além disso, uma vez que o BLUP apresentou precisão marcadamente superior na predição da CGC, este método deve ser preferido, apesar do viés médio maior, à semelhança de outros métodos estatísticos, como o da regressão em cristas (Draper e Smith, 1981). De qualquer maneira, se o objetivo principal da análise dialélica for o da predição de cruzamentos, o BLUP foi superior quanto a todos os parâmetros (Tabela 1).

As estimativas do viés médio e erro médio, para o valor do cruzamento pelo método BLUP, foram, em média, 3 vezes menores que as do

OLS, enquanto que, quanto ao erro quadrático médio, as estimativas do BLUP foram em média 9 vezes inferiores às do método OLS.

Estes aspectos de maior precisão com o BLUP, de uma maneira geral, foram coerentes com o critério de derivação do método, ou seja, o de minimizar o erro quadrático médio das predições, gerando as melhores predições lineares não viesadas.

Como era esperado, a eficiência das metodologias de análise aumenta com o acréscimo da herdabilidade, em virtude do aumento de precisão, decorrente da redução relativa da variação ambiental. Foram observadas reduções de aproximadamente 57%, nas estimativas do viés médio e erro médio; e mais de 80%, no erro quadrático médio para a CEC, pelo método OLS. Reduções de mesma magnitude nas estimativas de viés médio, erro médio e erro quadrático médio também foram observadas para as predições do valor do cruzamento pelo método OLS. Com o BLUP a redução nas estimativas do viés médio e erro médio, com o aumento da herdabilidade de 0,3 para 0,7, foi de aproximadamente 50 e de 70%, para o erro quadrático médio.

As médias dos valores de correlação entre os valores de cruzamento e CGC e suas respectivas predições, acompanhadas de seus erros padrão, estimados a partir da variação ao longo das simulações, para as duas metodologias de análise (BLUP e OLS), são apresentadas na Tabela 2.

TABELA 2. Média e erro padrão (entre parênteses), da correlação simples (Pearson) entre o valor genético e o predito pelos métodos BLUP e OLS em ausência de dominância, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0,3$				
CGC	0,9187 (0,049)	0,9093 (0,064)	0,8570 (0,085)	0,8208 (0,109)
Cruzamento	0,9186 (0,049)	0,9093 (0,064)	0,5338 (0,105)	0,5086 (0,117)
$h^2 = 0,5$				
CGC	0,9538 (0,027)	0,9420 (0,039)	0,9352 (0,042)	0,9084 (0,061)
Cruzamento	0,9538 (0,027)	0,9420 (0,039)	0,7055 (0,084)	0,6623 (0,094)
$h^2 = 0,7$				
CGC	0,9755 (0,014)	0,9681 (0,020)	0,9696 (0,022)	0,9586 (0,033)
Cruzamento	0,9755 (0,014)	0,9681 (0,020)	0,8260 (0,068)	0,8042 (0,073)

Observa-se diferenças pouco marcantes entre as estimativas, considerando os métodos II e IV de Griffing, embora possa ser vista uma ligeira superioridade com o método II, especialmente com o OLS. Além disso, o método II apresentou erros padrão menores em todos os casos.

Os valores médios de correlação entre o valor paramétrico e o predito, para os valores de CGC e de cruzamento obtidos pelo BLUP, foram sempre mais elevados e com menor erro padrão que aqueles do método dos quadrados mínimos ordinário. As diferenças entre os dois métodos foram mais marcantes, no tocante à predição dos cruzamentos, onde, com herdabilidade igual a 0,3, os valores médios de correlação para o método BLUP foram 72,09 e 78,78% maiores que os do OLS, para os métodos II e IV de Griffing, respectivamente; e estas diminuem, à medida que se aumenta o valor da herdabilidade, passando para 18,10 e 20,38%, com a herdabilidade igual a 0,7.

Nejati-Javaremi et al. (1997), em estudo simulado da utilização do BLUP com informações de similaridade alélica, através de marcadores moleculares, e do método dos quadrados mínimos ordinário, observaram o mesmo comportamento para os valores da correlação, em função da variação do valor da herdabilidade, ou seja, em situações de baixa herdabilidade ($h^2 = 0,1$), as diferenças entre as duas metodologias foram máximas. Um fato ressaltado pelos autores, foi o de que a correlação entre o valor predito e o genético com o método OLS, é função apenas da herdabilidade da característica, enquanto que o BLUP faz uso de informações adicionais de similaridade genética, entre os indivíduos avaliados, aumentando assim a precisão das suas predições.

Com o método OLS, o valor médio das correlações entre o valor genético do cruzamento e o predito foi sempre inferior, embora sempre próximo à raiz quadrada das herdabilidades ao nível de médias de cruzamento (iguais a 0,5477, 0,7071 e 0,8367), que é o valor de referência no caso do método OLS, correspondendo à correlação entre o fenótipo e o genótipo. Com o método BLUP, é interessante notar que as correlações foram maiores que estes valores de referência, sendo resultado do uso da informação de similaridade genética entre os cruzamentos. As estimativas de correlação para a CGC foram as mais elevadas e com menor variação em torno da média, enquanto que, para o valor do cruzamento, foram observados menores valores médios e com uma maior variação (maiores erros padrão).

Os valores médios da correlação classificatória de Spearman para as 300 simulações, em ausência de dominância considerando os métodos II e IV de Griffing, em três níveis de herdabilidade, estão apresentados na Tabela 3.

TABELA 3. Média e erro padrão (entre parênteses), da correlação classificatória (Spearman) entre o valor genético e o predito pelos métodos BLUP e OLS em ausência de dominância, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0,3$				
CGC	0,8889 (0,053)	0,8844 (0,072)	0,7947 (0,084)	0,7583 (0,072)
Cruzamento	0,9114 (0,054)	0,9031 (0,040)	0,5127 (0,092)	0,4908 (0,100)
$h^2 = 0,5$				
CGC	0,9169 (0,034)	0,9115 (0,062)	0,8741 (0,062)	0,8482 (0,083)
Cruzamento	0,9457 (0,034)	0,9337 (0,040)	0,6827 (0,075)	0,6456 (0,099)
$h^2 = 0,7$				
CGC	0,9414 (0,029)	0,9358 (0,059)	0,9186 (0,062)	0,9067 (0,082)
Cruzamento	0,9678 (0,030)	0,9601 (0,038)	0,8044 (0,080)	0,7883 (0,089)

Nota-se que o desempenho das duas metodologias de análise seguem o mesmo padrão observado pela correlação simples de Pearson. O método BLUP foi o mais eficiente no ordenamento da capacidade geral de combinação e dos cruzamentos avaliados. A diferença entre as duas metodologias foi máxima para a situação de baixa herdabilidade ($h^2 = 0,3$), sendo que a eficiência dos métodos de análise, como era esperado, aumentou com o valor da herdabilidade, em virtude do aumento de precisão decorrente da menor variação ambiental. O método BLUP apresentou ordenamento quanto à CGC e ao valor dos cruzamentos de qualidade superior à observada pelo método OLS, mesmo no nível mais elevado de herdabilidade estudado. Isto reforça o aspecto de que, ainda que o BLUP tenha apresentado viés médio superior ao do OLS (Tabela 1), este método deva ser o preferido na seleção de genitores (quando esta é baseada na CGC), em virtude de sua maior precisão acarretar um ordenamento das predições da CGC mais semelhante ao dos valores paramétricos.

Bueno Filho (1997) obteve resultados semelhantes, ao comparar os métodos BLUP e OLS, para a seleção de progênes de meios-irmãos de *E. grandis*. O autor observou valores de correlações entre as predições obtidas pelo BLUP e OLS, para quatro características avaliadas, superiores a 0,94. Apesar dessas correlações elevadas entre os valores genéticos preditos pelas duas metodologias, foram observadas diferenças de 15 a 25%, nas progênes selecionadas pelas duas metodologias ao se utilizar intensidades seletivas de 20 a 30%. Em relação às estimativas de ganho com a seleção, o BLUP foi sempre superior ao método OLS. Este fato, somado às melhores propriedades do método, faz com que o BLUP seja preferível ao OLS.

Na análise dialélica, pode-se dizer que, de uma forma geral os resultados do presente estudo apontam claramente a superioridade do BLUP, quando não ocorrem efeitos de dominância.

4.1.2 Presença de dominância

As estimativas de erro quadrático médio, erro médio e viés médio para a predição das capacidades de combinação e do valor dos cruzamentos, considerando os métodos II e IV de Griffing, em três níveis de herdabilidade, em situações de grau de dominância igual a 0,5 e dominância completa estão apresentadas nas Tabelas 4 e 5, respectivamente.

Novamente, pode-se observar, que as estimativas, considerando apenas as variações de amostragem, foram as menores, conforme o esperado. Aqui, ao contrário da situação de ausência de dominância, os valores para a CEC foram diferentes de zero, refletindo o fato de a CEC ser função da ocorrência de dominância.

TABELA 4. Erro quadrático médio (EQM), erro médio e viés médio das predições de efeitos em 300 dialelos simulados com 12 pais, tomados de uma população de 101 genitores, desconsiderando efeitos ambientais (variações apenas de amostragem), e considerando-os em associação com os métodos BLUP e de quadrados mínimos ordinário (OLS) em presença de grau dominância igual a 0,5, e utilizando os métodos II e IV de Griffing.

