

.UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

MARCOS DE ARAUJO RODOVALHO

INFERÊNCIA BAYESIANA E MODELOS MISTOS NA CULTURA DO
MILHO: APLICAÇÕES EM DIALELOS, ADAPTABILIDADE E
ESTABILIDADE E GANHOS POR SELEÇÃO

MARINGÁ
PARANÁ – BRASIL
MAIO - 2012

MARCOS DE ARAUJO RODOVALHO

INFERÊNCIA BAYESIANA E MODELOS MISTOS NA CULTURA DO
MILHO: APLICAÇÕES EM DIALELOS, ADAPTABILIDADE E
ESTABILIDADE E GANHOS POR SELEÇÃO

Tese apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração em Produção Vegetal, para a obtenção do Título de Doutor.

MARINGÁ
PARANÁ – BRASIL
MAIO – 2012

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central – UEM, Maringá – PR., Brasil)

R695i Rodovalho, Marcos de Araujo
Inferência bayesiana e modelos mistos na cultura do milho : aplicações em dialelos, adaptabilidade e estabilidade e ganhos por seleção / Marcos de Araujo Rodovalho. -- Maringá, 2012.
96 f. : il.

Orientador: Prof. Dr. Carlos Alberto Scapim.
Tese (doutorado) - Universidade Estadual de Maringá, Centro de Ciências Agrárias, Departamento de Agronomia, Programa de Pós-Graduação em Agronomia, 2012.

1. Milho - Melhoramento genético. 2. Milho - Genética quantitativa. 3. Teoria Bayesiana. I. Scapim, Carlos Alberto, orient. II. Universidade Estadual de Maringá. Centro de Ciências Agrárias. Departamento de Agronomia. Programa de Pós-Graduação em Agronomia. III. Título.

CDD 22.ed. 633.15

MARCOS DE ARAUJO RODOVALHO

INFERÊNCIA BAYESIANA E MODELOS MISTOS NA CULTURA DE MILHO:
APLICAÇÕES EM DIALELOS, ADAPTABILIDADE E ESTABILIDADE E
GANHOS POR SELEÇÃO

Tese apresentada à Universidade Estadual de Maringá, como parte das exigências do Programa de Pós-Graduação em Agronomia, área de concentração Produção Vegetal, para a obtenção do Título de Doutor.

_____ em 15 de maio de 2012

Prof. Dr. Ronald Barth Pinto

Prof. Dr. Marcos Ventura Faria

Prof. Dr. Carlos Antônio Lopes de Oliveira

Prof. Dr. Freddy Mora Poblete

Prof. Dr. Carlos Alberto Scapim
(Orientador)

Ao meus pais, que incentivaram meus estudos desde o princípio.

DEDICO.

AGRADECIMENTOS

À Capes, pela concessão da bolsa de Doutorado.

Ao Programa de Pós-Graduação em Agronomia da Universidade Estadual de Maringá, pela oportunidade concedida.

À secretária do Programa de Pós-Graduação em Agronomia, Érika Sato, pelo indispensável.

Aos professores doutores Cássio Antonio Tormena, Pedro Soares Vidigal Filho e Ronald José Barth Pinto, pela amizade e conselhos durante o decorrer do curso.

Ao professor doutor Carlos Alberto Scapim, pela orientação durante o Curso de Graduação em Agronomia, com projetos de iniciação científica, e por toda a colaboração durante o Mestrado e o Doutorado.

BIOGRAFIA

MARCOS DE ARAÚJO RODOVALHO, filho de Otacílio Gonçalves Rodovalho e Nunciata de Araújo Rodovalho, nasceu em Campo Grande, Mato Grosso do Sul, aos 17 dias do mês de julho de 1980.

Em março de 2001, iniciou o Curso de Agronomia na Universidade Estadual de Maringá (UEM). No período de julho de 2003 a março de 2006, participou de diversos projetos de pesquisa na área de melhoramento de milhos especiais, sob a orientação do professor doutor Carlos Alberto Scapim e coorientação do professor doutor Ronald José Barth Pinto, como bolsista do Programa de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico (CNPq).

Graduou-se em Agronomia, em 17 de fevereiro de 2006. No mesmo mês e ano, ingressou no Curso de Pós-Graduação em Agronomia, em nível de Mestrado, área de concentração em Produção Vegetal, na Universidade Estadual de Maringá, obtendo o título de Mestre em fevereiro de 2008. Em março do mesmo ano, ingressou no Curso de Doutorado da mesma Instituição de Ensino Superior.

Atualmente, trabalha na empresa Syngenta Seeds como Assistente de Melhorista.

SUMÁRIO

LISTA DE TABELAS	vii
LISTA DE FIGURAS	ix
RESUMO	x
ABSTRACT	xi
CAPÍTULO 1-REVISÃO BIBLIOGRÁFICA	1
1.1 Cultura do milho	1
1.2 Avaliação genética usando inferência bayesiana.....	1
1.2.1 O Método Monte Carlo.....	3
1.2.2 O Método da cadeia de Markov	4
1.2.3 Método Monte Carlo via cadeia de Markov	7
1.2.4 O algoritmo de Metropolis-Hastings	7
1.2.5 Intervalo de credibilidade	11
1.3 Modelos lineares mistos no melhoramento genético.....	12
1.3.1 Estimadores de componentes de variâncias.....	14
1.3.2 Importância da escolha do estimador de variância na predição de valores genéticos	15
1.3.3 Predição de valores genéticos.....	16
1.3.4 Aplicativo computacional para REML-BLUP	17
1.3.5 Adaptabilidade e estabilidade	18
1.3.6 Média Harmônica do Valor Genético (MHVG).....	18
1.3.7 Performance relativa do valor genético (PRVG).....	19
1.3.8 Média Harmônica da Performance Relativa do valor Genético predito (MHPRVG).....	19
1.3.9 Análise de Deviance (Anodev).....	20
REFERÊNCIAS	21
CAPÍTULO 2 -ANÁLISE DIALÉLICA DE LINHAGENS DE MILHO COM ENFOQUE BAYESIANO	27
RESUMO	27
1 INTRODUÇÃO	29
2 MATERIAL E MÉTODOS	31
2.1 Material genético	31
2.2 Análise dialélica	32
2.3 Aplicação da técnica de MCMC.....	32

3 RESULTADOS E DISCUSSÃO	34
4 CONCLUSÕES.....	43
5 REFERÊNCIAS	44
CAPÍTULO 3 - AVALIAÇÃO GENOTÍPICA DE HÍBRIDOS SIMPLES DE MILHO EM MULTILOCAIS POR MEIO DE MODELOS MISTOS.....	46
RESUMO	46
1 INTRODUÇÃO	48
2 MATERIAL E MÉTODOS	50
2.1 Locais de avaliação.....	50
2.2 Genótipos avaliados.....	50
2.3 Modelo de avaliação genética.....	50
3 RESULTADOS E DISCUSSÃO	52
3.1. Seleção genotípica em todos os locais.....	54
3.2 Avaliações genotípicas com base em cada local	56
3.3 Análise de estabilidade dos genótipos avaliados por meio da Média Harmônica dos Valores Genotípicos	66
3.4 Análise de adaptabilidade por meio de performance relativa dos valores genotípicos.	67
3.5 Análise de adaptabilidade e estabilidade por meio da Média Harmônica da Performance Relativa dos Valores Genotípicos (MHPRVG).....	67
3.6 Acurácia na seleção de genótipos	71
4 CONCLUSÕES.....	73
5 REFERÊNCIAS	74
CAPÍTULO 4 - AVALIAÇÃO GENÉTICA DE FAMÍLIAS DE MEIOS-IRMÃOS DE MILHO PIPOCA USANDO A METODOLOGIA BAYESIANA.....	76
RESUMO	76
1 INTRODUÇÃO	78
2 MATERIAL E MÉTODOS	80
3 RESULTADOS E DISCUSSÃO	83
4 CONCLUSÃO.....	88
5 REFERÊNCIAS	89

LISTA DE TABELAS

Tabela 1	Relação das linhagens utilizadas nos cruzamentos.....	31
Tabela 2	Esquema de cruzamentos das 10 linhagens de milho para a geração dos híbridos simples.....	32
Tabela 3	Estimativas <i>a posteriori</i> e intervalo HPD de 95% para os efeitos genéticos do quadro de cruzamentos dialélicos, Capacidade Geral e Específica de Combinação (\hat{G}_i e \hat{S}_{ij}) por meio do método de MCMC com 100.000 iterações e “burn in” de 1.000, para a característica rendimento de grãos (MCSE: erro-padrão de Monte Carlo; SD: Desvio-padrão).....	36
Tabela 4	Estimativas dos componentes de variância (REML) e de parâmetros genéticos para rendimento de grãos kg ha ⁻¹	53
Tabela 5	Anodev (Análise de Deviance) componentes de variância, coeficientes de determinação referentes à análise conjunta (todos os locais e genótipos).....	54
Tabela 6	Valores de efeitos genótipos, valores genotípicos preditos (u + g), ganho com a nova média, médias melhoradas, e valores genotípicos médios nos vários ambientes u + g + gem.....	55
Tabela 7	Ordenamento dos genótipos avaliados nos locais 1 e 2, onde são apresentados os efeitos genotípicos (g + ge), os valores genotípicos preditos para os respectivos locais (u + g + ge).....	57
Tabela 8	Ordenamento dos genótipos avaliados nos locais 3e 4 onde são apresentados os efeitos genotípicos (g + ge), os valores genotípicos preditos para os respectivos locais (u + g + ge).....	59

Tabela 9	Ordenamento dos genótipos avaliados nos locais 5 e 6, onde são apresentados os efeitos genotípicos ($g + ge$), os valores genotípicos preditos para os respectivos locais ($u + g + ge$).....	61
Tabela 10	Ordenamento dos genótipos avaliados nos locais 7 e 8, onde são apresentados os efeitos genotípicos ($g + ge$), os valores genotípicos preditos para os respectivos locais ($u + g + ge$).....	63
Tabela 11	Ordenamento dos genótipos avaliados nos locais 1 e 2, onde são apresentados os efeitos genotípicos ($g + ge$), os valores genotípicos preditos para os respectivos locais($u + g + ge$).....	69
Tabela 12	Valores genéticos e limites para características rendimento de grãos.....	71
Tabela 13	Critérios de informação de Akaike (AIC) e Bayesiano de Schwarz (BIC), o teste da razão de verossimilhança (LRT) (comparando os modelos: nulo e completo) e estimativas paramétricas usando modelos lineares mistos, via Máxima Verossimilhança Restrita (REML).....	83
Tabela 14	Estimativas bayesianas dos componentes de variância e herdabilidade (nível de famílias) para o rendimento de grãos e capacidade de expansão em milho pipoca.....	84
Tabela 15	Estimativas do ganho genético (seleção familiar) considerando duas intensidades de seleção (20,12 e 10,06%) calculadas de acordo com os valores preditos Bayesianos das famílias de milho pipoca.....	85

LISTA DE FIGURAS

Figura 1	Traços de qualidade desejável para o histórico de simulações MCMC.....	9
Figura 2	Traços de qualidade indesejável para o histórico de simulações MCMC.....	10
Figura 3	Estimativa da capacidade geral de combinação <i>a posteriori</i> dos genótipos avaliados: linhagens 1, 2, 3, 4, 5 e 6.....	39
Figura 4	Estimativa da capacidade geral de combinação <i>a posteriori</i> dos genótipos avaliados: linhagens 7, 8, 9 e 10.....	40
Figura 5	Distribuições marginais <i>a posteriori</i> das herdabilidades (nível de famílias), para as características capacidade de expansão e rendimento de grãos, avaliadas em 169 famílias de milho pipoca. À esquerda é mostrado o processo de iteração do algoritmo de cadeias independentes (análise Bayesiana).....	86
Figura 6	Gráfico de dispersão com os valores genéticos de famílias, obtidos via predição Bayesiana para capacidade de expansão (CE) e rendimento de grãos, avaliados em 169 famílias de meios-irmãos de milho pipoca (população UEM-C1).....	87

RESUMO

RODOVALHO, Marcos de Araujo., D.S. Universidade Estadual de Maringá, Maio de 2012. **Inferência bayesiana e modelos mistos na cultura de milho: aplicações em dialelos, adaptabilidade e estabilidade e ganhos por seleção.** Orientador: Carlos Alberto Scapim. Coorientadores: Ronald José Barth Pinto e Freddy Mora Poblete.