	EQM		Erro Médio		Viés Médio	
	Método II	Método IV	Método II	Método IV	Método II	Método IV
Amostragem ($h^2 = 1.0$)						
CEC	60,9737	61,4757	5,5590	5,7683	4,5986	5,0821
CGC	159,2897	170,8767	7,9556	9,9995	4,3533	4,2476
$h^2 = 0.3$						
Capacidade Específica de Combinação						
BLUP	535,3400	612,8840	17,1968	18,2820	14,0308	16,8476
OLS	7882,3600	7644,1100	70,7531	69,8059	38,8085	57,7671
Capacidade Geral de Combinação						
BLUP	525,3490	701,3770	17,6895	20,2822	11,7580	12,9279
OLS	712,6620	1075,9100	21,1515	26,3407	5,4187	9,6827
Cruzamento						
BLUP	1221,3300	1430,3500	27,6936	29,9695	20,5234	26,5053
OLS	8961,5300	9302,9000	75,4322	77,0489	41,2377	64,2329
$h^2 = 0.5$						
Capacidade Específica de Combinação						
BLUP	390,8450	349,2290	14,9834	14,4535	11,6259	13,7503
OLS	3418,3500	3119,4100	46,4573	44,3287	25,9946	36,4738
Capacidade Geral de Combinação						
BLUP	373,5030	440,1700	14,8196	16,3163	8,4648	9,4358
OLS	411,4890	549,1950	16,0273	18,6576	4,9909	7,1142
Cruzamento						
BLUP	732,8500	772,3610	21,4950	22,2194	15,3916	19,5391
OLS	3841,5100	3740,9200	49,2359	48,5766	26,8200	39,8955
$h^2 = 0.7$						
Capacidade Específica de Combinação						
BLUP	291,5010	281,5480	12,9764	13,1443	9,7825	12,0153
OLS	1523,7800	1399,9300	31,1199	29,8800	17,6792	24,5650
Capacidade Geral de Combinação						
BLUP	296,9630	300,9240	13,3306	13,5233	7,2746	7,1188
OLS	290,5860	299,4790	13,3794	13,7761	5,0050	4,7069
Cruzamento						
BLUP	461,9500	494,2100	17,0779	17,8162	12,0402	15,5667
OLS	1700,9800	1665,1200	32,8739	32,5919	18,0226	26,7481

TABELA 5. Erro quadrático médio (EQM), erro médio e viés médio das predições de efeitos em 300 dialelos simulados com 12 pais, tomados de uma população de 101 genitores, desconsiderando efeitos ambientais (variações apenas de amostragem), e considerando-os em associação com os métodos BLUP e de quadrados mínimos ordinário (OLS) em presença de dominância completa, e utilizando os métodos II e IV de Griffing.

	EQM		Erro Médio		Viés Médio	
	Método II	Método IV	Método II	Método IV	Método II	Método IV
Amostragem ($h^2 = 1.0$)						
CEC	238,2543	238,5820	11,0917	11,3808	9,1753	8,8636
CGC	261.2887	310.8047	11.4883	12.6879	6.1321	3.7507
$h^2 = 0,3$						
Capacidade Específica de Combinação						
BLUP	1658,1600	1638,4700	30,5653	30,8647	23,8385	26,1954
OLS	12912,2300	12690,0900	90,5841	90,1021	49,8015	61,8665
Capacidade Geral de Combinação						
BLUP	1051,7900	1092,3400	24,2878	24,8439	17,7755	16,5549
OLS	1278,4300	1577,6400	28,3663	31,5074	7,2109	7,4199
Cruzamento						
BLUP	2679,1400	2757,0400	41,0072	41,6818	30,7878	33,9032
OLS	14802,1500	15097,1800	97,1556	97,9194	52,5298	67,1338
$h^2 = 0,5$						
Capacidade Específica de Combinação						
BLUP	1230,2700	1556,5100	26,5433	26,2301	20,1619	22,2104
OLS	5715,6800	5500,4900	60,3118	59,2015	33,3098	40,6877
Capacidade Geral de Combinação						
BLUP	776,7870	779,4390	20,8161	21,1285	14,4076	12,8418
OLS	651,3940	882,7900	20,2315	23,6849	6,6378	6,0469
Cruzamento						
BLUP	1677,9100	1715,6800	32,5203	33,1050	23,8509	26,7110
OLS	6327,9800	6349,8300	63,4455	63,6463	34,1157	53,5606
$h^2 = 0,7$						
Capacidade Específica de Combinação						
BLUP	917,3040	852,2650	23,0742	22,6565	16,5011	19,3806
OLS	2626,8300	2444,7200	40,8477	39,4266	23,6051	27,5319
Capacidade Geral de Combinação						
BLUP	587,8690	605,2810	18,3245	18,7930	11,5666	9,9293
OLS	439,6840	567,9710	16,0498	18,7366	6,6048	4,3508
Cruzamento						
BLUP	1025,4000	1033,5400	25,5124	25,5369	17,4375	19,8798
OLS	2755,5000	2717,2200	42,0830	41,5976	22,8419	28,4974

Comparando as Tabelas 1, 4 e 5, pode ser observado que o aumento no grau de dominância diminui a eficiência das predições dos efeitos, em ambos os métodos dialélicos estudados. Nota-se claramente um aumento nas estimativas de viés médio, erro médio e erro quadrático médio, para todas as situações em que foi considerado o modelo com dominância completa. Esta tendência era esperada, uma vez que o erro de predição é função da variância genética, e esta, necessariamente, aumenta com a elevação do grau de dominância.

À semelhança da situação de ausência de dominância, há uma tendência de maior precisão das predições dos efeitos do dialelo quando se utiliza o método II, embora não muito marcante. Esta tendência é um pouco mais evidente com o erro quadrático médio, devido ao fato de este ser função de desvios ao quadrado.

Em termos da eficiência relativa do métodos BLUP e OLS, tendências muito semelhantes foram observadas, ao se considerar a presença de dominância. As estimativas de viés médio, erro médio e erro quadrático médio para o método BLUP, foram menores que as do OLS, à exceção das predições da capacidade geral de combinação. No entanto, novamente aqui, não se deve concluir pela superioridade do método OLS, em virtude da maior precisão ocorrida com o BLUP, resultando em menores estimativas de erro quadrático médio e erro médio (Tabelas 4 e 5).

O efeito da herdabilidade foi semelhante ao observado com o modelo sem a presença de dominância, ou seja, à medida que os níveis de herdabilidades se tornam mais elevados, há uma melhoria da eficiência das metodologias estatísticas. Em termos de estimativas do viés médio para a

predição da capacidade específica de combinação, observou-se uma redução de 29,41 e 56,26% com os métodos BLUP e OLS, respectivamente, quando a herdabilidade aumentou de 0,3 para 0,7, na situação de grau de dominância 0,5. Para o caso de dominância completa, a redução foi da ordem de 28,28 e 54,21% para as duas metodologias. Ainda com relação à precisão das predições de CEC, o método BLUP mostrou-se bem superior ao OLS, sendo que as estimativas de viés médio observadas com o BLUP em condições de baixa herdabilidade ($h^2 = 0,3$), com grau de dominância 0,5 e dominância completa, são semelhantes às do método dos quadrados mínimos ordinário, com herdabilidade 0,7 (Tabelas 4 e 5).

Um aspecto que chama a atenção, é o da qualidade das predições dos cruzamentos com os dois métodos em questão. As estimativas de erro quadrático médio, para o método OLS, foram marcadamente elevadas, quando comparadas às observadas com o BLUP. Neste aspecto, o método BLUP, em situação de herdabilidade 0,3, tem desempenho superior ao OLS com herdabilidade 0,7, nas predições da capacidade específica de combinação e do valor do cruzamento. Para a CGC, os dois métodos apresentam desempenho semelhante (Tabelas 4 e 5), com ligeira superioridade de precisão com o BLUP.

As Tabelas 6 e 7 mostram as médias das estimativas de correlação entre o valor genético e o predito, acompanhadas dos respectivos erros padrão, para as duas metodologias de análise (BLUP e OLS), aplicadas aos dois métodos dialélicos, nas situações de grau de dominância 0,5 e dominância completa.

TABELA 6. Média e erro padrão (entre parênteses), da correlação simples (Pearson) entre o valor genético e o predito pelos métodos BLUP e OLS em presença de grau dominância igual a 0,5, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0.3$				
CGC	0,8950 (0,088)	0,8638 (0,144)	0,8356 (0,111)	0,7862 (0,137)
CEC	0,2871 (0,203)	0,2822 (0,225)	0,2094 (0,112)	0,1763 (0,134)
Cruzamento	0,8286 (0,100)	0,7952 (0,122)	0,5201 (0,113)	0,4976 (0,111)
$h^2 = 0.5$				
CGC	0,9384 (0,052)	0,9214 (0,053)	0,9232 (0,050)	0,8859 (0,086)
CEC	0,4555 (0,149)	0,3889 (0,193)	0,3212 (0,113)	0,2444 (0,121)
Cruzamento	0,8949 (0,043)	0,8710 (0,071)	0,6904 (0,092)	0,6481 (0,116)
$h^2 = 0.7$				
CGC	0,9629 (0,019)	0,9583 (0,031)	0,9581 (0,028)	0,9469 (0,050)
CEC	0,5812 (0,129)	0,5531 (0,151)	0,4408 (0,114)	0,3822 (0,146)
Cruzamento	0,9270 (0,029)	0,9239 (0,039)	0,8083 (0,075)	0,7974 (0,089)

TABELA 7. Média e erro padrão (entre parênteses), da correlação simples (Pearson) entre o valor genético e o predito pelos métodos BLUP e OLS em presença de dominância completa, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0.3$				
CGC	0,8568 (0,138)	0,8486 (0,144)	0,8307 (0,113)	0,8037 (0,118)
CEC	0,4437 (0,201)	0,4370 (0,208)	0,2869 (0,148)	0,2697 (0,139)
Cruzamento	0,8006 (0,091)	0,7646 (0,098)	0,5301 (0,110)	0,5240 (0,113)
$h^2 = 0.5$				
CGC	0,9169 (0,051)	0,9096 (0,054)	0,8989 (0,070)	0,8890 (0,074)
CEC	0,5490 (0,149)	0,5461 (0,155)	0,4128 (0,131)	0,3925 (0,146)
Cruzamento	0,8553 (0,068)	0,8355 (0,077)	0,6976 (0,101)	0,6765 (0,103)
$h^2 = 0.7$				
CGC	0,9686 (0,035)	0,9486 (0,031)	0,9591 (0,052)	0,9405 (0,046)
CEC	0,6561 (0,111)	0,6425 (0,130)	0,5048 (0,142)	0,4868 (0,158)
Cruzamento	0,9104 (0,044)	0,8983 (0,043)	0,8180 (0,077)	0,8049 (0,065)

Como pode ser observado, não ocorreram diferenças expressivas entre os métodos II e IV de Griffing, havendo uma leve tendência de que o método II seja superior ao método IV, nas duas situações de grau de dominância estudadas.

Os valores médios de correlação entre o valor genético e o predito, referentes às capacidades de combinação e ao valor do cruzamento nas duas situações de grau de dominância simulados, com o BLUP, foram mais elevados e com menores valores de erro padrão que os observados pelo método de quadrados mínimos ordinário, evidenciando a superioridade do BLUP, independente do grau de dominância.

O aumento do grau de dominância proporcionou uma diminuição da correlação das predições de capacidade geral de combinação e do valor do cruzamento; entretanto, para a CEC, os valores médios de correlação tiveram um aumento significativo. Esta tendência se deve provavelmente ao aumento dos valores absolutos da CEC, que é função dos efeitos de dominância. É interessante notar as menores correlações no tocante à CEC, em relação à CGC. Este aspecto, no entanto, não impediu que houvesse altas correlações na predição dos valores de cruzamento, ao se utilizar o BLUP.

O aumento da herdabilidade, como observado para a situação de ausência de dominância, proporcionou uma melhoria considerável na predição dos diferentes efeitos, em especial da capacidade específica de combinação, onde foram observados aumentos superiores a 100%, no valor da correlação média, quando a herdabilidade passa de 0,3 para 0,7 na situação de grau de dominância 0,5. Com dominância completa, este incremento reduziu-se para valores próximos de 50 e 75%, com os métodos BLUP e OLS, respectivamente.

Os valores médios acompanhados dos respectivos erros padrão, das estimativas de correlação classificatória de Spearman entre o valor genético e o predito para os parâmetros do dialelo, pelos métodos BLUP e OLS, nas situações de dominância 0,5 e 1,0, estão apresentados nas Tabelas 8 e 9, respectivamente.

Observa-se de imediato, quando se compara os dois métodos dialélicos (Griffing II e IV), que o ordenamento dos valores genéticos foi semelhante entre os dois métodos. Contudo, o ordenamento observado com o método II, em todas as situações estudadas, diferentes níveis de herdabilidade e graus de dominância, foi melhor que o observado com o método IV.

TABELA 8. Média e erro padrão (entre parênteses) da correlação classificatória (Spearman) entre o valor genético e o predito pelos métodos BLUP e OLS em presença de grau dominância igual a 0,5, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0.3$				
CGC	0,8415 (0,092)	0,8223 (0,089)	0,7823 (0,113)	0,7628 (0,104)
CEC	0,2619 (0,130)	0,2498 (0,124)	0,1910 (0,141)	0,1628 (0,142)
Cruzamento	0,7929 (0,112)	0,7706 (0,128)	0,4809 (0,124)	0,4432 (0,132)
$h^2 = 0.5$				
CGC	0,8912 (0,061)	0,8710 (0,085)	0,8118 (0,102)	0,7816 (0,112)
CEC	0,4019 (0,131)	0,3686 (0,149)	0,2986 (0,101)	0,2106 (0,150)
Cruzamento	0,8494 (0,015)	0,8429 (0,082)	0,6364 (0,051)	0,5919 (0,128)
$h^2 = 0.7$				
CGC	0,9019 (0,052)	0,9114 (0,078)	0,8419 (0,095)	0,8628 (0,101)
CEC	0,5284 (0,106)	0,4988 (0,161)	0,4039 (0,146)	0,3210 (0,139)
Cruzamento	0,8976 (0,043)	0,8994 (0,056)	0,7801 (0,066)	0,7426 (0,099)

TABELA 9. Média e erro padrão (entre parênteses) da correlação classificatória entre o valor genético e o predito pelos métodos BLUP e OLS em presença de dominância completa, três níveis de herdabilidade, e considerando os métodos II e IV de Griffing.

	BLUP		OLS	
	Método II	Método IV	Método II	Método IV
$h^2 = 0,3$				
CGC	0,8246 (0,131)	0,8024 (0,129)	0,7603 (0,111)	0,7836 (0,118)
CEC	0,4073 (0,147)	0,4129 (0,148)	0,2188 (0,141)	0,2167 (0,152)
Cruzamento	0,7009 (0,099)	0,7783 (0,081)	0,4812 (0,122)	0,4925 (0,111)
$h^2 = 0,5$				
CGC	0,8590 (0,063)	0,8595 (0,062)	0,8078 (0,115)	0,8339 (0,077)
CEC	0,5203 (0,161)	0,5145 (0,153)	0,4081 (0,140)	0,3812 (0,131)
Cruzamento	0,7835 (0,081)	0,8355 (0,077)	0,6630 (0,103)	0,6469 (0,113)
$h^2 = 0,7$				
CGC	0,9308 (0,060)	0,9268 (0,040)	0,9106 (0,071)	0,9099 (0,067)
CEC	0,6664 (0,121)	0,6466 (0,132)	0,4758 (0,150)	0,4704 (0,160)
Cruzamento	0,9013 (0,061)	0,8991 (0,064)	0,8106 (0,086)	0,7908 (0,098)

O ordenamento dos valores preditos pelo método BLUP foi bem superior ao observado pelo OLS, independentemente do grau de dominância, para as predições das capacidades de combinação e para os valores dos cruzamentos. Em todas as situações estudadas, o BLUP apresentou melhor ordenamento e menores estimativas de erros padrão que o método OLS.

A variação da herdabilidade proporcionou diferenças expressivas na eficiência de ordenamento dos métodos de análise, principalmente com o método dos quadrados mínimos ordinário nas predições da CEC e do valor do cruzamento. Com o aumento da herdabilidade de 0,3 para 0,7, na situação de grau de dominância 0,5, há um acréscimo de mais de 100 e 60%, nos valores médios da correlação classificatória com o método OLS, para a CEC e dos valores dos cruzamentos, respectivamente. Com o BLUP, estes aumentos foram de mesma magnitude para as predições da CEC, e próximos a 15%, para o valor do cruzamento.

O aumento do grau de dominância diminuiu a eficiência dos métodos de análise para as predições da capacidade geral de combinação e do valor do cruzamento. À semelhança da correlação simples (Tabelas 6 e 7), há uma melhoria significativa no ordenamento da CEC com os dois métodos (BLUP e OLS). Este fato, como comentado anteriormente, pode ser função, principalmente, do aumento dos valores absolutos dos desvios de dominância.

Considerando a predição dos cruzamentos, aumentos no grau de dominância diminuíram, em parte, as diferenças de eficiência entre as metodologias, sendo que as diminuições na correlação de Spearman, ao se aumentar o grau de dominância, foram um pouco mais pronunciadas com o BLUP.

4.2 Predição de cruzamentos não avaliados

4.2.1 *Dados simulados*

Nas Tabelas 10 e 11, são apresentados os valores mínimos, médios, máximos e erros padrão da correlação simples, para os efeitos da capacidade específica de combinação e dos cruzamentos, de dialelos amostrados das populações de referência. As correlações consideradas são dos valores preditos dos 30 cruzamentos retirados com os valores paramétricos, e também com os valores preditos na análise dos dialelos completos, quando se considerou graus de dominância 0,5 e 1,0, respectivamente.

As estimativas de correlação, para a predição dos cruzamentos não avaliados, foram, de maneira geral, moderadas, sendo mais elevadas quando os valores de referência são as predições, considerando os dialelos completos.

Pode-se dizer que o BLUP é, assim, uma metodologia satisfatória para a predição de cruzamentos não realizados. Além disso, é razoável admitir que, à medida que o número de genitores aumenta (melhorando a amostragem na população de genitores), as diferenças entre os dois tipos de correlação diminuem.

TABELA 10. Valores mínimos, médios, máximos e erros padrão da correlação simples (Pearson) entre os valores preditos referentes a híbridos retirados e as predições sem retirá-los, bem como entre aqueles e os valores paramétricos, ao longo de 100 dialelos simulados e analisados pelo método BLUP, em situação de dominância 0,5 e três níveis de herdabilidade.

	CEC		CRUZAMENTO	
	Paramétrico	Dialelo	Paramétrico	Dialelo
$h^2 = 0.3$				
Mínimo	-0,0325	0,1451	0,1016	0,1974
Média	0,1439	0,4860	0,5319	0,6421
Máximo	0,2879	0,5819	0,6813	0,7780
Erro Padrão	0,0943	0,1129	0,1725	0,1621
$h^2 = 0.5$				
Mínimo	0,0103	0,1053	0,1137	0,1742
Média	0,1726	0,5230	0,5814	0,6910
Máximo	0,3519	0,6901	0,7151	0,8468
Erro Padrão	0,0963	0,1426	0,1687	0,1827
$h^2 = 0.7$				
Mínimo	0,0364	0,1849	0,1997	0,2653
Média	0,2313	0,5684	0,6619	0,8421
Máximo	0,3797	0,7539	0,8012	0,9210
Erro Padrão	0,0721	0,1043	0,1415	0,1319

TABELA 11. Valores mínimos, médias, máximos e erros padrão da correlação simples (Pearson) entre os valores preditos referentes a híbridos retirados e as predições sem retirá-los, bem como entre aqueles e os valores paramétricos, ao longo de 100 dialelos simulados e analisados pelo método BLUP, em situação de dominância completa e três níveis de herdabilidade.

	CEC		CRUZAMENTO	
	Paramétrico	Dialelo	Paramétrico	Dialelo
$h^2 = 0,3$				
Mínimo	-0,0475	0,0153	0,0941	0,2502
Média	0,1843	0,4993	0,5663	0,6983
Máximo	0,3045	0,6204	0,6992	0,8314
Erro Padrão	0,1023	0,1613	0,1515	0,1324
$h^2 = 0,5$				
Mínimo	-0,0211	0,0902	0,1569	0,2305
Média	0,2210	0,5780	0,6163	0,7684
Máximo	0,4053	0,7094	0,7508	0,8597
Erro Padrão	0,1142	0,1135	0,1303	0,1407
$h^2 = 0,7$				
Mínimo	0,0912	0,1309	0,1940	0,3014
Média	0,2819	0,6433	0,6924	0,8106
Máximo	0,4001	0,8196	0,8257	0,9152
Erro Padrão	0,1203	0,1421	0,1232	0,1529

Tomando-se o valor médio das correlações com o método dos quadrados mínimos ordinário (OLS), e herdabilidade igual a 0,3 (Tabelas 6 e 7), o valor médio da correlação entre o valor predito e o paramétrico para o método IV de Griffing, nos dois graus de dominância simulados, foram inferiores aos obtidos com o BLUP, quando o cruzamento em questão não foi avaliado. Para níveis de herdabilidade mais elevados, o mesmo não ocorre, mas os valores de correlação são relativamente elevados.

Segundo Bernardo (1994), que igualmente simulou a perda de cruzamentos, deve-se tomar como referência a correlação entre os valores genotípico e o fenotípico do cruzamento, que nada mais é que a raiz quadrada da herdabilidade para a característica. Neste trabalho, em todas as situações

estudadas, três níveis de herdabilidade e dois graus de dominância, os valores médios da correlação entre os valores preditos dos cruzamentos considerando a perda, e aqueles dos dialelos completos, foram superiores aos níveis preconizados (iguais a 0,5477, 0,7071 e 0,8367), exceção feita para as situações de herdabilidade 0,5 com dominância 0,5 e herdabilidade 0,7 com dominância completa (Tabelas 10 e 11), ou seja, pode-se dizer que, com baixa herdabilidade e/ou pequeno grau de dominância, o BLUP possibilita predições de qualidade superior ao próprio fenótipo isoladamente.