O presente trabalho teve por objetivo aplicar a inferência bayesiana e modelos mistos em dialelo, adaptabilidade e estabilidade e ganhos por seleção na cultura do milho. O método Bayesiano foi implementado por meio do algoritmo de Cadeias de Independência (*Independence Chain*), com informações *a priori* informativas para os componentes de variância familiar e residual (distribuição *a priori* Inversa Gamma). As estimativas dos parâmetros genéticos da capacidade de expansão e rendimento de milho pipoca em um ensaio de 169 famílias de meios-irmãos, por meio da metodologia Bayesiana como preditor do mérito genético, demonstrou que a população em estudo, possui variabilidade genética significativa para os ciclos subsequentes de melhoramento e que a metodologia Bayesiana pode auxiliar no processo de seleção de famílias de milho pipoca. Para as estimativas de capacidades geral e específica de combinação do rendimento de grãos de dez linhagens de milho por meio de cruzamentos dialélicos indicaram que as linhagens 6, 5, 9 e 10 mostraram-se promissoras para serem utilizadas no programa de melhoramento. As combinações híbridas 2x6, 1x8, 1x5 foram as que mais destacaram-se para rendimento de grãos, com as maiores estimativas de capacidade específica de combinação. Para a adaptabilidade e a estabilidade de 43 híbridos simples de milho em ensaios de rede, pela metodologia de modelos lineares mistos. De acordo com os valores genotípicos, os dez melhores híbridos, considerando os efeitos interação de ambiente e a média genotípica, foram: 14, 32, 13, 4, 34, 6, 33, 22, 10 e 37. Conclui-se que a metodologia de modelos lineares mistos é alternativa útil e adequada para a avaliação genotípica de híbridos simples de milho em multilocais.

Palavras-chave: *Zea mays* L., adaptabilidade/estabilidade, dialelo, bayesiana, modelos mistos.

ABSTRACT

RODOVALHO, Marcos de Araujo., D.S. Universidade Estadual de Maringá, May 2012. **Bayesian inference and mixed models in crop corn: applications diallels, adaptability / stability and gains by selection.** Advisor: Carlos Alberto Scapim. Committee Members: Ronald José Barth Pinto e Freddy Mora Pobleto.

This study aimed to apply the Bayesian inference and mixed models in diallel, adaptability and stability and selection gains in corn. The Bayesian method has been implemented through the Chains of Independence (*Independence Chain*) algorithm, with the informative prior information for the components of family and residual variance (Inverse Gamma *prior* distribution). Estimates of genetic parameters of capacity expansion and yield of popcorn in an essay of 169 half-sib families, through the Bayesian methodology as a predictor of genetic merit, demonstrated that the study population, has significant genetic variability for cycles subsequent improvement and that the Bayesian methodology can assist in the selection of families popcorn process. For estimates of general and specific combining ability of the yield of ten lines of maize using diallel indicated that strains 6, 5, 9 and 10 were promising for use in the breeding program. The hybrid combinations 2x6, 1x8 and 1x5 were the ones that stood out most to yield, with the highest estimates of specific combining ability . For adaptability and stability of 43 corn simple hybrids in trials network, the methodology of mixed linear models. According to the genotypic values, the ten best hybrids, considering the interaction of genotype and environment effects medium were: 14, 32, 13, 4, 34, 6, 33, 22, 10 and 37 . It is concluded that the methodology of linear mixed models is helpful and suitable for genotypic evaluation of corn hybrids in multisite alternative.

Keywords: Zea mays L., adaptability/stability, diallel, Bayesian, mixed models.

CAPÍTULO 1

REVISÃO BIBLIOGRÁFICA

1.1 Cultura do milho

O Brasil é um dos maiores produtores de milho, com uma área semeada de 15,6 milhões de hectares, equivalente a uma produção de 65,1 milhões de toneladas e produtividade de 4,1 t ha⁻¹ (COMPANHIA NACIONAL DE ABASTECIMENTO, 2012). Dentro da evolução mundial de produção de milho, o Brasil tem se destacado como terceiro maior produtor, ficando atrás apenas dos Estados Unidos e da China.

A produtividade média de grãos de milho no Brasil é considerada baixa, pois se verificam na literatura produtividades superiores a 10 t ha⁻¹ (FERREIRA et al., 2009), obtidas por meio de híbridos de linhagens. Um dos fatores do baixo nível de produtividade, no Brasil, é o grande número de pequenos produtores que cultivam esse cereal. Porém, o aumento da produtividade vem sendo possível graças ao avanço tecnológico proporcionado pelo desenvolvimento de híbridos com genética superior e a adoção de práticas de manejo que proporcionem maior nível de respostas e segurança aos híbridos atualmente comercializados. A área cultivada com milho híbrido atinge milhões de hectares em todo o mundo, sendo o Brasil o segundo país a adotar a tecnologia de variedades híbridas (VENCOVSKY & RAMALHO, 2000; FERREIRA 2009).

1.2 Avaliação genética usando inferência bayesiana

No melhoramento vegetal, o enfoque estatístico mais utilizado na avaliação genética é conhecido como estatística frequentista ou clássica. Entretanto, o procedimento bayesiano, usando diferentes métodos de Monte-Carlo, via Cadeias de Markov (MCMC), é indicado como alternativa ao método de estimação/predição frequentista.

O termo "bayesiano" vem do uso predominante do teorema de Bayes, criado pelo reverendo Thomas Bayes, no século XVIII. O primeiro exemplo de inferência estatística bayesiana, isto é, "a razão por trás dos efeitos de causas", foi publicada por Thomas Bayes, em 1763, quando definiu pela primeira vez a probabilidade condicional, observando que o resultado, agora denominado Teorema Bayesiano, era uma consequência trivial de sua

definição. A obra de Thomas Bayes só foi publicada em 1764, por Richard Price, na Royal Society of London (BAYES, 1763).

A análise bayesiana, por meio de diferentes métodos, tem sido frequentemente utilizada para realizar inferências científicas e obter conclusões preditivas em diversos experimentos biológicos. Quando ocorre uma junção entre as técnicas frequentistas e bayesianas, esta última contribui, ampliando as informações sobre determinado parâmetro e subsidiando maior segurança na tomada de decisão, pois permite uma conclusão mais acurada e assertiva (BALDING, 2002; BAYARRI & BERGER, 2004; MOON & SCHORFHEIDE, 2011).

Os métodos frequentistas assumem que os parâmetros desconhecidos são constantes fixas, definindo a probabilidade como frequências relativas. Decorre desta premissa que as probabilidades são objetivas, não podendo inferir afirmações probabilísticas sobre os parâmetros, pois os mesmos são fixos (EAVES & ERKANLI, 2003; HOTI & SILLANPAA, 2006; BOX & TIAO, 1973).

Na teoria bayesiana, qualquer incerteza sobre os parâmetros desconhecidos de um modelo determinístico é expressa em termos de probabilidades (CORDEIRO, 1999). Trata-se de um método que utiliza parâmetros como variáveis aleatórias e probabilidade como “grau de credibilidade” (a probabilidade de um evento ocorrer é o grau de crença de que o evento seja verdadeiro).

Estudos que tem por base a filosofia bayesiana não são determinados como observações pontuais. São utilizadas probabilidades associadas aos parâmetros, assim como as suas distribuições. Logo, a tomada de decisão pode ser realizada por meio de intervalos de confiança, ou regiões de credibilidade no contexto bayesiano (WEERAHANDI & ZIDEK, 1981; GOLDSTEIN & WOOF, 2007).

Na abordagem Bayesiana, a inferência sobre θ (parâmetro de interesse) é baseada na distribuição *a posteriori* $f(\theta|Y)$, mediante a aplicação do teorema de Bayes com a informação *a priori* $\pi(\theta)$ e da distribuição de amostragem $f(Y|\theta)$, para o cálculo da probabilidade:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)\pi(\theta)}{p(y)} = \frac{p(y|\theta)\pi(\theta)}{\int (y|\theta)\pi(\theta)d\theta}$$

O denominador $p(y)$ é a probabilidade marginal para os dados, tendo em vista que o mesmo integra os parâmetros da função de verossimilhança:

$$p(y) = \int p(y|\theta)\pi(\theta) d\theta$$

O termo *a posteriori* $p(\theta|y)$ incorpora ambas as informações *a priori* e as informações dos dados.

Simplificando, o teorema de Bayes atualiza os conhecimentos existentes com novas informações, ou seja, dados novos atualizam o estado atual de conhecimento, sendo o resultado um novo estado de conhecimento, representado pela distribuição *a posteriori*. Estes são os elementos essenciais da abordagem Bayesiana para a análise dos dados (O'HARA & SILLANPAA, 2009).

Atualmente, tanto os modelos frequentistas quanto os modelos bayesianos estão em ampla utilização. Nas últimas décadas, o uso de análise bayesiana era bem inferior à análise frequentista, como pode ser observado pela proporção de artigos científicos publicados para cada assunto. No entanto, este cenário vem mudando e o uso da abordagem bayesiana tem ganhado cada vez mais espaço nos artigos científicos publicados em revistas especializadas (VALLVERDÚ, 2008). O fato principal que estimulou o aumento do uso da abordagem bayesiana foi o desenvolvimento da ciência da computação, possibilitando a inserção de novas metodologias para que a abordagem bayesiana fosse usada na solução de problemas (GAMMERMAN, 1997).

De acordo com Gelfand & Smith (1990), as distribuições *a posteriori* dos parâmetros de interesse são impossíveis de se obter pela usual via computacional. Devido a isto, são utilizados os métodos de Monte Carlo, via cadeias de Markov (MCMC), para resolver as integrações numéricas e permitir a análise das distribuições *a posteriori* (GELFAND & SMITH, 1990; GILKS et al., 1996; METROPOLIS et al., 1953).

1.2.1 O Método Monte Carlo

O método Monte Carlo foi originalmente empregado em aplicações físicas pelos físicos Stam Ulam, John Von Neumann e Roberts Richtmyer em Los Alamos, ligados à fabricação da bomba atômica durante a segunda guerra mundial (MAKIYAMA, 2006). O termo era aplicado como uma palavra-código para as simulações com números aleatórios (ULAM et al., 1947). Monte Carlo é, em essência, a aproximação de uma expectativa pela amostra média de uma função de variáveis aleatórias simuladas (FAN et al., 2002). Esta definição é ampla o suficiente para cobrir tudo o que tem sido denominado como métodos de Monte Carlo (GILKS et al., 1996).

A abordagem original do Método Monte Carlo foi empregada para a geração de números aleatórios e cálculo de integrais complexas:

$$\int_a^b h(x) dx$$

A solução de $h(x)$ gera uma função $f(x)$ e uma função de densidade de probabilidade $p(x)$ definida sobre o intervalo (a; b), em que:

$$\int_a^b h(x) dx = \int_a^b f(x) p(x) dx = E_{p(x)}[f(x)]$$

As igualdades acima demonstram que a integral pode ser expressa como uma esperança de $f(x)$ sobre a densidade de $p(x)$. Assim, x_1, \dots, x_n são variáveis aleatórias de densidade $p(x)$, logo, a expressão a seguir é referida como integração de Monte Carlo:

$$\int_a^b h(x) dx = E_{p(x)}[f(x)] \cong \frac{1}{n} \sum_{i=1}^n f(x_i)$$

O método de Monte Carlo é útil justamente por gerar uma grande quantidade de informações de interesse, as quais podem ser expressas como expectativas (SORENSEN & GIANOLA, 2002).

Enquanto a maioria das simulações de Monte Carlo são feitas por aplicativos computacionais, existem várias aplicações do método que utilizam baralhos ou dados de arremesso de moedas, em vez de sequências de números aleatórios gerados por computador, que datam de séculos atrás, muito antes do surgimento do nome Monte Carlo (GAMMERMAN, 1997).

1.2.2 O Método da cadeia de Markov

O método da cadeia de Markov é basicamente a idéia de dependência dentre estados e é atribuída ao matemático Andrey Andreyevich Markov, que desenvolveu um modelo probabilístico em que os resultados sucessivos dependem apenas do estado estritamente anterior (GAMERMAN & LOPES, 2006).

Este é um caso especial de processo estocástico, que trata de variáveis aleatórias. Um processo estocástico pode ser definido como um conjunto de variáveis aleatórias

$\{\theta^{(t)}: t \in T\}$, em que t é um conjunto de números inteiros positivos consecutivos, o que implica intervalos espaçados no tempo: $\{\theta^{t=0}, \theta^{t=1}, \theta^{t=2}, \dots, \}$.

O processo estocástico também deve ser definido com relação a um estado do espaço θ , que identifica o intervalo de valores possíveis de θ . Este estado do espaço pode ser discreto ou contínuo, dependendo de como a variável de interesse é mensurada.

A cadeia de Markov é um processo estocástico com a propriedade de que qualquer estado especificado na série θ^t dependa apenas do valor anterior da cadeia $\theta^{[t-1]}$. Portanto, condicionalmente, independente de todos os outros valores anteriores: $\theta^{[0]}, \theta^{[1]}, \dots, \theta^{[t-2]}$.