As correlações entre o valor predito e o observado foram mais elevadas nas situações de dominância completa. Esta tendência já havia sido observada anteriormente (Tabelas 6 e 7). Contudo, este aumento da correlação não proporcionou maior precisão na predição da CEC, na qual foram constatados maior amplitude e maior erro padrão, associados para as situações de dominância completa.

Os valores de correlação, observados neste estudo, são superiores aos relatados por Bernardo (1996a), que avaliou a eficiência do método BLUP, com informações de parentesco genético por genealogia, em 16 combinações de cruzamentos entre linhagens de nove grupos heteróticos de milho. As estimativas de correlação entre o valor predito e o observado, para a produção de grãos, variou de 0,426 a 0,762, para valores de herdabilidade entre 0,3505 a 0,7465. Por outro lado, neste estudo, a correlação média entre o valor predito e o observado, variou de 0,6421 a 0,8421 e de 0,6983 a 0,8106, para os graus de dominância 0,5 e 1,0, respectivamente, em intervalos de herdabilidade simulada de 0,3 a 0,7.

O aumento da herdabilidade, resultante da redução relativa da variação ambiental, proporcionou uma melhoria considerável na predição dos

diferentes efeitos dos cruzamentos não avaliados, em especial para a capacidade específica de combinação, com a qual foram registrados os maiores incrementos relativos na correlação média. Na situação de dominância 0,5, ocorreram aumentos de 60,74%, na correlação média entre o valor predito e o paramétrico, e 16,95%, na correlação com o valor observado pela análise do dialelo completo, quando o valor da herdabilidade gerada variou de 0,3 para 0,7. Com dominância completa, o incremento na correlação média foi de 52,96 e 28,84%, para as comparações com os valores paramétricos e os obtidos com a análise do dialelo sem falhas, respectivamente.

Os valores mínimos, médios, máximos e erros padrão da correlação classificatória (Spearman) para a CEC e cruzamento entre os valores preditos dos 30 cruzamentos retirados e os valores paramétricos e também entre aqueles e os preditos na análise com o BLUP dos dialelos completos, considerando graus de dominância 0,5 e 1,0, estão apresentados nas Tabelas 12 e 13, respectivamente.

Nota-se que o desempenho do método BLUP no ordenamento dos efeitos da CEC e do valor do cruzamento, quando estes não são avaliados, é semelhante ao observado para a correlação simples de Pearson (Tabelas 10 e 11), sendo que os valores médios da correlação classificatória foram sempre inferiores. Para a situação de baixa herdabilidade ($h^2 = 0,3$), nos dois graus de dominância simulados, o ordenamento pelo método BLUP, das predições dos cruzamentos não avaliados, foi superior ao obtido pelo método dos quadrados mínimos ordinário, no mesmo nível de herdabilidade (Tabelas 8 e 9).

Deve-se ressaltar que as baixas correlações, observadas entre os valores preditos e os paramétricos (Tabelas 12 e 13), também devem estar associadas ao número não muito elevado de genitores (15) dos dialelos amostrados. É de se esperar que, à medida que esse número aumenta, diminuam as diferenças observadas entre as correlações, considerando-se valores paramétricos, e aqueles com os valores dos dialelos completos.

TABELA 12. Valores mínimos, médios, máximos e erros padrão da correlação classificatória (Spearman) entre os valores preditos referentes a híbridos retirados e as predições sem retirá-los, bem como entre aqueles e os valores paramétricos, ao longo de 100 dialelos simulados e analisados pelo método BLUP, em situação de dominância 0,5 e três níveis de herdabilidade.

	CEC		CRUZAMENTO	
	Paramétrico	Dialelo	Paramétrico	Dialelo
	$h^2 = 0,3$			
Mínimo	-0,0412	0,0237	0,1223	0,1568
Média	0,1293	0,4615	0,5191	0,6143
Máximo	0,2743	0,5911	0,6441	0,7672
Erro Padrão	0,0968	0,1239	0,1344	0,1391
	$h^2 = 0,5$			
Mínimo	-0,0248	0,0971	0,1005	0,1689
Média	0,1629	0,5344	0,5440	0,7162
Máximo	0,3716	0,6839	0,7010	0,9013
Erro Padrão	0,1014	0,1266	0,1398	0,1912
	$h^2 = 0,7$			
Mínimo	0,0814	0,1557	0,1380	0,2409
Média	0,2001	0,5592	0,6096	0,7828
Máximo	0,3680	0,7029	0,7519	0,9156
Erro Padrão	0,1142	0,1635	0,1402	0,1371

TABELA 13. Valores mínimos, médios, máximos e erros padrão da correlação classificatória (Spearman) entre os valores preditos referentes a híbridos retirados e as predições sem retirá-los, bem como entre aqueles e os valores paramétricos, ao longo de 100 dialelos simulados e analisados pelo método BLUP, em situação de dominância 0,5 e três níveis de herdabilidade.

	CEC		CRUZAMENTO	
	Paramétrico	Dialelo	Paramétrico	Dialelo
$h^2 = 0.3$				
Mínimo	-0,0198	0,0205	0,0630	0,2938
Média	0,1629	0,4510	0,5481	0,6693
Máximo	0,2708	0,6169	0,6792	0,8136
Erro Padrão	0,1143	0,1816	0,1774	0,1416
$h^2 = 0.5$				
Mínimo	0,0591	0,0607	0,1954	0,2172
Média	0,2069	0,4917	0,5877	0,7254
Máximo	0,3716	0,6958	0,7701	0,8743
Erro Padrão	0,1018	0,1161	0,1547	0,1624
$h^2 = 0.7$				
Mínimo	0,1033	0,1602	0,0241	0,2601
Média	0,2691	0,5820	0,6896	0,7985
Máximo	0,3951	0,7235	0,8629	0,9015
Erro Padrão	0,1123	0,1212	0,1943	0,1513

Em todas as situações estudadas, diferentes níveis de herdabilidade e graus de dominância, o valor da correlação classificatória para a predição do cruzamento considerando os valores dos dialelos completos como referência, foi superior a 50%, mesmo nas situações de baixa herdabilidade (Tabelas 12 e 13). Estes valores sugerem que o método BLUP, em associação com informações de similaridade genética obtidas por marcadores moleculares, pode ser empregado com eficiência razoável na predição de cruzamentos não realizados ou não avaliados em esquemas dialélicos.

Como esperado, a eficiência na predição do valor de cruzamentos não realizados pelo método BLUP, aumenta com a elevação do nível de herdabilidade. À semelhança da correlação simples de Pearson (Tabelas 12 e

13), foram observados acréscimos significativos, nos valores médios da correlação classificatória para as predições da capacidade específica de combinação e para os valores dos cruzamentos, nas duas situações estudadas, de grau de dominância.

A variação no ordenamento das predições da CEC e valor do cruzamento, em função do grau de dominância, apresentou comportamento semelhante ao observado para a correlação simples de Pearson, ou seja, à medida que se aumentou a participação relativa dos efeitos de dominância no valor do cruzamento, observou-se maior eficiência no ordenamento pelo método BLUP.

4.2.2 Dados experimentais

Os valores mínimos, médios, máximos e erros padrão dos coeficientes de correlação simples (Pearson) ao longo de 100 simulações, para a predição dos efeitos da capacidade específica de combinação (CEC) e valor do cruzamento, considerando diferentes números de cruzamentos retirados, são apresentados na Tabela 14.

As estimativas da correlação, para os efeitos da capacidade específica de combinação, foram menores e com maior variação (maiores estimativas de amplitude e erro padrão), que para a predição do valor do cruzamento. Todavia, podem ser observados valores máximos semelhantes para as predições da CEC e do cruzamento, especialmente nas situações em que foram simulados os menores níveis de perda (10 e 20 cruzamentos). É interessante notar que estes valores foram até maiores que o observado para a situação em que foram preditos os valores da CEC, retirando-se um híbrido

por vez. Este valor, calculado a partir de 120 pares de valores (pois são 120 cruzamentos), foi igual a 0,5968.

Vale ressaltar a ocorrência de estimativas bem inferiores ao valor médio e até mesmo correlações negativas, como é o caso das situações onde foram retirados 50 e 60 cruzamentos. Entretanto, adotando-se o valor 0,5968 (correlação entre o valor predito e o observado, quando foi retirado um híbrido de cada vez), como um padrão de referência, correspondendo a um desbalanceamento mínimo, foram observados 63, 61, 58, 41, 31 e 28 estimativas de correlação superiores a este limite estabelecido, para as situações em que foram retirados 10, 20, 30, 40, 50 e 60 híbridos, respectivamente. É razoável, assim, admitir que o método BLUP pode ser eficiente na predição da capacidade específica de combinação de híbridos não avaliados, ainda que sob elevados níveis de desbalanceamento.

TABELA 14. Valores mínimos, médios, máximos e erros padrão da correlação simples (Pearson) entre o valor genético observado e o predito ao longo de 100 simulações, nos diversos níveis de retirada de cruzamentos para o dialelo envolvendo 22 linhagens de feijão.

	Número de Híbridos Retirados					
	10	20	30	40	50	60
	CEC					
Mínimo	0,0186	0,1432	0,0379	0,0695	-0,0531	-0,1224
Média	0,6258	0,6128	0,6072	0,5367	0,4855	0,4833
Máximo	0,9521	0,9205	0,8492	0,8215	0,8217	0,7771
Erro Padrão	0,2226	0,1581	0,1412	0,1566	0,1886	0,1932
	CRUZAMENTO					
Mínimo	0,2325	0,2909	0,4013	0,3593	0,0594	0,0901
Média	0,7517	0,7427	0,7440	0,7035	0,6658	0,6635
Máximo	0,9624	0,9482	0,9052	0,8940	0,9100	0,8533
Erro Padrão	0,1630	0,1276	0,0982	0,1125	0,1537	0,1482

Os valores médios das estimativas da correlação, entre o valor predito e o observado, para a capacidade específica de combinação, variou de 0,4833 a 0,6258, para as situações em que foi simulada a retirada de 10 a 60 cruzamentos. Estes valores são bem superiores aos reportados por Bernardo (1995), ao estudar vários modelos genéticos na predição dos mesmos efeitos, em um dialelo parcial com 67 híbridos de milho. Contudo, no estudo realizado, a contribuição dos efeitos de dominância foi considerada relativamente pequena, quando comparada com outros estudos com a cultura do milho.

A estimativa do coeficiente de correlação entre o valor predito de cruzamento e o observado, para os 120 cruzamentos, quando foi retirado 1 cruzamento por vez, foi igual a 0,7852, sendo este valor mais elevado que os valores encontrados por Bernardo (1996a), ao avaliar a eficiência do método BLUP, com informações de parentesco genético por genealogia, em 16 combinações de cruzamentos entre linhagens de nove grupos heteróticos de milho, em cruzamentos dialélicos parciais, utilizando o mesmo processo de validação utilizado neste estudo, sendo que os valores, encontrados pelo autor, variaram de 0,426 a 0,762.

Para as situações em que foi simulada a retirada de 10 a 60 híbridos por vez, os valores médios da correlação entre o valor predito de cruzamento e o observado, variou de 0,6635 a 0,7517. Ao se adotar o valor 0,7852 (correspondente à perda de apenas 1 cruzamento), como valor de referência para a validação do método, foram observadas 53, 45, 40, 28, 15 e 15 estimativas de correlação superiores ao limite estabelecido para os níveis de retirada de 10 até 60 híbridos. Embora estas frequências tenham sido inferiores às observadas para a predição da capacidade específica de

combinação, deve-se considerar que este valor limite (0,7852), como comentado anteriormente, é elevado, superior ao relatado na literatura.

Com estimativas de correlação entre o valor observado e o predito, inferiores às observadas neste estudo, Bernardo (1994, 1995 e 1996a) afirmou que a utilização do método BLUP, associado a estimativas de parentesco, pode ser uma alternativa viável na identificação de cruzamentos promissores, na cultura do milho. O presente estudo contribui para a generalização dessa idéia em outras culturas, utilizando informações de similaridade genética, a partir de marcadores moleculares.

No presente experimento, as estimativas dos componentes de variâncias foram 14272,55, 5242,79 e 10890,72 kg²/ha², para os efeitos de capacidade geral de combinação, específica e erro, respectivamente. A estimativa de herdabilidade (h^2), ao nível de médias de cruzamento, foi igual a 66,91%. As médias de correlação entre o valor predito e o observado pela análise com o BLUP, aqui observadas, são inferiores às obtidas com dados simulados para os níveis de herdabilidade 0,7 nas duas situações de grau de dominância estudados (Tabelas 10 e 11). Entretanto, foram superiores aos valores médios observados com a herdabilidade 0,5. Certamente, as falhas originalmente ocorridas e, conseqüentemente, a não avaliação de todos os cruzamentos possíveis para cada linhagem, podem ter contribuído para a redução das estimativas de correlação. Neste experimento, o número médio de cruzamentos, em que cada linhagem participou, foi de aproximadamente 12, variando de 6 até 17 (Abreu, 1997).

Na Tabela 15, estão apresentados os valores mínimos, médios, máximos e erros padrão das estimativas de correlação classificatória (Spearman), ao longo de 100 simulações, para a predição dos efeitos da

capacidade específica de combinação (CEC) e dos valores dos cruzamentos, considerando diferentes números de cruzamentos retirados. A estimativa da correlação classificatória, quando foi simulada a retirada de um híbrido por vez, para as predições da CEC e do valor do cruzamento, foram iguais a 0,6065 e 0,7852, respectivamente. Para as situações em que foram simuladas retiradas de 10 a 60 cruzamentos, o valor médio da correlação classificatória, variou de 0,4591 a 0,6198 e de 0,6303 a 0,7492, para as predições da CEC e do cruzamento, respectivamente. Este valores foram bastante próximos aos observados para a correlação simples (Tabela 14). Em linhas gerais, quando se aumentou o número de híbridos com desempenho a ser predito, foram observadas tendências de reduções, na eficiência do ordenamento, tanto para a predição da CEC, quanto do valor do cruzamento, observando-se menores médias para a correlação classificatória e também maiores estimativas dos erros padrão.

TABELA 15. Valores mínimos, médios, máximos e erros padrão da correlação classificatória (Spearmam) entre o valor genético observado e o predito ao longo de 100 simulações, nos diversos níveis de retirada de cruzamentos para o dialelo envolvendo 22 linhagens de feijão.

	Número de Híbridos Retirados					
	10	20	30	40	50	60
	CEC					
Mínimo	-0,0151	0,1392	0,1351	0,1051	0,0241	0,0485
Média	0,6198	0,6056	0,5924	0,5409	0,4763	0,4591
Máximo	0,9183	0,8973	0,8963	0,8021	0,8037	0,8114
Erro Padrão	0,2134	0,1614	0,1507	0,1679	0,1740	0,1827
	CRUZAMENTO					
Mínimo	0,2146	0,2650	0,2367	0,1350	0,1493	0,1051
Média	0,7492	0,7391	0,6970	0,6636	0,6315	0,6303
Máximo	0,9074	0,8980	0,9122	0,8530	0,8812	0,8306
Erro Padrão	0,1401	0,1387	0,1353	0,1485	0,1632	0,1649

A eficiência no ordenamento das predições, foi semelhante à observada para os dados simulados nas situações de alta herdabilidade ($h^2 = 0,7$), para os dois graus de dominância estudados (Tabelas 12 e 13), principalmente para a capacidade específica de combinação, nas situações em que foi simulada a retirada de até 30 híbridos, nas quais os valores médios das estimativas do coeficiente de correlação classificatória foram superiores aos observados, para os dados simulados. É interessante salientar a concordância de resultados, considerando dados simulados e experimentais.

Uma metodologia eficiente, para predição de híbridos não realizados, pode contribuir sobremaneira no sucesso de programas de obtenção de híbridos comerciais, na medida em que permite avaliar um número economicamente viável de cruzamentos por linhagem, prever o desempenho dos híbridos possíveis e então partir para avaliações das combinações mais promissoras, em menor número e em testes mais extensivos, com maior precisão experimental.

Os processos de predição de híbridos têm sido propostos e avaliados por diversos autores, sendo sua eficiência dependente, tanto do modelo adotado, como da precisão das estimativas dos parâmetros envolvidos. Jenkins (1934), utilizando dados de híbridos de milho, sugeriu a utilização de quatro alternativas para predição de híbridos duplos, conhecidos como métodos A, B, C e D de Jenkins.

Os métodos diferem com respeito ao tipo de ação gênica envolvida, sendo que os métodos A, C e D pressupõem apenas efeitos aditivos, enquanto que o método B considera efeitos aditivos e não aditivos (dominância e vários tipos de epistasia). O autor apresentou um estudo sobre a eficiência dos

métodos em questão, obtendo valores de correlação entre a produção predita e a observada iguais a 0,75, 0,76, 0,73 e 0,61, para os quatro métodos, respectivamente; porém, sendo o valor da correlação ligeiramente superior e também o fato de considerar o método B mais completo, o autor recomendou a utilização deste como o mais eficiente.

Diversos pesquisadores têm verificado a eficiência do método B na predição de médias de híbridos duplos. Segundo Hallauer e Miranda Filho (1988), o método tornou-se extensamente utilizado na predição do desempenho de híbridos duplos, nos Estados Unidos, nos anos 40 a 70, mesmo com estudos realizados, mostrando a forte influência da interação genótipo por ambiente.

Neste estudo, as correlações entre o valor predito e o observado, foram superiores aos normalmente relatados na literatura, para outros métodos preditivos. Embora não conclusivos, os resultados obtidos neste estudo sugerem a potencialidade da utilização do método BLUP, na predição dos efeitos da capacidade específica de combinação e do valor de cruzamentos não realizados em cruzamentos dialélicos. Deve-se ressaltar, contudo, que certos aspectos sobre o processo de predição não foram abordados, como a existência de interação genótipo \times ambiente, efeitos epistáticos, grau de relacionamento genético entre as linhagens e também o número de híbridos a serem avaliados para predições eficientes.

Certamente, estudos teóricos e práticos mais detalhados sobre a utilização do método BLUP, em cruzamentos dialélicos, poderão contribuir para elucidar pontos importantes, especialmente quanto ao número de linhagens e de cruzamentos a serem realizados, bem como modelos genéticos adequados para obtenção de predições com níveis adequados de precisão.

A complexidade das análises envolvidas em qualquer metodologia é um ponto de importância que deve ser considerado. A difusão do método está relacionado com a disponibilidade de "softwares" e algoritmos eficientes, para realização das análises. O método BLUP tem grande aceitação entre os melhoristas da área animal, onde programas e algoritmos eficientes para modelos genéticos específicos foram desenvolvidos, ao longo dos anos 80 e 90. Contudo, sua adaptação ao fitomelhoramento, muitas vezes, torna-se inviável.

Ao que parece, até o presente momento não se tem relatos, na literatura, a respeito de programas desenvolvidos, considerando os modelos genéticos comumente utilizados no fitomelhoramento, como é o caso de cruzamentos dialélicos e/ou para casos em que a similaridade genética é quantificada por marcadores moleculares. Neste estudo, um conjunto de rotinas foram escritas em linguagem de programação SAS[®], através do módulo de matrizes (SAS Institute, 1995), que estão apresentados no Anexo. Entretanto, sua utilização é restrita às situações e modelos genéticos aqui apresentados.

Na Tabela 16, estão apresentados os valores preditos de CEC e do valor genotípico de cruzamentos, referentes aos híbridos não obtidos originalmente por Abreu (1997), em virtude de problemas de incompatibilidade entre as linhagens. Isto não significa que tais híbridos não possam ser obtidos, mas que são necessárias técnicas especiais para tanto, como o uso de cruzamentos-ponte (Sousa, 1993). Este aspecto ilustra a aplicabilidade potencial do BLUP, que pode fornecer indicativos de quais os cruzamentos promissores que justificariam o esforço dispendido, para a realização dos cruzamentos-ponte.

TABELA 16. Valores preditos da capacidade específica de combinação e valor do cruzamentos realizados e não realizados (em negrito), pelo método BLUP em um o dialelo envolvendo 22 linhagens de feijão^{1/}.