Isso pode ser declarado de maneira formal:

$$P(\theta^t \in A | \theta^0, \theta^1, \dots, \theta^{t-2}, \theta^{t-1}) = P(\theta^t \in A | \theta^{t-1})$$

em que: A é qualquer conjunto completo identificado no estado do espaço.

Assim, uma cadeia de Markov desloca-se em torno do estado e do espaço parâmetro, e como memória tem apenas o que foi no último período. Esta propriedade passa a ser extremamente útil na geração de amostras de distribuições desejada da cadeia, pois, quando a cadeia finalmente encontra a região do estado e espaço com maior densidade, irá produzir uma amostra a partir desta distribuição, que é apenas levemente dependente. Os valores das amostras serão usados para descrever *a posteriori* de interesse.

Um ponto fundamental é o processo de transição entre cadeias, no qual se define a probabilidade para a cadeia se mover para outros pontos no espaço parâmetro, dada a posição atual da cadeia.

A estrutura que define a região K de transição é um mecanismo geral que descreve a probabilidade da cadeia mudar para algum outro estado com base em seu status atual (CASELA & GEORGE, 1992). A vantagem desta definição é que ela congrega tanto os casos contínuos de estado e espaço, assim como os casos de estado e espaço discreto. É necessário que $K(\theta, A)$ seja uma medida de probabilidade definida para todos pontos θ no estado e espaço, para definir $A \in \theta$. Assim, $K(\theta, A)$ mapeia os potenciais eventos de transição para a sua probabilidade ocorrente.

Quando o estado e espaço são discretos, K é mapeado por uma matriz, $k \times k$ para elementos k discretos em A . Nesta matriz, cada combinação ij define a probabilidade de uma transição de estado ocorrer a partir do primeiro termo para todos os estados possíveis:

$$P_A \begin{bmatrix} p(\theta_1, \theta_1) & \dots & p(\theta_1, \theta_k) \\ \vdots & & \vdots \\ p(\theta_k, \theta_1) & \dots & p(\theta_k, \theta_k) \end{bmatrix}$$

A linha indica o período atual em que a cadeia se encontra e a coluna indica para onde a cadeia se deslocará para o próximo período.

A soma da primeira linha de P_A define uma função massa de probabilidade (*PMF*) condicional, onde todos são definidos com o mesmo valor inicial, e cobrem todos os destinos possíveis no estado e espaço para a linha i : $\sum_{j=1}^k p(\theta_i, \theta_j)$. Cada elemento da matriz é uma probabilidade: $p(\theta_i, \theta_j) \geq 0, \forall i, j \in A$.

Quando o estado e espaço são contínuos, K é uma função de densidade de probabilidade condicional, *PDF*: $f(\theta|\theta_i)$, adequadamente definida para todos os $\theta \in A$, dado o estado atual de θ_i . Uma característica importante do núcleo de transição é que as probabilidades de transição entre dois estados selecionados podem ser calculadas para um número arbitrário de passos m . Por exemplo, a probabilidade de transição do estado $\theta_i = x$ no tempo 0 para o estado $\theta_j = y$, em exatamente m passos é fornecida por uma série multiplicativa:

$$p^m(\theta_i^{[0]} = x, \theta_j^{[m]} = y) = \underbrace{\sum_{\theta_1} \sum_{\theta_2} \dots \sum_{\theta_{m-1}}}_{\text{todos os possíveis caminhos}} \underbrace{p(\theta_i, \theta_1)p(\theta_1, \theta_2) \dots p(\theta_{m-1}, \theta_j)}_{\text{produtos de transição}}$$

Assim, $p^m(\theta_i^{[0]} = x, \theta_j^{[m]} = y)$ é também uma matriz de transições estocásticas e essa propriedade é válida para todas as cadeias discretas, exatamente conforme determinado.

Para as cadeias de Markov contínuas, há apenas uma ligeira modificação, pois, em vez de integrais, utilizam somatórios. A ideia básica é que a probabilidade completa para a transição de x para y seja o produto de todos os passos intermediários necessários, na qual se somam todos os caminhos possíveis para que a cadeia passe de x para y .

O método da cadeia de Markov apresenta algumas propriedades, como: ser homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes; ser irreduzível, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações; e por último necessitar ser aperiódica, não haver estados absorventes (HUANG & NEVMYVAKA, 2001).

1.2.3 Método Monte Carlo via cadeia de Markov

O método de Monte Carlo via cadeia de Markov (MCMC) tornou-se uma importante ferramenta da estatística Bayesiana, uma vez que permite inferir sobre distribuições posteriores complexas, nas quais as técnicas de integração analítica ou numérica não podem ser aplicadas (CONGDON, 2003). A ideia subjacente a estes métodos é gerar uma cadeia de Markov por meio iterativo da simulação de Monte Carlo, que tenha pelo menos a distribuição desejada *a posteriori* com equilíbrio e distribuição estacionária (TIERNEY, 1994).

Normalmente, o estimador de Monte Carlo é melhor, pois, para a mesma quantidade de esforço computacional, apresenta menor variação que seus concorrentes. A variância de um estimador de Monte Carlo é facilmente definida por: g_n como uma variável aleatória X e um estimador de Monte Carlo $E[g(X)]$.

Como em todas as variáveis aleatórias, pode-se calcular a variância pela fórmula padrão:

$$\text{Var}(\bar{g}_n(X)) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n g(X^{(i)})\right) = \frac{\text{Var}(g(X))}{n} = \frac{1}{n} \int_{x \in \mathcal{X}} [g(x) - \mathbb{E}(g(X))]^2 f_x(x) dx$$

Os métodos MCMC mais utilizados são o algoritmo de Metropolis-Hastings e amostrador de Gibbs, que é um derivado do primeiro. O princípio básico é simular um movimento aleatório no espaço de θ que converge para uma distribuição estacionária, que é a distribuição de interesse (ROBERT & CASELLA, 2004).

1.2.4 O algoritmo de Metropolis-Hastings

Nicholas Metropolis foi um matemático americano, inventor do algoritmo que leva seu sobrenome (METROPOLIS et al., 1953; METROPOLIS, 1987). Este algoritmo é o mais completo método de MCMC, pois os outros métodos são derivados dele (ROBERT & CASELLA, 2011). O algoritmo foi proposto por Metropolis, em 1953 (METROPOLIS et al., 1953), e generalizado por Hastings em 1970 (HASTINGS, 1970).

A metodologia de Metropolis-Hastings utiliza a mesma idéia de aceitação e rejeição utilizada em metodologias clássicas, em que o valor é gerado por meio de uma distribuição e aceito ou rejeitado em termos de probabilidade (ROBERT & CASELLA, 2004).

A função do algoritmo de Metropolis-Hastings é gerar, via simulação, uma cadeia de Markov dentro de um espaço determinado e, após, atingir uma distribuição estacionária. Só então é realizada a extração das amostras dos valores desejados (BROOKS & ROBERT, 1998).

O algoritmo de Metropolis-Hastings pode ser resumido da seguinte maneira: inicializando o contador de interação, ou seja, $t = 0$ especificando um ponto de início de θ^0 . Este ponto pode ser arbitrário contanto que $f(\theta^0|y) > 0$, o que gera uma nova amostra θ_{novo} pelo uso da distribuição proposta $q(\cdot | \theta^t)$, posteriormente é realizado o cálculo de aceitação da probabilidade desta nova distribuição.

Calcula-se da seguinte maneira:

$$r = \min \left\{ \frac{f(\theta_{novo}|y)}{f(\theta^t|y)}, 1 \right\}$$

Amostra da nova média μ com distribuição uniforme $U(0, 1)$,

Caso $\theta^{t+1} = \theta_{novo}$ se $\mu < r$, aceite o novo valor; caso contrário, rejeite e use $\theta^{t+1} = \theta^t$.

É incrementado o contador de t para $t = t + 1$, disponibilizando um novo valor de θ_{novo} da distribuição proposta $q(\cdot | \theta^t)$. Se, $t < T$ o número de amostras desejadas é retornado ao passo anterior, caso contrário o processo finaliza. Nota-se que o número de repetição continua aumentando mesmo se a amostra é aceita.

Para o bom funcionamento deste amostrador, é necessário possuir uma proposta que seja, no mínimo, próxima da distribuição verdadeira, para conseguir uma taxa de convergência aceitável (BROOKS & ROBERTS, 1998).

Na inferência sobre os dados simulados, para resumir *a posteriori*, pode-se utilizar qualquer distribuição para calcular os parâmetros de interesse, porém isto deve ser após uma série de testes de convergência e a utilização da distribuição real é quem produz os parâmetros de melhor qualidade (KAUFMAN & SAIN, 2010).

Após o início de uma simulação, é realizada a prática de descartar uma parte inicial de uma cadeia de Markov simulada, de tal maneira que nenhuma amostra seja realizada utilizando valores iniciais, para que estes pontos da cadeia não afetem a amostragem *a posteriori*. Por exemplo, a cadeia pode rapidamente oscilar para as regiões em torno de zero em algumas iterações (GEMAN & GEMAN, 1984; CHEN, 2009). Este descarte intencional no início da cadeia é chamado de aquecimento ou “burn-in”. Não existe uma recomendação de quantas iterações devem ser descartadas, porém quanto melhor forem suas informações, menor será o número de descartes (FAN et al., 2002).

É necessário tratar a simulação com cuidado. Normalmente existem duas questões, a saber: a) deve-se primeiramente decidir se a cadeia de Markov atingiu estacionariedade, ou a distribuição *a posteriori* desejada (WALSH, 2004); b) em segundo lugar, precisa ser determinado o número de iterações a serem mantidas após ser atingida a estacionariedade da cadeia de Markov para a convergência. Existem testes diagnósticos que ajudam a resolver este problema (GARETH et al., 2010).

Por outro lado, porém, não existem testes conclusivos que possam dizer quando a cadeia de Markov converge para a distribuição estacionária. Para o teste de convergência, deve-se checar todas as distribuições de todos os parâmetros e não apenas um parâmetro isolado. Por isso é necessário ter cautela em assumir que a cadeia atingiu a estacionariedade e prosseguir com testes de “*burn-in*”, até que isso ocorra (GAMERMAN & LOPES, 2006).

A análise dos traços da simulação é um indicador para verificar se a cadeia convergiu para uma distribuição estacionária. Este rastreamento de amostras em relação ao índice de simulação pode ser muito útil na avaliação da convergência.

O fato de não haver uma convergência estacionária para a distribuição é um indicativo de mau ajuste ou necessidade de um maior período de aquecimento (*burn-in*). Os traços de simulação também podem dizer se a cadeia está convergindo bem e se torna ferramenta de avaliação da qualidade da cadeia, o que indica média e variância constante (BROOKS & ROBERTS, 1998).

Algumas cadeias apresentam características típicas que podem ser observadas nos traços para cada parâmetro. Ilustra-se uma distribuição dos traços simulados, os quais são bastante coesos e próximos à média (Figura 1). Isso indica que a cadeia está atingindo a distribuição corretamente, ou seja, a cadeia está convergindo de maneira adequada (COWLES & CARLIN, 1996).

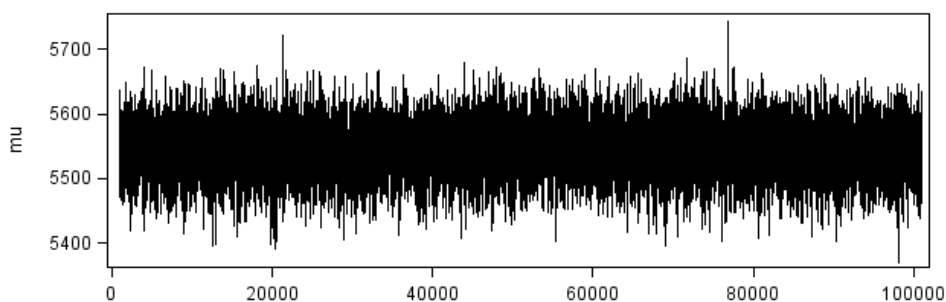


Figura 1. Traços de qualidade desejável para o histórico de simulações MCMC.

Ao contrário do caso anterior, um exemplo de simulação de baixa qualidade pode ser evidenciado quando os traços estão dispersos e com tendências que não estejam de acordo com a tentativa de simulação (Figura 2). Necessita-se efetuar ajustes *a priori* ou fornecer um período maior de aquecimento (BROOKS & ROBERTS, 1998).