P1	P2	Cruz.	CEC	P1	P2	Cruz.	CEC	P1	P2	Cruz.	CEC
6	6	457.1196	35.3659	4	8	393.3804	44.3897	9	13	364.4001	26.1030
12	17	441.8586	40.7014	6	16	392.2336	1.0205	2	5	364.2377	36.6147
4	6	436.0193	38.5674	15	15	391.0872	9.7111	6	18	363.2490	-28.3880
6	12	429.2644	41.7043	13	22	388.8628	47.8266	1	17	362.8388	17.7723
4	12	428.4066	58.4658	3	8	387.9267	51.4885	10	15	362.6294	25.7774
12	12	425.6374	48.4640	2	4	387.4779	37.9908	7	18	362.6053	28.7683
4	17	424.6370	24.7551	7	17	385.3422	19.6947	6	11	361.4484	24.5256
3	6	424.2493	39.9569	2	9	384.4897	39.7560	9	17	361.3106	-9.8290
17	17	423.2933	-1.8478	3	9	383.9662	33.9071	4	11	361.1514	40.5671
4	4	420.4419	45.8193	9	12	383.4707	35.6005	5	18	360.9715	14.8084
8	12	420.2365	70.2979	3	5	381.5614	41.4658	1	12	360.9672	41.6255
5	17	417.6865	39.7130	5	9	379.2577	50.7348	6	10	360.5801	15.8110
6	22	414.5215	40.9191	2	2	377.0824	44.1156	3	7	360.3466	34.4920
3	12	412.9138	49.2408	7	12	376.7193	30.3957	4	10	359.2025	31.8185
6	17	412.6757	-10.7717	4	16	376.2602	8.6126	2	13	359.0187	11.5570
5	6	412.4091	37.0260	8	17	375.8167	1.8942	5	16	357.4089	11.6697
6	15	412.3805	10.8156	16	17	374.3158	-18.7567	15	16	357.2300	-10.7322
6	8	409.5398	37.8040	4	7	374.2576	33.4222	17	22	356.7341	8.4607
3	17	406.3143	18.6574	9	16	373.9298	35.0246	3	18	355.6145	-0.2320
2	12	405.4685	50.3984	8	15	373.9224	21.8824	3	13	355.0033	6.2368
12	15	405.0269	25.7521	6	13	373.9010	5.9401	7	9	354.1795	25.7866
4	9	405.0172	52.4138	5	5	373.8985	43.0925	2	8	353.5437	25.7084
18	19	404.6636	71.3227	8	9	373.8923	44.4721	10	13	352.4763	23.5972
12	13	404.3279	34.5321	8	16	373.5822	31.8941	1	4	352.3941	32.2125
2	6	404.1358	33.4239	13	13	373.0641	16.4429	12	18	351.8563	-17.4906
6	9	404.1089	35.3739	5	8	370.6662	43.2331	10	17	351.5569	-7.1776
3	4	403.8554	42.5020	20	20	369.8714	33.1352	12	16	351.5213	-17.4016
15	17	403.7532	2.8549	1	6	369.5664	27.0088	5	7	351.3659	32.0018
5	12	402.6443	56.7136	13	15	369.3358	0.3371	2	11	350.7779	49.0567
18	22	398.7593	49.9253	4	13	368.8575	20.9389	17	18	350.5474	-46.7363
3	3	397.8620	47.6895	3	15	368.5618	2.7875	5	19	350.4493	32.4656
4	5	397.1506	44.4075	7	15	368.3030	24.5381	17	20	349.5279	-5.9913
13	17	396.8088	5.9276	8	8	366.2330	43.5292	13	16	349.2244	-9.4224
6	7	395.6892	32.1493	10	12	365.9019	30.5254	7	8	347.6352	33.2838
5	15	395.3320	39.2410	2	3	365.6534	24.0837	2	15	347.5252	-9.6463
4	15	395.1936	17.1942	9	15	364.4048	15.1478	19	20	347.1776	23.6272

Continua...

TABELA 16. Cont.

PI	P2	Cruz.	CEC	PI	P2	Cruz.	CEC	PI	P2	Cruz.	CEC
2	10	346.4086	23.0049	17	21	327.0974	-9.4441	13	14	300.5321	11.2992
11	17	344.2301	-3.5782	2	18	326.2571	-20.9864	7	19	298.2865	-7.3711
13	20	344.1114	6.4399	1	15	325.6663	2.4824	10	11	298.0376	16.0176
2	17	343.9583	-35.0957	7	13	325.4699	-1.4432	8	14	297.5486	25.2744
2	7	343.4553	27.2816	2	19	325.3232	6.2591	1	11	296.8581	27.4940
5	13	343.1577	19.1821	1	8	325.3181	30.9872	1	10	295.5992	18.6495
3	10	343.1148	16.9492	16	16	325.3027	-35.3697	15	19	294.7213	-29.0380
4	18	342.7582	-25.3133	10	22	324.9837	16.0942	3	21	294.4931	-12.6230
11	13	342.5095	23.0882	1	16	324.4211	11.5890	9	22	293.8069	-27.4877
12	19	342.4039	1.2365	8	11	324.3340	27.7443	16	19	293.6067	-22.8515
11	15	340.2638	14.3380	1	2	324.1216	28.7040	10	20	292.5471	-13.0999
21	22	339.3115	44.5560	9	10	323.3691	19.9341	18	18	291.9868	-69.5335
8	13	339.2874	13.3051	13	19	322.1353	-8.1045	15	21	291.2262	-17.7815
9	9	338.9036	21.7657	10	16	321.7847	-4.7154	19	21	290.9099	6.2993
20	22	338.7287	7.6350	2	16	320.6632	-26.1564	16	21	289.7783	-13.3493
13	21	337.7910	27.4505	20	21	318.6829	18.2850	14	15	289.4704	-9.8964
1	9	337.5515	36.3870	16	20	317.7944	-6.4063	11	11	288.0956	17.6200
1	3	336.8754	30.5171	11	16	314.4391	-1.1348	1	1	287.7077	22.7160
6	20	335.3360	-26.8857	8	18	313.2763	-28.8358	3	20	287.2766	-39.7876
5	11	335.0994	38.6851	12	22	312.3396	-38.9726	5	14	286.6708	10.3455
15	22	334.6448	-4.8786	15	20	312.2568	-18.1113	21	21	285.1742	21.1144
4	22	334.4478	-15.5891	5	22	311.9379	-16.1907	11	22	284.2053	-13.7581
3	11	334.0450	23.7210	7	11	311.7782	14.7002	4	19	283.0409	-56.8511
18	20	333.5768	-1.5682	4	21	311.6648	-7.6764	1	20	282.8559	-18.0080
15	18	333.4337	-38.1135	7	10	311.6329	7.2378	9	20	282.7139	-44.2232
16	22	333.1132	13.0409	12	20	311.3544	-33.3484	22	22	282.2024	-43.2488
5	10	333.0944	29.8696	9	11	310.5558	17.6147	8	22	281.5708	-42.5067
3	16	332.7894	-22.6331	16	18	309.7198	-52.8702	8	19	280.6619	-33.2707
19	22	332.7046	17.3983	19	19	309.6591	4.4976	2	22	279.0142	-50.1948
11	12	332.3846	8.5602	10	10	309.2825	16.9546	2	20	278.3823	-56.4692
7	16	332.2964	-1.1167	1	13	308.6914	10.1501	5	20	275.1361	-39.4942
9	18	331.8955	-7.4336	1	7	307.8643	22.2176	3	14	274.9824	-11.0262
6	19	331.5569	-31.9007	4	14	306.7798	13.3037	2	14	274.3086	-3.0971
6	14	330.2083	8.4091	6	21	305.5780	-37.3287	10	14	273.6471	16.5609
1	5	329.6748	31.1411	1	18	305.1191	-8.1369	8	20	271.5002	-40.3672
17	19	328.7993	-21.6028	14	17	304.8746	-18.2930	7	22	271.1968	-44.6057
13	18	327.7890	-31.2818	11	18	304.0555	-11.9424	3	22	269.2467	-68.5651
8	10	327.3185	23.1701	12	14	303.2122	3.7032	10	18	269.1866	-57.7375
7	7	327.2138	21.0601	4	20	300.7522	-36.7640	11	20	268.4978	-24.3507

Continua...

TABELA 16. Cont.

PI	P2	Cruz.	CEC	PI	P2	Cruz.	CEC	PI	P2	Cruz.	CEC
7	20	267.9528	-36.5496	2	21	249.8313	-48.6819	18	21	238.6676	-74.1225
9	14	265.2214	-4.2699	9	21	248.2363	-42.3625	7	21	235.1180	-49.9887
10	21	260.5851	-17.6087	1	22	248.2264	-46.9950	1	19	234.6640	-50.4125
14	16	257.8959	-33.1917	11	19	242.1926	-38.7823	9	19	226.0513	-85.0984
11	14	257.2321	11.0720	11	21	240.3591	-27.5976	14	19	223.9991	-25.7660
12	21	256.1573	-64.4593	10	19	239.3542	-59.3904	8	21	218.8418	-74.5399
3	19	253.8276	-73.8394	1	14	239.1224	-4.2957	14	21	218.1555	-16.9783
5	21	253.7749	-43.6579	1	21	238.9870	-25.5387	14	14	183.9464	-37.8982
7	14	253.2639	-10.7353	14	20	238.9499	-16.2973				
14	18	252.3689	-43.0116	14	22	238.6758	-13.6002				

^{1/} Pais: 1.Sanilac; 2.Brasil 2; 3.A-62; 4.Porrillo; 5.Jamapa; 6.Milionário; 7.IPA 7419; 8.A-175; 9.Rojo de seda; 10.A140; 11.Carioca; 12.Compuosto Negro Chimaltenango; 13.Puebla 152; 14.Creat Northern; 15.Flor de Mayo; 16.A114; 17.Ojo de Liebre; 18.Apetito Blanco; 19.Mont Calm; 20.Jalo; 21.Diacol Andino; 22.Perry Marrow.

A título de ilustração, observa-se na Tabela 16, que dentre os 10 (dez) cruzamentos mais promissores (incluídas aí algumas linhagens), quatro estiveram ausentes e tiveram o seu valor genotípico predito pelo BLUP. São cruzamentos que, possivelmente, poderiam originar populações promissoras para a obtenção de cultivares superiores. Trata-se de um exemplo real que permitiu ilustrar claramente o potencial do BLUP, na análise dialélica, dado que se tenha algum tipo de informação sobre a similaridade genética entre as linhagens. No tocante a este último aspecto, os marcadores moleculares do tipo RAPD podem ser de grande valia, pelo menor custo e maior rapidez. Por se tratar de marcador tido como de baixa repetibilidade, em virtude de sua alta sensibilidade às condições de laboratório (Michelmores et al., 1994) além da desvantagem da herança dominante, ele tem sido mais extensivamente

empregado em estudos básicos de diversidade. Contudo, os resultados do presente estudo permitem inferir que o BLUP pode vir a corresponder a uma das principais técnicas pelas quais o RAPD tenha expressiva aplicação em etapas avançadas de programas de melhoramento.