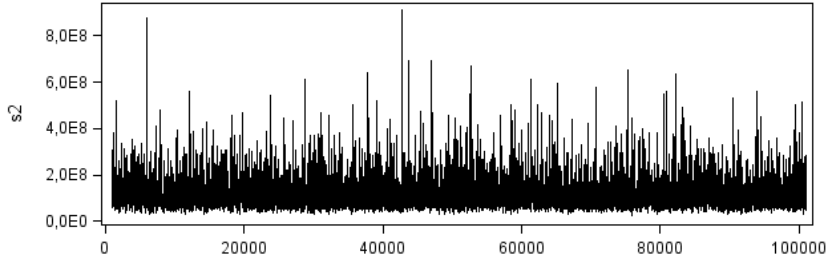


Figura 2. Traços de qualidade indesejável para o histórico de simulações MCMC.

Para minimizar o efeito de autocorrelação durante a simulação e para obter amostras independentes por meio da seleção de valores da distribuição *a posterior* da cadeia de Markov para um número “n” de iteração, uma técnica denominada *thin* é utilizada para melhorar o cálculo recursivo da covariância proposta da cadeia de Markov durante a execução MCMC.

Outro indicativo que pode ser utilizado para testar a convergência da cadeia é o teste de Geweke, que compara os valores no início da cadeia de Markov com os obtidos por meio da simulação da última parte da cadeia, com o intuito de detectar possíveis falhas de convergência (GEWEKE, 1992).

O teste de Geweke é estatístico e tem por base o descrito a seguir: dadas duas subsequências da cadeia de Markov $\{\theta^t\}$, em que são extraídos $\{\theta_1^t: t = 1, \dots, n_1\}$ e $\{\theta_2^t: t = n_a, \dots, n\}$ onde $1 < n_1 < n_a < n$.

Em que $n_2 = n - n_a + 1$, é definido por :

$$\bar{\theta}_1 = \frac{1}{n_1} \sum_{t=1}^{n_1} \theta^t \text{ e } \bar{\theta}_2 = \frac{1}{n_2} \sum_{t=n_a}^{n_1} \theta^t.$$

Temos $\hat{s}_1(0)$ e $\hat{s}_2(0)$ que indicam coerência das estimativas da densidade espectral na frequência zero para as duas cadeias de MCMC, respectivamente. Se as relações de n_1/n e n_2/n são fixas, $(n_1 + n_2)/n < 1$ a cadeia é estacionária, então a estatística a seguir converge para uma distribuição normal para $n \rightarrow \infty$:

$$Z_n = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\frac{\hat{s}_1(0)}{n_1} + \frac{\hat{s}_2(0)}{n_2}}}$$

Este é um teste bilateral e os escores de z indicam rejeição ou não da estacionariedade da cadeia simulada.

1.2.5 Intervalo de credibilidade

Uma das grandes vantagens da abordagem Bayesiana é a construção de regiões de credibilidade, o que torna a inferência não mais em um parâmetro pontual, mas sim um intervalo.

Para estimar o intervalo de credibilidade é utilizado o intervalo de maior densidade - HPD (Highest Posterior Density) (CHEN & SHAO, 1999).

Para o parâmetro θ_i adotam-se as seguintes condições:

A ordem para obter os valores de θ_i^t são ordenados:

$$\theta_{i(1)} \leq \theta_{i(2)} \leq \dots \leq \theta_{i(n)}$$

O intervalo de credibilidade $100(1 - \alpha)\%$ que corresponde a $100(\alpha/2)\%$ e $100(1 - \alpha/2)\%$ é a porcentagem da distribuição a posteriori.

O cálculo do intervalo de credibilidade $100(1 - \alpha)\%$ é dado por:

$$R_{j(n)} = (\theta_{i(j)}, \theta_{i(j+[(1-\alpha)n])})$$

em que: $j = 1, 2, \dots, n - [(1 - \alpha)n]$.

O termo $100(1 - \alpha)\%$ é o intervalo de alta densidade (HPD), denotado por $R_{j(n)}$, ou seja, aquele com menor intervalo entre os intervalos de credibilidade.

O HPD é um intervalo em que encontram-se os pontos com maior probabilidade de densidade, em relação a outros pontos fora do intervalo (BESAG & HIGDON, 1999). A aplicabilidade dessa teoria é interessante. Exemplificando, por meio do MCMC é possível contar o número de vezes que um particular genótipo apresenta o melhor desempenho médio e, portanto, calcular a probabilidade de que seja o de melhor desempenho. Assim, o cálculo desta probabilidade ajuda no melhoramento genético, como forma de selecionar variedades com maior aptidão e desempenho.

A escolha do intervalo de maior densidade (HPD) *a posteriori*, fornece conjuntos com intervalos iguais da distribuição *a posteriori* (CHEN & SAHO, 1999).

Uma grande distinção entre as abordagens Bayesiana e frequentista está na sua interpretação. A probabilidade Bayesiana reflete a opinião subjetiva de uma pessoa. Seguindo esta abordagem, um estatístico pode fazer a alegação de que um parâmetro esteja dentro de um intervalo de credibilidade com probabilidade mensurável. Esta propriedade é interessante, pois permite que se faça uma declaração de probabilidade direta sobre os parâmetros. Um intervalo de credibilidade, por outro lado, permite que se faça uma alegação de que o parâmetro verdadeiro esteja dentro do intervalo.

1.3 Modelos lineares mistos no melhoramento genético

Um modelo linear que apresenta somente fatores de efeitos fixos, além do erro experimental, sempre aleatório, é denominado modelo fixo. Modelos que apresentam apenas fatores de efeitos aleatórios, exceto a constante μ , sempre fixa, são denominados modelos aleatórios. Um modelo misto é aquele que apresenta tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ (CAMARINHA FILHO, 2002).

A avaliação genética usando modelos lineares mistos pode ser realizada com dados balanceados e desbalanceados, que podem ser obtidos por meio de delineamentos completos ou incompletos (RESENDE, 2002).

Outro motivo de se adotar um modelo linear misto é a possibilidade de se fazer a predição de efeitos aleatórios na presença de efeitos fixos, por meio dos BLUP's (*Best Linear Unbiased Prediction*), que são de grande valia em genética e melhoramento (CAMARINHA FILHO, 2002).

No contexto de modelos lineares mistos, pode-se fazer uma abordagem diferenciada em cada efeito do modelo matemático – estatístico, pois é possível declarar dentro de um modelo quais serão de efeito aleatório e fixo, estratificando os parâmetros nas matrizes de incidência diferenciadas para efeito fixo e outra para efeito aleatório, obtendo-se diferentes soluções para cada parâmetro do modelo (SEARLE et al., 1992).

No melhoramento vegetal a utilização dos modelos lineares mistos é recomendada quando o efeito de genótipo é tido como aleatório, e efeito como o de bloco pode ser considerado como fixo. Dessa forma há uma correção matricial dos efeitos aleatórios pelos efeitos tidos como fixo, visto que se eliminam os ruídos de um efeito dentro do outro. Assim, são eliminados os ruídos de blocos associados aos efeitos de genótipos, permanecendo apenas os efeitos de interesse para inferências (RESENDE, 2002).

Em análises e estudos de dados provenientes de diferentes experimentos com diferentes níveis de tratamentos, é recomendada a utilização de modelos lineares mistos, pois estes permitem a predição dos efeitos aleatórios e fornecem estimativas precisas dos componentes de variância (RESENDE, 2002).

O sucesso na obtenção das estimativas é possível, pois a solução do modelo é dada de forma independente no nível de incidência de parâmetros em matrizes diferentes (matrizes de incidência de efeitos fixos e aleatórios) (RESENDE, 2002).

A metodologia de modelos lineares mistos tem a vantagem de ser bastante flexível para a estimativa de parâmetros e predição de valores genéticos, pois pode ser aplicado a dados desbalanceados e provenientes de diferentes gerações (KALIL FILHO et al., 2000).

Casos de desbalanceamento, frequentes em experimentos de campo, podem ser promovidos por várias condições que levam a perdas de parcelas ou impossibilidade de avaliar todos os genótipos ou locais (RESENDE & BARBOSA, 2006).

Por ser uma análise mais elaborada e minuciosa, o uso de modelos lineares mistos requer um conhecimento adequado da metodologia de estimadores de variância e preditores de valores associados aos tratamentos. Por sua vez, a associação dos mesmos pode levar a resultados superiores aos obtidos em metodologias clássicas de análise, porém na pior das hipóteses podem-se obter resultados similares com outras metodologias de abordagem linear normal (FARIAS NETO & RESENDE, 2001; RESENDE, 2002).

Modelos lineares mistos utilizados com estimadores de variância e preditores de valores de efeito aleatório podem ser usados para evitar distorções, por sua maior flexibilidade e pelos efeitos de repetição dos blocos, anos e locais poderem ser convenientemente considerados como de efeitos fixos ou aleatórios, independente das variações do efeito de genótipo ser considerados como aleatórios (RESENDE, 2002; 2004).

A adoção de efeito de bloco como fixo é bastante discutida em casos de ensaios de campo de programas de melhoramento, quando o efeito de blocos é de caráter aleatório, pois tem-se a premissa de que os blocos em um determinado local representa uma amostra de todos os blocos possíveis naquele local (RAMALHO et al., 2000).

Resende & Higa (1994) estimaram os efeitos de blocos na seleção de espécies perenes e constataram que o efeito de blocos não contribui para o efeito genético e que indivíduos podem ser prontamente comparados. Portanto, o efeito de bloco tem de ser tomado como fixo. Para Piepho (1994) delineamentos de blocos casualizados, devem preferencialmente ser considerados como efeito fixo para ensaios de competição.

É grande o número de ensaios de cultivares que são tidas cujo efeito é considerado como fixo, porém tratar tal efeito como aleatório traz algumas vantagens como a obtenção de estimativas de componentes de variância, que são ferramentas de grande valia para o melhorista (RESENDE, 2002 b). Esses estimadores tornam possível maximizar os efeitos genéticos conduzindo a uma seleção mais apropriada (PIEPHO & MORHRING, 2005).

1.3.1 Estimadores de componentes de variâncias

São vários os métodos existentes para obter as estimativas de componentes de variâncias, cada qual com suas particularidades, vantagens e desvantagens, sendo basicamente nove métodos derivados de três conceitos clássicos da estimação estatística: os momentos, a função de verossimilhança e as funções quadráticas (RESENDE, 2007).

Os primeiros métodos com intuito de estimar variâncias baseiam-se nos métodos dos momentos. Nesta classe podem ser enquadrados o método de análise de variância, Anova de Fisher – 1918, e os métodos I, II e III, de Henderson (1953). Segundo Valério Filho (1983), o método I é uma extensão do método Anova e o método II é uma modificação do método I, visando à aplicação em modelos mistos que não incluam interação entre efeitos fixos e aleatórios. O método III pode ser aplicado tanto a modelos mistos quanto a modelos aleatórios.

Sobre o enfoque de funções quadráticas, temos os métodos dos estimadores quadráticos não-viesados de norma mínima (MINQUE) e de variância mínima (MIVQUE), ambos propostos por Rao (1971a) e também o método iterativo de norma mínima (I-MINQUE) (SEARLE, 1987).

Os métodos estimadores quadráticos não-viesados MINQUE e MIVQUE exigem o conhecimento e a atribuição de valores *a priori* para os componentes de variância, sendo as alternativas mais comuns MIVQUE(A)/MINQUE(A) e MIVQUE(0). O primeiro utiliza como *a priori* as estimativas da Anova, e o segundo utiliza como valor inicial “um” para componente de erro e zero aos demais componentes. Para o método iterativo I-MINQUE, utiliza-se um vetor com pesos pré-estabelecidos com variâncias (SEARLE et al., 1992).

Os métodos que têm por base a função de verossimilhança dos dados são a Máxima Verossimilhança (ML), proposta por Hartley & Rao (1967), e a Máxima Verossimilhança Restrita (REML), proposta por Patterson & Thompson (1971).

O método de ML apresenta vícios na estimativa de variâncias, os quais são atribuídos ao fato de o método não levar em conta a perda de graus de liberdade, associados às estimativas dos efeitos fixos, em decorrência da imposição de restrição à não-negatividade nos algoritmos de cálculo, tornando-se um viés (SEARLE et al., 1992). Para eliminar esse vício de não-negatividade, é necessário aumentar o número de observações do fator aleatório, cujo componente é alvo de estimação, o que leva à eliminação do viés (SEARLE, 1987).

O método da ML maximizam as observações (efeitos aleatórios) em função dos efeitos fixos e dos componentes de variância, porém é necessário assumir que as observações são normalmente distribuídas (SEARLE, 1987).

Alterações foram sugeridas por Patterson & Thompson (1971) na metodologia de ML, que passou a ser conhecida como Máxima Verossimilhança Restrita (REML), hoje tida como a melhor forma de se obter as estimativas de variâncias (RESENDE, 2002). Diferente do que ocorre no método ML, o método REML leva em consideração os graus de liberdade dos efeitos fixos, eliminando, assim, o vício do método ML.

O método REML baseia-se no fato da verossimilhança ser fracionada em dois componentes. O primeiro é a verossimilhança de um ou mais parâmetros estatísticos e envolve todos os parâmetros fixos, como μ . O segundo é composto pela verossimilhança residual e envolve apenas os parâmetros de variância dos efeitos aleatórios. Temos então cada componente maximizado separadamente.

As estimativas dos parâmetros de variância são conhecidas como as estimativas REML.

Para amostras de uma população normal, o primeiro componente torna-se a verossimilhança da média da amostra y e o segundo componente a verossimilhança da variável associada com a variância amostral, especificamente:

$$\text{Log } L = \left[-\frac{1}{2} \ln \left(\frac{2\pi\sigma^2}{n} \right) - \frac{1}{2} \left(\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \right)^2 \right] + \left[-\frac{n-1}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \ln(n) - \frac{1}{2} \sum \left(\frac{Y_i - \bar{y}}{\sigma} \right)^2 \right]$$

A utilização do algoritmo REML em experimentos que envolvem efeitos fixos e efeitos aleatórios não é restritiva a dados independentes, ou dados com a mesma variância em qualquer nível. Logo, torna-se uma ferramenta extremamente flexível para obter estimativas, para análise de dados provenientes de ensaios agrícolas.

1.3.2 Importância da escolha do estimador de variância na predição de valores genéticos

A escolha do estimador de variância para a predição do valor genotípico é de fundamental importância, tendo em vista que os preditores lineares das variáveis aleatórias são funções diretas dos componentes de variância presentes no modelo de Searle et al. (1992).

Os delineamentos tidos como aumentados ou em blocos incompletos são mais frequentes em ensaios de pesquisas de melhoramento de plantas e, em muitos casos, são

analisados como de efeitos fixos para todos os componentes do modelo. Contudo, já se sabe que estes delineamentos têm melhor análise com o enfoque de modelos lineares mistos (SEARLE et al., 1992). Vale ressaltar que a escolha do estimador dos componentes de variância vai refletir diretamente na qualidade dos valores preditos dos genótipos (DUARTE et al., 2001).

1.3.3 Predição de valores genéticos

Do ponto de vista da inferência clássica de estimação/predição, quando o efeito genético é considerado como de efeito fixo, os valores genéticos (*breeding values*) são estimados. Quando os efeitos de genótipo são considerados como de efeitos aleatórios os valores são preditos (RESENDE, 2007).

Quando o efeito genético é tido como fixo com variâncias homogêneas, utiliza-se o método dos mínimos quadrados para estimar os valores genéticos. Em casos de efeito genético fixo com heterogeneidade de variâncias, para obtenção de estimativas, utiliza-se o método dos quadrados mínimos ponderados ou generalizados (COTTERILL et al., 1983).

Para genótipos considerados de efeito aleatório, que é desejável, as predições podem ser feitas de três maneiras: BP, “*Best Prediction*”; BLP “*Best linear Prediction*” e BLUP “*Best Linear Unbiased Prediction*”, sendo este último maior em eficiência (RESENDE, 2002).

Segue a descrição dos três preditores, classificados como “Best”:

“BP” - Melhor Preditor (*Best Prediction*) - quantidades balanceadas e precisões de informações associadas a todos os indivíduos avaliados, em que as médias e variâncias são previamente conhecidas ou estimadas com precisão.

“BLP” - Melhor Predição Linear (*Best Linear Prediction*)- pode ser usado em observações desbalanceadas e fornece precisão nas informações vinculadas aos indivíduos avaliados, sendo as médias e componentes de variâncias conhecidas ou estimadas com precisão.

“BLUP” - Melhor Preditor Linear não-tendencioso (*Best Linear Unbiased Prediction*) - com este preditor é possível trabalhar dados desbalanceados e mesmo assim obter informações precisas dos indivíduos avaliados, sendo os componentes de variâncias conhecidos ou estimados com precisão, sem o conhecimento da média (BERNARDO, 2002; MAYOR & BERNARDO, 2009).

Assim, BLUP pode ser, resumidamente, definido como resultado da regressão dos efeitos de um fator aleatório (g), em função das observações (y) corrigidas para os efeitos fixos do modelo (MARTINS et al., 1993).

O BLUP é o preditor de máxima correlação com o verdadeiro valor genotípico (g), sob a normalidade do caráter e dos valores genotípicos; μ é o critério ótimo de seleção, pois apresenta o ordenamento de candidatos que maximizam a média genotípica populacional (DUARTE et al., 2001).

O preditor BLUP maximiza a correlação entre o valor genético predito e o verdadeiro, ou seja, elevada acurácia. Este preditor minimiza o quadrado da diferença entre valores genéticos preditos e verdadeiros ou, ainda, minimiza a variância do erro de predição, tendo o requisito para um preditor ser “best” $E(\hat{a} - a)^2$. Por esta razão, pode otimizar a probabilidade de selecionar o melhor entre dois indivíduos avaliados e otimizar a probabilidade de selecionar o melhor indivíduo entre vários indivíduos avaliados (ROBINSON, 1991).

Segundo Henderson (1984), dentre uma classe de funções lineares que geram predições não-viciadas, o BLUP maximiza a correlação ($\gamma - \hat{\gamma}$). Logo, é possível verificar que quando se comparam os preditores, o mais completo é o BLUP conjuntamente com o melhor estimador de variâncias REML, pois tais procedimentos conduzem à maximização do valor genético predito, levando a uma precisão adequada na tomada de decisão.

1.3.4 Aplicativo computacional para REML-BLUP

A utilização da metodologia de modelos lineares mistos via REML-BLUP é realizada por uma série de aplicativos computacionais, como, por exemplo, o SAS (1992), com os procedimentos: PROC MIXED, PROC IML, PROC GLIMMIX, dentre outros (CAMARINHA FILHO, 2006). Outro aplicativo que pode ser utilizado é o ASREML, programa bastante versátil que constitui boa ferramenta para o melhoramento animal e vegetal (GILMOUR et al., 1995). No âmbito nacional, temos o SELEGEN-REML/BLUP (RESENDE, 2002 b), desenvolvido para o melhoramento de espécies perenes, em que é possível realizar análise com mais de 150 variações de modelos lineares mistos (RESENDE, 2007 b).

1.3.5 Adaptabilidade e estabilidade

É de fundamental importância no programa de melhoramento vegetal a seleção de cultivares/genótipos com aptidões específicas, isto é, com comportamento diferenciado ou não, em ambientes distintos. Para entender este comportamento utilizam-se diferentes técnicas que estudam a adaptabilidade e estabilidade de um grupo de cultivares. Por meio destes estudos, é possível recomendar genótipos em ambientes específicos ou amplos.

A adaptabilidade do genótipo é a capacidade de o indivíduo sobreviver, crescer e se reproduzir nas condições ambientais do local onde está inserido. A estabilidade é a capacidade de um genótipo exibir um desempenho constante em função de variações ambientais (CRUZ et al., 2004).

São vários os procedimentos utilizados no estudo da adaptabilidade e estabilidade e comportamento diferencial de genótipos em ambientes diferentes. Os primeiros métodos descritos na literatura têm por base a utilização da análise de variância, por meio do uso de componentes quadráticos do desdobramento da interação. Outros métodos utilizam regressões lineares, simples, quadráticas, múltiplas e modelos não-lineares (RESENDE, 2007). Podem também ser utilizados métodos multivariados, como a técnica de componentes principais e análise fatorial.

Uma nova metodologia para o estudo de adaptabilidade e estabilidade foi descrito por Resende (2004), denominada Média Harmônica da Performance Relativa do Valor Genético (MHPRVG), que torna possível praticar seleção simultânea para a média genética, adaptabilidade e estabilidade, sobre o enfoque de modelos mistos, utilizando o estimador REML e o preditor BLUP.

1.3.6 Média Harmônica do Valor Genético (MHVG)

A predição baseada na média harmônica permite selecionar famílias simultaneamente, por produtividade e estabilidade. Pode ser feita pelo emprego dos mesmos preditores BLUP e das equações dos modelos lineares mistos utilizados na tradicional seleção baseada na média aritmética (RESENDE, 2004).

Para tanto, é composto o vetor de dados, para as análises conjuntas de locais, como a recíproca de cada dado observado (y), ou seja, $1/y$, conforme sugere Resende (2004). A média de $(1/y)$ fornece uma função $(1/H)$ da média harmônica (H) desses dados. Assim, os

resultados dos valores genéticos gerados pelo SELEGEM-REML/BLUP são invertidos, visando à obtenção de:

$$MHVG = \frac{l}{\sum_{i=1}^l \frac{1}{VG_j}}$$

1.3.7 Performance relativa do valor genético (PRVG)

Genericamente, a performance relativa do valor genotípico tem sido utilizada há muito tempo (WRIGHT et al., 1966) em termos de dados fenotípicos e constitui a base do método de Annicchiarico (1992). No cálculo da PRVG, os valores fenotípicos ou dados originais são expressos como proporção da média geral de cada local e, posteriormente, obtém-se BLUP contendo o valor médio dessa proporção entre os locais (RESENDE, 2004):

$$PRVG = \frac{1}{l} \left(\frac{\sum VG_j}{M_j} \right) \pi r^2$$

1.3.8 Média Harmônica da Performance Relativa do valor Genético predito (MHPRVG)

A metodologia MHPRVG apresenta algumas particularidades vantajosas ao melhorista, como o fato de praticar a seleção simultânea para três caracteres, considerar o efeito de genótipo como aleatório, praticidade no estudo de dados com heterogeneidade de variância e dados desbalanceados que podem ser amplamente utilizados até mesmo em delineamentos mais elaborados. A MHPRVG considera erros correlacionados dentro de locais e faz inferência em valores genéticos já decompostos das instabilidades, da mesma grandeza numérica das variáveis estudadas:

$$MHPRVG = \frac{l}{\sum_{i=1}^l \frac{1}{PRVG_j}}$$

A aplicação da técnica de seleção simultânea por rendimento de grãos, média genética ($\mu + g_i$), adaptabilidade e estabilidade pode ser aplicada à cultura do milho, tendo por base os modelos lineares mistos utilizando o método da Média Harmônica da Performance Relativa dos Valores Genéticos (MHPRVG) preditos.

1.3.9 Análise de Deviance (Anodev)

Na análise de variância clássica, as fontes de variação dos modelos são testadas por meio do teste F. Como alternativa para teste de modelos lineares mistos, é utilizada a Análise de Deviance (Anodev).

A utilização da Análise de Deviance é proposta como medida de discrepância para os modelos lineares mistos, por meio da elaboração de uma tabela similar a de análise de variância, com intuito de avaliar a contribuição de cada componente ao modelo proposto.

No entanto, a interpretação do quadro de Anodev é um pouco mais complexa pela falta de ortogonalidade dos componentes. Cada medida de deviance representa a variação correspondente a cada componente, excluindo o efeito do componente incluído antes dele, e ignorando os efeitos dos termos que são incluídos posteriormente (STURION & RESENDE, 2010).

A diferença entre deviances segue assintoticamente uma distribuição qui-quadrado, com $n-p$ graus de liberdade, sendo que um modelo com p parâmetros é considerado correto (MCCULLAGH & NELDER, 1989).

A vantagem de utilizar a deviance como medida de discrepância quando se utiliza a máxima verossimilhança é que a deviance é aditiva e, ao acrescentar variáveis no modelo, a deviance deve decrescer.

Neste contexto, a deviance desempenha o papel do resíduo da análise de variância, sendo a deviance proporcional a $-2x\text{Log}L$ em que o $\text{Log}L$ é o logaritmo da verossimilhança do componente de variância. Uma mudança na deviance de um modelo em comparação com um modelo maior segue uma distribuição de χ^2 (RESENDE, 2007).