5 CONCLUSÕES

1. Os resultados deste trabalho permitiram concluir que a metodologia da melhor predição linear não tendenciosa (BLUP) é altamente eficiente na análise dialélica, quando existe informação sobre a similaridade genética entre os materiais sendo cruzados, a partir de marcadores moleculares.
2. A melhor predição linear não tendenciosa (BLUP) possibilita uma precisão marcadamente superior, em relação ao método dos quadrados mínimos ordinário, para a predição de efeitos de cruzamentos e de capacidades de combinação, em diferentes condições de herdabilidade e de grau médio de dominância, independentemente de se utilizar os métodos II ou IV de Griffing.
3. A disponibilidade de informação molecular sobre o grau de similaridade entre materiais genéticos, em associação ao BLUP, permite a predição dos valores genéticos de cruzamentos não realizados e de efeitos de capacidade específica de combinação, com considerável acurácia em diferentes condições de herdabilidade e grau de dominância, bem como sob elevados níveis de desbalanceamento.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU, A. F. B. Predição de potencial genético de populações segregantes do feijoeiro utilizando genitores inter-raciais, Lavras: UFLA, 1997. 80p. (Tese de Doutorado – Genética e Melhoramento de Plantas).
- BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop. Sci.*, Madison, v.34, n.1, p.20-5, jan.-feb., 1994.
- BERNARDO, R. Genetic models for predicting maize performance in unbalanced yield trial data. *Crop Sci.*, Madison, v.35, n.1, p.141-47, jan.-feb., 1995.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance. *Crop Sci.*, Madison, v.36, n.1, p.50-6, jan.-feb., 1996a.
- BERNARDO, R. Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. *Crop Sci.*, Madison, v.36, n.4, p.862-6, aug.-sep., 1996b.
- BERNARDO, R. Best linear unbiased prediction of the performance of crosses between untested maize inbreds. *Crop Sci.*, Madison, v.36, n.4, p.872-6, aug.-sep., 1996c.
- BOUCHEZ, A.; GOFFINET, B. Evaluation of selection index: application to the choice of an indirect multitrait selection index for soybean breeding. *Theor. Appl. Genet.*, v.79, n.1, p.261-67. 1990.

- BROWN, K. G.** Asymptotic behavior of MINQUE - type estimation of variances components. *Annals of Statistics*, Hayward, v.4, p.746-54, 1976.
- BUENO FILHO, J. S. de S.** Modelos mistos na predição de valores genéticos aditivos em testes de progênies florestais. Piracicaba, ESALQ. 1997. 118p. (Tese de Doutorado).
- CRUZ, C. D.; VENCOVSKY, R.** Comparação de alguns métodos de análise dialélica. *Revista Brasileira de Genética*. Ribeirão Preto. V.12, n.2, p.432-38, 1989.
- CRUZ, C. D.; REGAZZI, A. J.** Modelo biométricos aplicados ao melhoramento genético. Viçosa, MG: UFV, Imprensa Uiversitária, 390p. 1994.
- DRAPER, N. R.; SMITH, H.** *Applied regression analysis*. 2 ed. New York: J. Wiley, 1981. 709p.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIM, D. B.** Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Ser. B.*, Oxford. v.39, n.1, p.1-38. 1977.
- DUARTE, J. M.** Estudo da divergência genética em raças de feijão por meio de marcadores RAPD. Lavras: UFLA. 1998. 78p. (Dissertação de Mestrado- Genética e Melhoramento de Plantas).
- EUCLYDES, R. F.** Uso do sistema para simulação *Genesys* na avaliação de métodos de seleção clássicos e associados a marcadores moleculares. Viçosa: UFV, 1996. 150p. (Tese de Doutorado).

- FALCONER, D. S. **Introduction to quantitative genetics**. 2nd Edition. Longman Group Limited, New York. 1981, 661p.
- FERREIRA, P. E. On Jinks-Hayman's analysis of half dialles. **Revista Brasileira de Genética**. Ribeirão Preto. v.8, n.1, p.149-55. 1985.
- GARDNER, C. O.; EBERHART, S. A. Analysis and interpretation of the variety cross diallel and related population. **Biometrics**. Raleigh, v.22, n.3, p.439-52. 1966
- GERALDI, I. O.; MIRANDA FILHO, J. B. de. Adapted models for the analysis of combining ability of varieties in partial diallel crosses. **Revista Brasileira de Genética**. Ribeirão preto, v.11. p.419-30. 1988.
- GIANOLA, D.; FOULLEY, J. L. Variance estimation from integrated likelihood (VEIL). **Génétique, Sélection, Évolution**, Paris, v.22, p.403-17, 1990.
- GRAYBILL, F. A. **Theory and applications of linear model**. Massachusetts, Duxbury Press, 1976. 704p.
- GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences**, Melbourne, v.9, p.463-93, 1956.
- HALLAUER, S. R.; MIRANDA FILHO, J. B. de. **Quantitative genetics in maize breeding**. Ames, Iowa State University Press, 1988. 468p.
- HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, London, v.54, n.1, p.93-108, 1967.

- HAYMAN, B. I. The analysis of variance of diallel tables. **Biometrics**, Raleigh, v.10, n.2, p.235-44. 1954a
- HAYMAN, B. I. The theory and analysis of diallel crosses. **Genetics**, Baltimore, v.39, n.6, p.789-809. 1954b
- HAYMAN, B. I. The theory and analysis of diallel crosses II. **Genetics**, Baltimore, v.43, n.1, p.63-85. 1958.
- HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**, Raleigh, v.17, n.1, p.226-52, jan.-mar., 1953.
- HENDERSON, C. R. General flexibility of linear models for sire evaluation. **J. Dairy Sci.**, Champaign, v.57, n.8, p.963-72, aug., 1974.
- HENDERSON, C. R. Best Linear Unbiased Estimation and Prediction under a selection model. **Biometrics**, Raleigh, v.31, n.2, p.423-447, apr.-jun., 1975a.
- HENDERSON, C. R. Comparison of alternative sire evaluation methods. **J. Anim. Sci.**, Menasha, v.41, n.3, p.760-7, mar., 1975b.
- HENDERSON, C. R. **Applications of linear models in animal breeding.** Ontario, University of Guelph, 1984. 462p.
- HENDERSON, C. R. Recent developments in variance and covariance estimation. **J. Anim. Sci.** Menasha, v.63, p.208-216. 1986.
- HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; VON KROSIG, C. M. Estimation of environmental and genetic trends from records subjecting to culling. **Biometrics**, Raleigh, v.15, n.1, p.192-218, 1959.

- HILL, R. R.; ROSENBERG, J. L. Methods for combining data from germoplasm evaluation trials. *Crop. Sci., Madison*, v.25, n.2, p.267-40, mar.-apr., 1985.
- HUANG, H.; HARDING, J.; BYRNE; FAMULA, T. Estimation of long-term genetic improvement for gerbera using the best linear unbiased prediction (BLUP) procedure. *Theor. Appl. Genet.*, v.91, n.3, p.790-94. 1995.
- JENKINS, M. T. Methods of estimating the performance of double crosses in corn. *Journal of American Society of Agronomy. Madison*, v.26, n.1, p.199-204. 1934.
- JEYARUBAN, M. G.; GIBSON, J. P.; GOWE, R. S. Comparison of index selection and Best Linear Unbiased Prediction for simulated layer poultry data. *Poultry Science*. v.74, n.7. p.1566-1576, 1995.
- JINKS, J. L. The analysis of continuous variation in an diallel cross of *Nicotiana rustica* varieties. *Genetics, Baltimore*, v.39, n.6, p.767-88, 1954.
- JINKS, J. L. The F₂ and backcross generation from um set of diallel crosses. *Heredity, Edinburgh*, v.10, n.1, p.1-30, 1956
- JINKS, J. L.; HAYMAN, B. I. The analysis of diallel crosses. *Mayze Genetics Cooperation Newsletter, Ithaca*, v.37, n.1, p.48-54, 1953.
- JONES, R. M. Analysis of variance of the half diallel table. *Heredity, Edinburgh*, v.20, n.1, p.117-21. 1965.
- KEMPTORN, O.; CURNOW, R. N. The partial diallel crosses. *Biometric Raleigh*. v.17. n.1. p.229-50. Jan-Feb. 1961.

- LAIRD, N. LANGE, N.; STRAM, D.** Maximum likelihood computation with repeated measures: application of EM algorithm. *Journal of the American Statistical Association*, Chicago, v.82. n.1. p.97-105. 1987
- LOPES, P. S.; MARTINS, E. N.; SILVA, M. A. RAGGI, L. A.** Métodos de resolução de sistemas de equações lineares. Viçosa, MG, UFV - Imp. Univ., 1993. 55p.
- MARTINS, E. N.** Desenvolvimento de uma estratégia computacional para seleção de coelhos, usando a melhor predição linear não-viesada. Viçosa, UFV, 1995. 117p. (Tese de Doutorado).
- MICHELMORE, R. W.; KESSEL, R. V.; RIDER, E. J.** Genetic mapping in lettuce. In: **PHILLIPS, R. L.; VASIL, I. K.** (Ed) DNA-based markers in plants. Dordrecht: Klumer Academic Publishers, 1994. p.223-239.
- MIRANDA FILHO, J. B. de.; GERALDI, I. O.** An adapted model for the analysis of partial diallel crosses. *Revista Brasileira de Genética*. Ribeirão Preto. v.7. p.677-88, 1984.
- NEJATI-JAVAREMI, A.; SMITH, C.; GIBSON, J. P.** Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.*, Menasha, v.75, n.7. p.1738-45, July., 1997.
- PANTER, D. M.; ALLEN, F. L.** Using best linear unbiased predictions to enhance breeding for yield in soybean: I Choosing Parents. *Crop Sci.*, Madison, v.35, n.2, p.397-405, mar.-apr., 1995a.
- PANTER, D. M.; ALLEN, F. L.** Using best linear unbiased predictions to enhance breeding for yield in soybean: II Selection of superior crosses from a limited number of yield trials. *Crop Sci.*, Madison, v.35, n.2, p.405-10, mar.-apr., 1995b.

- PATERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, London, v.58, p.545-54, 1971.
- RAMALHO, M. A. P.; SANTOS, J. B. dos; ZIMMERMAN, M. J. de O. **Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro**. Goiânia: UFG, 1993. 271p.
- RAO, C. R. Estimation of heteroscedastic variances in linear models. *J. Am. Statistical Ass.*, v.65, n.1, p.161-72, 1970.
- RAO, C. R. Estimation of variance and covariance components. *J. Mult. Anal.*, v.1, n.1, p.257-75, 1971a.
- RAO, C. R. Minimum variance quadratic unbiased estimation of variance components. *J. Mult. Anal.*, v.1, n.2, p.445-56, 1971b.
- RESENDE, M. D. V. Melhoramento genético de essências florestais. In: **Simpósio sobre atualização em genética e melhoramento de plantas**. Lavras, MG. p.59-93. 1997.
- RESENDE, M. D. V.; PRATES, D. F.; JESUS, A.; YAMADA, C. K. Melhor predição linear não viciada (BLUP) de valores genéticos no melhoramento de Pinus. *Boletim de Pesquisa Florestal*, Colombo-PR, v.32/33, p.3-22, jan.-dez., 1996a.
- RESENDE, M. D. V.; PRATES, D. F.; JESUS, A.; YAMADA, C. K. Estimção de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (REML) e melhor predição linear não viciada em Pinus. *Boletim de Pesquisa Florestal*, Colombo-PR, v.32/33, p.23-39, jan.-dez., 1996b.

- SAS Institute. SAS language and procedure: Usage. Version 6. 1st ed. Inst., Cary, North Caroline. 1995.**
- SEARLE, S. R. Linear models. New York: John Wiley & Sons, 1971. 532p.**
- SEARLE, S. R. Linear models for unbalanced data. New York: John Wiley & Sons, 1987. 536p.**
- SEARLE, S.; CASELA, G.; McCULLOCH, C. E. Variance components. John Willey & Sons, 1992. 501p.**
- SOH, A. C. Ranking parents by best linear unbiased prediction (BLUP) breeding values in oil palm. Euphytica, v.76, n.1, p.13-21. 1994.**
- SORENSEN, D. A.; KENNEDY, B. W. Estimation of response to selection using least-square and mixed model methodology. J. Anim. Sci., Menasha, v.58, n.4, p.1097-106, apr., 1984.**
- SORENSEN, D. A.; KENNEDY, B. W. Analysis of selection experiments using mixed models methodology. J. Anim. Sci., Menasha, v.63, n.2, p.245-58, feb., 1988.**
- SOUSA, E. A. Controle genético de alguns caracteres do feijoeiro. Lavras: ESAL, 1993. 110p. (Dissertação de Mestrado em Genética e Melhoramento de Plantas).**
- SPRAGUE, G. F.; TATUM, L. A. General vs specific combining ability in single crosses of corn. Journal of the American Society of Agronomy, Washington, v.34, n.3, p.923-32, mar., 1942.**
- STELL, R. G. D.; TORRIE, J. H. Principles and procedures of statistics: A biometrical approach. McGraw-Hill. 2 ed. 1981. 633p.**

- THOMPSON, R. The problem of negative estimates of variance components. *Annals of Mathematics Statistics*, Hayward. v.33, p.273-5, 1962.
- VAN VLECK, L. D. Variance of prediction error with mixed model equations when relationships are ignored. *Theor. Appl. Genet.* v.85, p.545-549. 1993a.
- VAN VLECK, L. D. Selection index na introduction to mixed model methods for genetic improvement of animals: the green book. CRC Press. Boca Raton, Florida. 1993b. 481p.
- VERNEQUE, R. da S. Procedimentos numéricos e estimação de componentes de covariância em análise multivariada pelo método da máxima verossimilhança restrita - modelos mistos aplicados ao melhoramento animal. Piracicaba, ESALQ. 1994. 157p. (Tese de Doutorado).
- VENCOVSKY, R.; BARRIGA, P. *Genética biométrica no fitomelhoramento*. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 486p.
- WHITE, T. L.; HODGE, G. R. *Predicting breeding values with applications in forest tree improvement*. Kluwer Academic Publishers: Netherlands. 1989. 367p.
- WOOD, C. M.; CHRISTIAN, L. L.; ROTHSFIELD, M. F. Use of animal model in situations of limited subclass numbers and high degrees of relationships. *J. Anim. Sci., Menasha*, v.69, n.6, p.1420-7, jun., 1991.
- ZHU, J.; WEIR, B. Diallel analysis of sex-linked and maternal effects. *Theor. Appl. Gen.*, v.92, n.1, p.1-9, jan., 1996.

...the problem of ...
...the ...

...of prediction error ...
...the ...

...of ...
...the ...

...of ...
...the ...

ANEXO

...of ...
...the ...

...of ...
...the ...

...of ...
...the ...

...of ...
...the ...

...of ...
...the ...

These April 2000, vol. 1, p. 1-9, Jan. 1999

[REDACTED]

PROGRAMA SAS DESENVOLVIDO PARA REALIZAÇÃO DE ANÁLISE DE
DIALELOS MÉTODOS II E IV DE GRIFFING SOB ABORDAGEM BLUP COM
ESTIMATIVAS DOS COMPONENTES DE VARIÂNCIAS PELO MÉTODO DA
MÁXIMA VEROSSIMILHANÇA RESTRITA (REML).

```

options nodate nonumber ps=1000 ls=90;
data dados;
  input efix p1 p2 cruz y1;
  cards;

;
run;
proc sort;
  by cruz;
proc means noprint;
  by cruz;
  var p1 p2;
  output out=dados1 mean=;
run;
proc iml;
START REML;
  n=nrow(x1);
  /*conta o número de observações */
  x=design(x1);
  /*cria a matriz de incidência dos efeitos fixos*/
  nf=ncol(x);
  /*conta o número de efeitos fixos*/
  vp1=unique(p1);
  vp2=unique(p2);
  vp=union(vp1, vp2) `;
  q1=nrow(vp);
  /*conta o número de pais */
  z1=j(n, q1, 0);
  do i=1 to q1;
    do j=1 to n;
      if p1[j,1]=vp[i,1] then z1[j,i]=z1[j,i]+1;
      if p2[j,1]=vp[i,1] then z1[j,i]=z1[j,i]+1;
    end;
  end;
  /*cria a matriz de incidência de pais (CGC) */
  z2=design(p3);
  /*cria a matriz de incidência para a CEC*/
  q2=ncol(z2);
  /*conta o número de cruzamentos*/
  na=q1+q2;
  /*define o número de efeitos aleatórios*/
  z=(z1||z2);
  /*cria a mat. de incidência dos efeitos aleat.*/
  k=(x||z);
  /*cria a matriz de incidência total*/
  a=1945;
  /*valores iniciais dos componentes da variância*/
  b=1260;
  r=12129;

```

```

iter=0;
/*inicializa o contador de iterações*/
maxiter=2000;
/*define o número máximo de iterações*/
conv=1E-5;
/*define o critério de convergência*/
do until (dif<conv);
  g=block((g1)/a, (g2)/b);
  /*define a matriz G inversa*/
  t=inv(((x`*x)|| (x`*z))//((z`*x)|| (z`*z+g`*r)));
  /*concatena e inverte a matriz núcleo (LHS) */
  solucao=t*(x`*y//z`*y);
  /*solução do sistema*/
  cgc=solucao[nf+1:nf+q1];
  /*solução do para a CGC*/
  cec=solucao[nf+q1+1:nf+q1+q2];
  /*solução do para a CEC*/
  r1=(y`*y-solucao`*k`*y)/(n-nf);
  /*componente da variância residual*/
  a1=(cgc`*g1*cgc+trace(g1*(t[nf+1:nf+q1,nf+1:nf+q1]))*r)/(q1);
  /*componente da variância CGC*/
  b1=(cec`*cec+trace(t[nf+q1+1:nf+q1+q2,nf+q1+1:nf+q1+q2])*r)/(
q2);
  /*componente da variância CEC*/;
  dif=max(abs(r1-r), abs(a1-a), abs(b1-b));
  /*cálculo da diferença entre as estimativas*/
  if iter=maxiter then dif=0;
  /*condição para o máximo de iterações*/
  if iter<1 then do;
    var_cgc=a;
    var_cec=b;
    erro=r;
    difer=dif;
    n_iter=iter;
    razao=raz;
  end;
  else do;
    var_cgc=var_cgc//a;
    var_cec=var_cec//b;
    erro=erro//r;
    difer=difer//dif;
    n_iter=n_iter//iter;
    razao=razao//raz;
  end;
  iter=iter+1;
  a=a1;
  b=b1;
  r=r1;
end;
print , 'Estimativas dos Componentes da Variância (REML)',
/* imprime as estimativas dos componentes de variância*/
  n_iter [format= 3.]
  var_cgc [format= 12.6]
  var_cec [format= 12.6]
  erro [format= 12.6]

```

```

                difer    [format= 12.8]
                razao    [format= 12.8];
stder=sqrt(vecdiag(t)*r);
t_val=solucao/stder;
gl=n-nf-q2+1;
p_alfa=1-probf(t_val#t_val,1,gl);
cod=(unique(x1))`//(unique(p1))`//(unique(p3))`;
ef=repeat('efix',nf,1)//repeat('  g',q1,1)//repeat('
s',q2,1);
gl=j((nf+q1+q2),1,gl);
print , 'Solução dos efeitos fixos e Predição dos Gis e Sijs
'
```

```

                ef        [format= 4.]
                cod       [format= 3.]
                solucao  [format= 12.7]
                stder    [format= 10.5]
                gl       [format= 4.]
                t_val    [format= 7.4]
                p_alfa   [format= 8.6];
```

```

FINISH;
use dados;
read all into data;
x1=data(1,1);
/*define o vetor de efeitos fixos*/
p1=data(1,2);
/*define o vetor de efeitos do pai 1 (CGC)*/
p2=data(1,3);
/*define o vetor de efeitos do pai 2 (CGC) */
p3=data(1,4);
/*define o vetor de efeitos da CEC*/
y=data(1,5);
/*define o vetor de obsevações*/
g0={matriz de similaridade entre as linhagens};
gl=inv(g0);
use dados1;
read all into data;
do i=1 to nrow(b);
  do j=1 to nrow(b);
    if i^=j then
g00[i,j]=((g0[b[i,1],b[j,1]]*g0[b[i,2],b[j,2]]+g0[b[i,1],b[j,
2]]*g0[b[i,2],b[j,1]]))/4;
end;
  end;
g2=inv(g00);
RUN REML;
quit;
```