REFERÊNCIAS

- ANNICCHIARICO, P. Cultivar adaptation and recommendation from alfalfa trials in Northern Italy. **Journal Genetics Breeding**, New Delhi, v. 46, n. 1, p. 269-278, 1992.
- BALDING, D. J. Discussion on the meeting on 'Statistical modelling and analysis of genetic. **Royal Statistical Society**. 64, Parte. 4. , p. 737-775, 2002.
- BAYARRI, M. J.; BERGER, J O. The Interplay of Bayesian and Frequentist Analysis. **Statistical Science**, Beachwood, v. 19, n.1, p. 58-80, 2004.
- BAYES, T. An essay towards solving a problem in the doctrine of Chances. **Philosophical Transactions of the Royal Society**. v. 53, p. 370-418, 1763.
- BERNARDO, R. **Breeding for quantitative traits in plants**. Woodbury, MN. : Stemma Press, 2002. 369p.
- BESAG, J. E.; HIGDON, D. Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society - Statistical Methodology*. 61, Parte. 4, p. 691-746, 1999.
- BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: Wiley Classic Library, 1973. 608p.
- BROOKS, S. P.; ROBERTS, G. O. Diagnosing convergence of Markov chain Monte Carlo algorithms. **Statistics and Computing**, London, v. 8, p. 319-335, 1998.
- CAMARINHA FILHO, J. A. **Modelos lineares misto: estruturas de matrizes de variância e covariância e seleção de modelos**. Piracicaba: Universidade de São Paulo. 2002. 96p. Tese de doutorado.
- CAMARINHA FILHO, J. A. **Uma aplicação de modelos lineares mistos**. Curitiba: Universidade Federal do Paraná, 2006. 28p. Tese de doutorado.
- CASELA, G.; GEORGE, E. I. Explaining the gibbs sampler. **The American Statistician**, Washington, v. 46, n. 3, p. 167-174, 1992.
- CHEN, F. Bayesian Modeling Using the MCMC Procedure. **SAS Institute Inc.** - Cary, NC , 2009. 257p.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of Bayesian credible and HPD intervals. **Journal of Computational and Graphical Statistics**, Alexandria, v, 8, p. 69-92, 1999.
- CONGDON, P. **Applied Bayesian Modeling**. New York: John Wiley & Sons, 2003. 478p.
- CONAB. **Companhia Nacional de Abastecimento. Acompanhamento de safra brasileira: grãos, sétimo levantamento**, abril/2012 - Companhia Nacional de Abastecimento. Brasília: Conab 2012. Acesso em: 20 de abril de

2012:http://www.conab.gov.br/OlalaCMS/uploads/arquivos/12_04_11_15_04_18_boletim_abril_2012.pdf.

CORDEIRO, G. M. **Introdução à teoria assintótica**. Rio de Janeiro: 22º Colóquio Brasileiro de Matemática, IMPA, 1999. 166p.

COTTERILL, P. P.; CORRELL, R. L.; BOARDMAN, R. Methods of estimating the average performance of families across-incomplete open-pollinated progeny tests. **Silvae Genetica**, Frankfurt, vol. 32, n. 1-2, p.28-33, 1983.

COWLES, M. K. E.; CARLIN, B. P. Markov chain Monte Carlo convergence diagnostics: A comparative review. **Journal of the American Statistical Association**, Washington, v. 91, p. 883-904, 1996.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Universidade Federal de Viçosa. 2004. 480p.

DUARTE, J. B.; VENCOVSKY, R.; DIAS, T. S. Estimadores de componentes de variância em delineamento de blocos aumentados com tratamentos novos de uma ou mais populações. **Pesquisa Agropecuária Brasileira**, Brasília, v. 36, n. 9, p. 1155-1167, 2001.

EAVES, L.; ERKANLI, A. Markov chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and G x E interaction. **Behavior Genetics**, Boulder, v. 33, n. 3, p. 279-299, 2003.

FAN, X.; ÁKOS, F. I.; SIVO, S. A.; KEENAN S. C. **SAS for Monte Carlo Studies "A Guide for Quantitative Researchers"**. Cary, NC : SAS Institute Inc., 2002. 272p.

FARIAS NETO, J. T.; RESENDE, M. D.V. Aplicação da metodologia de modelos mistos (REML/BLUP) na estimação de componentes de variância e predição de valores genéticos em pupunheira (*Bactris gasipaes*). **Revista Brasileira de Fruticultura**, Jaboticabal, v. 23, n. 2, p. 320-324, 2001.

FERREIRA, E.A.; PATERNIANI, M.E.A.G.Z.; DUARTE, A.P.; GALLO, P.B.; SAWAZAKI, E.; AZEVEDO FILHO, J.A.; GUIMARÃES, P.S. Desempenho de híbridos top crosses de linhagens S3 de milho em três locais do Estado de São Paulo. **Bragantia**, Campinas, v.68, n. 2, p. 319-327, 2009.

GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for bayesian inference**. London: Chapman & Hall, 2. ed. edição, 2006. 264p.

GAMMERMAN, D. **Markov chain Monte Carlo: stochastic simulation for bayesian inference**. London: Chapman and Hall , 1997. 344p.

GARETH, W. P.; KANNANY, B.; LASSCOCKY, B.; MELLEN, C. Model selection and Adaptive Markov Chain Monte Carlo for Bayesian Cointegrated VAR Models. **International Society for Bayesian Analysis**, Durham, vol, 3, p. 465-492, 2010.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Washington, v. 85, p. 398-409, 1990.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, vol. 6, p. 721-741, 1984.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. **Bayesian Statistics**. Oxford, UK: Clarendon Press, v. 4, 1992.

GILKS, W.R.; RICHARDSON, S.; SPIEGELHALTER, D.J. Markov Chain Monte Carlo in Practice. London: Chapman & Hall/CRC Interdisciplinary Statistics, 1996 . 512p.

GILMOUR, A.R.; THOMPSON, R.; CULLIS, B. ASREML, an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, p. 1440-1450, 1995.

GOLDSTEIN, M.; WOOF, D. **A bayes linear statistics: theory and methods**. New York: John Wiley & Sons, 2007. 536p.

HARTLEY, H. O.; RAO, J. N. K. Maximum likelihood estimation for the mixed analysis of variance model. **Biometrika**. Oxford, v. 54, p. 93-108, 1967.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**. Oxford, v. 57, p. 97-109, 1970.

HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**. Arlington, v. 9, p. 226-252, 1953.

HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.

HOTI, F.; SILLANPÄÄ, M. J. Bayesian mapping of genotype x expression interactions in quantitative and qualitative traits. **Heredity**, v, 97, p. 4-18, 2006.

HUANG, T.; NEVMYVAKA, Y. A practical markov chain monte carlo approach to decision problems. **Research Society Conference**. Middlebury College, p. 520-524, 2001.

KALIL FILHO, A. N.; RESENDE, M. D. V.; KALIL, G. P. C. Componentes de variância e predição de valores genéticos em seringueira pela metodologia de modelos mistos (REML/BLUP). **Pesquisa Agropecuária Brasileira**. Brasília, v, 35, n. 9, p. 1883-1887, 2000.

KAUFMAN, C. G. E.; SAIN, S. R. Bayesian functional ANOVA modeling using gaussian process prior distributions. **International Society for Bayesian Analysis**, Durham, v. 5, n. 1, p. 123-150, 2010.

MAKIYAMA, F.A. **Cenários bayesianos para análise de riscos**. Projeto de Pesquisa para elaboração do Trabalho de conclusão de Curso apresentado como exigência para a obtenção

do título de bacharel em Ciências da Computação à Universidade Federal de Santa Catarina – UFSC, no curso de Ciências da Computação. 60 p. 2006.

MARTINS, E.N.; LOPES, P. S.; SILVA, M. A.; REGAZZI, A. J. **Modelo linear misto**. Viçosa: Universidade Federal de Viçosa, 1993. 46p.

MAYOR, P. J.; BERNARDO, R. Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F2 populations. **Crop Science**, Madison, v. 49, n. 5, pp. 1719-1725, 2009.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models (mathematical statistics of generalized linear model)**. London : Chapman and Hall, 1989.

METROPOLIS, N. The beginning of the Monte Carlo method. **Los Alamos Science**, Los Alamos, v, 15, p. 125–130, 1987.

METROPOLIS, N.; ROSENBLUTH, A. W.; MARSHALL, N. R.; TELLER, A. H. Equations of state calculations by fast computing machines. **The Journal of Chemical Physics**, Melville, v, 21, p. 1087–1092, 1953.

MOON, H. R.; SCHORFHEIDE, F. **Bayesian and frequentist inference in partially identified models. department of economics**. 2011. Acesso em: 10 de Janeiro de 2012. <http://www.ssc.upenn.edu/~schorf/papers/partialid-paper.pdf>.

O'HARA, R. B.; SILLANPÄÄ, M. J. A review of bayesian variable selection methods: what, how and which. **International Society for Bayesian Analysis**. Durham, v, 4, n. 1, p. 85-118, 2009

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Oxford, v. 58, p. 545-554, 1971.

PIEPHO, H. P. Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects and multiplicative interaction (AMMI) analysis. **Theoretical and Applied Genetics**, Berlin, v. 89, n. 5, p. 647-654, 1994.

PIEPHO, H. P.; MOHRING, J. Best Linear Unbiased Prediction of Cultivar Effects for Subdivided Target Regions. **Crop Science**, Madison, v. 45, n.3, p.1151-1159, 2005.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: Universidade Federal de Lavras, 2000. 326p.

RAO, C. R. Estimation of variance and covariance components — MINQUE theory. **Journal of Multivariate Analysis**, Los Angeles, v. 1, p. 257-275, 1971.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975p.

RESENDE, M. D. V. **Software Selegen–REML/BLUP**. Colombo: Embrapa Florestas, Documentos 77, 2002a. 67p.

RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004. 65p.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RESENDE, M. D. V. **SELEGEN–REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Colombo: Embrapa Florestas, 2007a. 361p.

RESENDE, M. D. V. **SELEGEN–REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Colombo: Embrapa Florestas, 2007b. 361p.

RESENDE, M.D.V.; BARBOSA, M.H.P. Selection via simulated BLUP based on family genotypic effects in sugarcane. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, p.421-429, 2006.

RESENDE, M. D. V.; HIGA, A. R. Estimaco de parâmetros genéticos no melhoramento de Eucalyptus: seleço em um caráter com base em informaçes do indivíduo e de seus parentes. **Boletim de Pesquisa Florestal**, Colombo, n. 28-29, p. 11-36, 1994.

ROBERT, C. P.; CASELLA G. **Monte Carlo Statistical Methods**. 2ª Edition. New York: Springer -Verlag, 2004. 645p.

ROBERT, C. P.; CASELLA, G. A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. **Statistical Science**, Beachwood, v. 26, n. 1, p.102-115. 2011.

ROBINSON, G. K. That BLUP is a good thing: the estimation of random effects. **Statistical Science**, Beachwood, v. 6, n.1, p. 15-51, 1991.

SAS, INSTITUTE. **Software: statistical analysis system institute**. Cary, NC, 1992. 663p.

SEARLE, S. R. **Linear models for unbalanced data**. New York : John Wiley & Sons. 1987. 536p.

SEARLE, S. R.; CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York: John Wiley& Sons. 1992. 501p.

SORENSEN, D.; GIANOLA, D. **Likelihood, bayesian, and MCMC methods in quantitative genetics**. New York: Springer, 2002. 757p.

STURION, J. A.; RESENDE, M. D. V. Avaliaco genética e análise de deviance em um teste desbalanceado de procedência e progênie de *Ilex paraguariensis*. **Brazilian Journal of Forestry Research**, Colombo, v. 30, 60, p. 157-160, 2010.

TIERNEY L. Markov chains for exploring posterior distributions. **Annals of Statistics**, v. 22, p. 1701–1786, 1994.

ULAM, S.; RICHTMYER, D.; NEUMANN, J. VON. Statistical methods in neutron diffusion. **LAMS-551, Los Alamos National Laboratory**, p. 18-36, 1947.

VALLVERDÚ J. The false dilemma: bayesian vs. frequentist. **Electronic Journal For Philosophy**. ISSN:1211-0442, 2008.

VALÉRIO FILHO, W.V. **Métodos de Henderson para componentes da variância de dados não balanceados**. Piracicaba: Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, 1983. 137p. Dissertação de Mestrado.

VENCOVSKY, R.; RAMALHO, M.A.P. Contribuição do melhoramento genético de plantas no Brasil. In: PATERNIANI, E. (ed.). **Agricultura brasileira e pesquisa agropecuária**. Brasília: EMBRAPA, 2000. 194p.

WALSH, J. B. Markov Chain Monte Carlo and Gibbs Sampling. **Lecture Notes for EEB 581**. version 26, April 2004.

WEERAHANDI, S.; ZIDEK, J. V. Multi-Bayesian statistical decision theory. **Journal of the Royal Statistical Society** - Series A. v,144, p. 85–93, 1981.

WRIGHT, J. W. S. S.; PAULEY, R. B. POLK, J. J.; READ, R. A. Performance of Scotch pine varieties in the north-central region. **Silvae Genetica**, Frankfurt, v. 15, p. 101-110, 1966.

CAPÍTULO 2

ANÁLISE DIALÉLICA DE LINHAGENS DE MILHO COM ENFOQUE BAYESIANO

RESUMO

O presente trabalho teve por objetivo avaliar as capacidades geral e específica de combinação do rendimento de grãos de dez linhagens de milho por meio de cruzamentos dialélicos. Para tal, utilizou-se o procedimento Bayesiano, implementado pelo algoritmo de simulação estocástica de Metropolis-Hastings. O método de Markov Chain Monte Carlo (MCMC) concilia probabilidade com simulação e possibilita inserir informações *a priori* para obtenção das distribuições marginais *a posteriori* dos parâmetros de interesse, assim como a construção de regiões de credibilidade. Os resultados do intercruzamento em esquema dialélico das dez linhagens de milho foram analisados primeiramente pelo método 4 (modelo I) proposto por Griffing (1956), o qual leva em consideração apenas os F1's. Os resultados indicaram que as linhagens 6, 5, 9 e 10 mostraram-se promissoras para serem utilizadas no programa de melhoramento. As combinações híbridas 2x6, 1x8, 1x5 foram as que mais destacaram-se para rendimento de grãos, com as maiores estimativas de capacidade específica de combinação.

. **Palavras-chave:** *Zea mays* L., capacidade de combinação, MCMC.

DIALLEL ANALYSIS OF MAIZE LINES WITH BAYESIAN APPROACH

ABSTRACT

The present study aimed to estimate the General Combining Ability (GCA) and the Specific Combining Ability (SCA) of the grain yield of ten maize inbred lines through diallel crosses. For such a purpose, the Bayesian procedure was used, implemented by the Metropolis-Hastings stochastic simulation algorithm. Markov Chain Monte Carlo methods (MCMC) arrange probability with simulation and enables the insertion of prior information toward obtaining posterior marginal distributions of the parameters of interest, as well as the construction of Highest Posterior Density (HPD) intervals. The results of the diallel intercross of all ten maize inbred lines and their hybrids were analyzed by method 4 (model I) described by Griffing (1956). Maize inbred lines 6, 5, 9 and 10 are promising to be used in breeding programs. Hybrid combinations 2x6, 1x8 and 1x5 showed the highest estimates of specific combining abilities for grain yield.

Keywords: *Zea mays* L., combining ability, MCMC

1 INTRODUÇÃO

O objetivo maior de um programa de melhoramento de cultivares é selecionar linhagens aos pares com elevada complementação e especificidade nos cruzamentos, o que contribui para o incremento de rendimento de grãos (BERNARDO, 1995).

A principal metodologia para a seleção de linhagens superiores baseia-se em cruzamentos dialélicos que fundamentam a escolha por meio de estimativas da capacidade geral de combinação de cada linhagem (VENCOVSKY & BARRIGA, 1992). A capacidade geral de combinação é o parâmetro que determina o valor genético de cada linhagem.

Os esquemas de cruzamentos dialélicos baseiam-se em modelos matemáticos que fornecem estimativas que auxiliam na escolha dos melhores genitores e na definição de grupos heteróticos.

Em muitos casos durante um programa de melhoramento é desejável conhecer a expressão média de um caráter, o qual pode ser um indicativo do valor genético de um determinado genótipo individual ou de uma população (VENCOVSKY & BARRIGA, 1992).

Em características quantitativas, os valores avaliados provem das médias fenotípicas observadas. Quanto maior o número de n indivíduos observados, mais a média geral aproxima-se da média populacional, sendo, portanto, fidedigno ao valor genotípico do seu indivíduo ou população (VENCOVSKY & BARRIGA, 1992). A escolha do efeito fixo ou aleatório depende da inferência que se deseja fazer para o conjunto de dados, na qual o valor genotípico é mais constantemente inferido na natureza aleatória (HENDERSON, 1984).

A metodologia de Griffing (1956) é a mais usual para estimar os valores genéticos dos parentais e seus cruzamentos, ou seja, a capacidade geral e específica de combinação. Também são encontradas na literatura outras metodologias, como Gardner & Eberhart (1966), a qual tem por finalidade avaliar cruzamentos com variedades em equilíbrio de Hardy-Weinberg. Por meio desta metodologia, é possível estimar os efeitos das variedades e a heterose varietal. Outro método que pode ser utilizado se trata do dialelo proposto por Hayman (1954), que fornece informações sobre o mecanismo de herança da característica avaliada (CRUZ & REGAZZI, 1994).

O termo dialelo é utilizado para expressar o conjunto de $p(p-1)/2$ híbridos, no qual p é o número de linhagens, variedades, etc, podendo este, incluir os genitores e os F_1 s recíprocos. A solução deste modelo geralmente se dá por meio de modelos lineares fixos (VENCOVSKY & BARRIGA, 1992).

As estimativas de capacidade geral e específica de combinação são obtidas de maneira que os resultados são pontuais. Uma opção a esta característica é a construção de intervalos de credibilidade, que podem ser estimados por meio de técnicas com abordagem Bayesiana.

Para a construção de espaço amostral para obter as estimativas de intervalos de credibilidade, que contenham informações *a posteriori*, são utilizadas metodologias de simulação que levam em conta as informações *a priori*, como o método de Monte Carlo via Cadeia de Markov (MCMC).

O método MCMC é uma alternativa aos métodos não-iterativos para problemas complexos, que envolvem informações *a priori* e simulações com incremento de probabilidade associada às distribuições *a posteriori* (GAMERMAN, 1997). A idéia é obter uma amostra da distribuição *a posteriori* e calcular estimativas amostrais de características desta distribuição (CHEN & SHAO, 1999).

Os métodos MCMC mais utilizados são o algoritmo de Metropolis-Hastings, (METROPOLIS et al., 1953) e o amostrador de Gibbs (GILKS et al., 1996). O intuito é simular uma distribuição aleatória no espaço de θ que converge para uma distribuição estacionária, no caso, a distribuição de interesse (BROOKS & ROBERTS, 1998).

As respostas para esta abordagem são dadas por intervalos de credibilidade, disponíveis por meio do método de MCMC, os quais formam regiões simultâneas credíveis para que conjuntos de rendimentos possam ser determinados (BESAG & KOOPERBERGH, 1995).

O método de MCMC é uma técnica de uso geral para a geração de valores com probabilidade no espaço de alta dimensão, que se utiliza de números aleatórios, extraídos de probabilidade uniforme em determinado intervalo (GAMERMAN, 1997).

A utilização de técnicas Bayesianas em análises dialélicas na cultura do milho ainda é escassa. Neste contexto, o objetivo deste trabalho foi obter a capacidade geral de combinação de linhagens de milho e a capacidade específica das combinações híbridas para rendimento de grãos, usando o procedimento Bayesiano, implementado pelo algoritmo de Metropolis-Hastings, como metodologia de inferência científica para a predição dos parâmetros de interesse.

2 MATERIAL E MÉTODOS

O presente trabalho foi conduzido na Estação Experimental da Syngenta em Cascavel, no Estado do Paraná – Brasil. A unidade de pesquisa situa-se no município de Cascavel, região Oeste do Estado do Paraná (Latitude 24°53' Sul; Longitude 53° 23' a Oeste de Greenwich e a 800 m de altitude. De acordo com a classificação climática de Köppen, Cascavel situa-se na região climática Cfb (IAPAR, 1987). O solo do local é classificado como Latossolo Vermelho distroférico típico, conforme Embrapa (2006).

2.1 Material genético

O campo de polinização foi instalado em setembro de 2010, para o qual utilizaram-se dez linhagens de milho para a obtenção de híbridos (Tabela 1). Constituiu-se, portanto, um arranjo dialélico $n(n - 1/2)$ completo de milho com 45 híbridos simples F1's.

Tabela 1. Relação das linhagens utilizadas nos cruzamentos.

Linhagem	Genealogia	Origem	Código
L-111	Super precoce (Americanas)	B73	L1
L-721	Super precoce (Americanas)	B73	L2
L-883	Super precoce (Americanas)	Lancaster	L3
L-976	Precoce Tropical	Stiff Stalk Sintetic	L4
L-101	Precoce Tropical	Stiff Stalk Sintetic	L5
L-114	Precoce Tropical	Pop Suwan	L6
L-127	Precoce Tropical	Tuxpeno	L7
L-139	Precoce Tropical	Pop Suwan	L8
L-14	Ciclo normal Tropical	No Stiff Stalk Sintetic	L9
L-15	Ciclo normal Tropical	No Stiff Stalk Sintetic	L10

Os cruzamentos foram realizados manualmente entre as plantas dos pares de fileiras. Cada fileira teve cinco metros de comprimento. Foram realizadas todas as polinizações possíveis de espigas por par de fileiras, para a obtenção de suficiente quantidade de sementes para os experimentos (Tabela 2).

Tabela 2. Esquema de cruzamentos das 10 linhagens de milho para a geração dos híbridos simples.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
L1	---	HS1	HS2	HS3	HS4	HS5	HS6	HS7	HS8	HS9
L2			HS10	HS11	HS12	HS13	HS14	HS15	HS16	HS17
L3				HS18	HS19	HS20	HS21	HS22	HS23	HS24
L4					H25	H26	H27	H28	H29	H30
L5						HS31	HS32	HS33	HS34	HS35
L6							HS36	HS37	HS38	HS39
L7								HS40	HS41	HS42
L8									HS43	HS44
L9										HS45
L10										---

2.2 Análise dialélica

A abordagem por meio do método Bayesiano, via cadeias de Markov, fornece as estimativas *a posteriori* das informações sobre as regiões de credibilidade. Esta ferramenta MCMC é de bastante utilidade no estudo de técnicas Bayesianas por ser possível inferir uma região de aceitação para determinado valor paramétrico.

Com o intuito de estudar o efeito da capacidade geral e específica de combinação (CGC e CEC) para a característica rendimento de grãos das dez linhagens e dos 45 híbridos, aplicou-se a técnica de MCMC com 100.000 iterações e “burn in” de 1.000, para determinar o efeito de CGC e CEC, construindo intervalos de credibilidade para cada informação. Para a análise, utilizou-se o aplicativo computacional SAS 9.2, por meio do procedimento PROC MCMC.

2.3 Aplicação da técnica de MCMC

Para aplicar a técnica Bayesiana, utilizou-se um número de iterações de 100 mil amostras. Para o aquecimento inicial da cadeia utilizou-se um valor inicial de “burn-in” de mil iterações. A amostragem foi realizada a cada cinco pontos (*thin*) para evitar possível autocorrelação. Estimaram-se as seguintes medidas de cada parâmetro de interesse: média,

moda, mediana, máximo e mínimo, desvios-padrão e teste de Geweke (CHEN & SHAO, 1999).

As informações *a priori* utilizadas para cada parâmetro, das estimativas de g_i e s_{ij} , foram obtidas por meio do modelo de Griffing (1956) método 4. As F1's foram estimadas no aplicativo computacional Genes (CRUZ, 2006).

Método 4: são incluídas $p(p-1)/2$ combinações, com ausência dos progenitores e dos híbridos F1's recíprocos, ou seja, avalia-se apenas os F1's,

Modelo matemático:

$$y_{ij} = m + g_i + g_j + s_{ij} + \bar{\varepsilon}_{ij}$$

y_{ij} : valor médio da combinação híbrida (ij) ($i, j = 1, 2, \dots, p, i > j$),

m : constante;

g_i, g_j : efeitos da capacidade geral de combinação do i -ésimo ou j -ésimo respectivamente;

s_{ij} : efeito da capacidade específica de combinação para as combinações híbridas entre genitores de ordem i e j ;

$\bar{\varepsilon}_{ij}$: erro experimental médio associado a ij ;

Considera-se que : $s_{ij} = s_{ji}$

3 RESULTADOS E DISCUSSÃO

Por meio do teste de Geweke, foi possível inferir que houve convergência da cadeia durante o processo de simulação, com alfa de 5%, fornecendo um indício de qualidade do modelo Bayesiano para a obtenção das informações *a posteriori* dos parâmetros genéticos de interesse.

De acordo com as estimativas obtidas após as simulações MCMC, foi possível construir os intervalos de credibilidade (95% de probabilidade), como alternativa da estimativa dos parâmetros de interesse, que permitem a tomada de decisão para as dez linhagens avaliadas em cruzamentos (Tabela 3). De acordo com estes intervalos de credibilidade, é possível observar diferenças significativas entre as linhagens.

De acordo com as estimativas, a linhagem 6 apresentou maior CGC, com valor da média *a posteriori* de 1246,7 e intervalo de credibilidade Bayesiano (HPD) de 963,4 e 1567,0 inferior e superior, respectivamente. Foi possível constatar que este intervalo que compreende o valor médio da CGC da linhagem 6 não contém o valor zero, indicando sua superioridade, acima da média do ensaio (P=95%). Os valores da moda e mediana da distribuição marginal *a posteriori* foram de 1069,5 e 1249,7, respectivamente (Tabela 3).

Nas Figuras 3 e 4, é possível verificar um resumo da distribuição marginal *a posteriori* para estimativa da capacidade geral de combinação, determinada por meio do procedimento Bayesiano, implementado pelo método de Monte Carlo via cadeias de Markov. Pode-se observar a convergência da cadeia para todas as linhagens avaliadas, isto é, atingindo a estacionaridade da distribuição marginal *a posteriori*. Com isso, pode-se inferir que as estimativas de CGC são, com certeza, as médias da distribuição *a posteriori*, uma vez que as estimativas encontram-se no espaço parâmetro.

É possível ainda verificar, na Figura 3 e 4, a ausência de autocorrelação da cadeia MCMC (com Thin=2), o que demonstra que o amostrador de Metropolis-Hastings foi eficiente na geração dos dados *a posteriori*. Segundo Ehlers & Brooks (2008), o gráfico de autocorrelação, deve apresentar um “decaimento” o mais imediato possível nas barras do gráfico, a fim de se obter estimativas mais confiáveis para os parâmetros de interesse.

As linhagens 5, 9 e 10 apresentaram valor médio posterior de CGC de 893,5; 470,5 e 468,5, respectivamente, com regiões de credibilidade (HPD) de 602,1-1202,5, para a linhagem 5, 163,1-769,0, para a linhagem 9, e 158,4-758,0 para a linhagem 10. Observa-se que os intervalos de credibilidade não incluem a média posterior da CGC da linhagem 6,

confirmando a sua superioridade para CGC. Porém, os intervalos de credibilidade da CGC das linhagens 5, 9 e 10, observados na simulação do MCMC, não contêm o valor zero, indicando também superioridade, acima da média do ensaio (P=95%).

As linhagens 4 e 8 apresentaram valor de CGC de 232,8 e 171,4, respectivamente. No entanto, as estimativas dos intervalos de credibilidade *a posteriori* (95% de probabilidade) apresentaram valor negativo no seu limite inferior (-64,8 e -129,8, respectivamente), indicando que se encontram abaixo da média do ensaio. Estas linhagens seriam descartadas do processo de seleção. O fato de que o valor zero poder ser assumido no processo de simulação da distribuição *a posteriori* implica a não contribuição para o incremento ou redução da característica rendimento de grãos, do ponto de vista da capacidade geral de combinação, neste caso.

A linhagem 2 apresentou o pior desempenho para CGC, com valor médio *a posteriori* de -2849,7 e intervalo de credibilidade Bayesiano de -3154,3–2555,1 (95% de probabilidade). É possível inferir com confiabilidade que o valor estimado de CGC *a posteriori* não compreende o valor de zero dentro da região de credibilidade (HPD), indicando que a CGC seria negativa e, portanto, a linhagem contribui para a redução da característica rendimento de grãos.

Tabela 3. Estimativas *a posteriori* e intervalo HPD de 95% para os efeitos genéticos do quadro de cruzamentos dialélicos, capacidade geral e específica de combinação (\hat{G}_i e \hat{S}_{ij}) por meio do método de MCMC com 100.000 iterações e “burn in” de 1.000, para a característica rendimento de grãos (MCSE: erro-padrão de Monte Carlo; SD: Desvio-padrão).

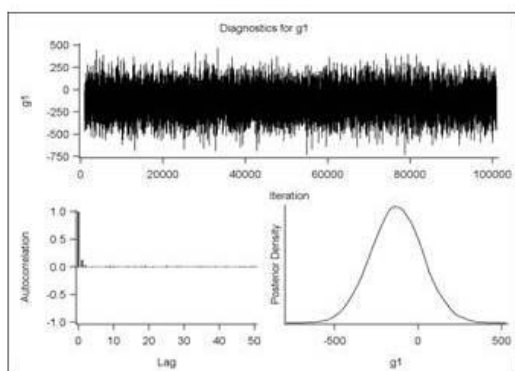
Par.	Linhasgens	Resumo da distribuição <i>a posteriori</i>						Intervalos <i>a posteriori</i>		Erro-padrão de Monte Carlo		Diagnóstico de Geweke	
		Média	Perct. 50%	Moda	Mediana	Máximo	Mínimo	HPD Intervalo		MCSE	SD	z	Pr > z
g ₁	L1	-128,6	-128,8	-90,3	-128,8	470,6	-734,2	-426,7	176,4	1,24	154,3	0,13	0,9
g ₂	L2	-2849,0	-2849,7	-2838,7	-2849,7	-2256	-3432,5	-3154,3	-2555,1	1,24	153,6	-1,24	0,22
g ₃	L3	-66,9	-68,3	-141	-68,3	584	-649,6	-370,5	228,4	1,24	153	-1,17	0,24
g ₄	L4	232,8	230,0	196,4	230	809,8	-298,3	-68,9	526,3	1,52	152,4	-1,46	0,15
g ₅	L5	893,5	894,6	829,9	894,6	1519,9	346	602,1	1202,5	1,54	153,5	0,29	0,77
g ₆	L6	1246,9	1249,7	1069,5	1249,7	1862,1	688,8	963,4	1567	1,55	154,7	-0,31	0,76
g ₇	L7	-89,5	-90,1	-154,3	-90,1	555	-738,8	-386,8	209,6	1,55	153,6	-2,17	0,03
g ₈	L8	171,4	171,5	99,2	171,5	846,1	-437,2	-129,8	470,1	1,58	153,3	-0,39	0,69
g ₉	L9	470,5	471,0	412,6	471	1027,4	-58,8	163,1	769	1,55	154,5	-0,67	0,5
g ₁₀	L10	468,7	467,9	422,4	467,9	1059,5	-104,3	158,4	758,7	1,61	154,5	-1,17	0,24
Híbridos													
s _{1x2}	H1	-3039,2	-3039,16	-3046,5	-3039,2	-2959,3	-3121,2	-3077,8	-2998,5	0,22	20,1	-1,25	0,21
s _{1x3}	H2	-2775,8	-2777,53	-2915,5	-2777,5	-1118,6	-4625,2	-3578,6	-1957,3	4,64	412,2	0,74	0,46
s _{1x4}	H3	1228,7	1231,35	1205	1231,4	2768,5	-352,9	440,7	2030,6	5,15	408,8	-1,3	0,19
s _{1x5}	H4	1619,3	1617,93	1839,8	1617,9	3109	-120,5	801,3	2433,8	5,11	418	1,85	0,06
s _{1x6}	H5	692,6	682,88	-84,4	682,9	2367	-941,7	-132,4	1461,3	5,05	409,1	-0,57	0,57
s _{1x7}	H6	1403,4	1402,18	539,5	1402,2	3050,3	-162,7	637,1	2261,2	5,38	416,5	-0,23	0,81

Tabela 3, continuação...

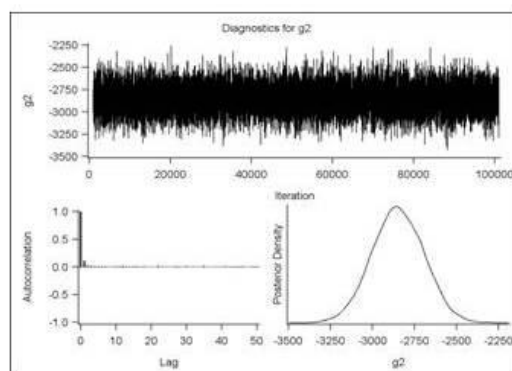
S _{1x8}	H7	2160,3	2163,92	1321,4	2163,9	3692,1	666,9	1290,8	2938,3	4,99	417,8	-0,57	0,57
S _{1x9}	H8	305,8	299,29	285,4	299,3	2020,2	-1313	-505,2	1119,8	4,95	415,4	0,99	0,32
S _{1x10}	H9	334,3	292,17	128,8	331,9	1903,3	-1266,7	-494,1	1092,3	4,88	409,3	-1,38	0,17
S _{2x3}	H10	-324,2	-320,94	-575,7	-320,9	1252,3	-1810,5	-1136,8	465,7	4,61	409,8	0,43	0,67
S _{2x4}	H11	692,2	691,50	545	691,5	2441,5	-939,4	-123,7	1497,9	4,6	412,8	-0,1	0,92
S _{2x5}	H12	456,6	456,81	453,9	456,8	544,3	375,4	418,1	496,2	0,22	19,9	0,17	0,87
S _{2x6}	H13	2615,0	2614,83	2598,3	2614,8	2698,7	2539,5	2574,4	2652,8	0,22	20,1	-0,99	0,32
S _{2x7}	H14	-494,3	-494,53	-516	-494,5	1153,7	-2090,3	-1283,6	308,0	4,64	408,1	0,59	0,55
S _{2x8}	H15	-695,1	-694,79	-1007,4	-694,8	1011,1	-2340,3	-1524,5	79,0	4,61	412,5	-0,53	0,59
S _{2x9}	H16	1348,8	1349,50	1004,4	1349,5	2886,3	-228,1	538,2	2153,1	4,55	412,6	0,42	0,68
S _{2x10}	H17	904,6	906,09	816,7	906,1	2390,6	-738,1	104,3	1690,1	4,69	409,6	-1,1	0,27
S _{3x4}	H18	618,7	618,78	277	618,8	2387,7	-1068,7	-178	1432,4	4,45	412,7	-0,68	0,5
S _{3x5}	H19	1210,2	1212,21	1072,7	1212,2	2747	-531,1	387,1	1976,1	4,63	409,4	-1,26	0,21
S _{3x6}	H20	567,2	569,32	962,6	569,3	2218,7	-1143,7	-297	1353,4	4,55	419,2	-0,56	0,58
S _{3x7}	H21	370,9	372,83	127	372,8	1941,2	-1442	-427,3	1182,7	4,78	415	-0,68	0,5
S _{3x8}	H22	1226,6	1225,34	1302,8	1225,3	2906,2	-324,7	423,8	2040,2	4,55	412,4	-1,32	0,19
S _{3x9}	H23	737,8	735,44	598,9	735,4	2425,9	-738,5	-36,9	1560,9	4,78	410,2	1,39	0,16
S _{3x10}	H24	1061,6	1058,16	1203,2	1058,2	2559	-641,8	259,5	1880,4	4,52	416,8	-1,5	0,13
S _{4x5}	H25	-834,8	-833,60	-1453,9	-833,6	792,1	-2338,2	-1682,1	-62,7	4,65	412,9	0,02	0,98
S _{4x6}	H26	84,2	85,59	-226	85,6	1749,3	-1507,2	-727,6	893,6	4,63	415,2	-1,11	0,27
S _{4x7}	H27	-360,3	-360,04	-203,1	-360	1089,4	-1959,3	-1166,9	425,6	4,48	409	1,26	0,21
S _{4x8}	H28	701,4	698,99	749,9	699	2402,7	-801,4	-110,3	1514,8	4,66	415	2,37	0,02
S _{4x9}	H29	348,3	345,30	320,8	345,3	2076,8	-1353,3	-478,2	1161,9	4,86	420,6	-0,89	0,37
S _{4x10}	H30	259,5	262,27	628,3	262,3	1863,8	-1377,8	-573,5	1061,9	4,75	415,5	0,01	0,99

Tabela 3, continuação...

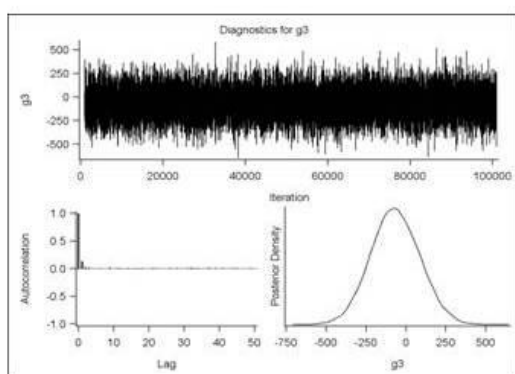
S _{5x6}	H31	-596,4	-592,97	-659,3	-593	975,7	-2244,6	-1408,3	194,2	4,85	410,5	0,16	0,87
S _{5x7}	H32	547,2	543,71	112,4	543,7	2247,7	-1041,7	-291,5	1332,8	4,69	415,1	-2,67	0,01
S _{5x8}	H33	559,7	559,99	714,1	560	2113,8	-1071,1	-249,3	1363,5	4,62	412,6	0,28	0,78
S _{5x9}	H34	-208,1	-205,55	95,4	-205,5	1587,1	-1810,2	-1013,1	586,8	4,63	413,9	-0,02	0,98
S _{5x10}	H35	-176,1	-174,72	137,8	-174,7	1440,2	-1713,8	-971,2	609,6	4,56	408,7	-1,37	0,17
S _{6x7}	H36	-39,5	-40,52	-19,1	-40,5	1599,9	-1638,7	-852,2	764,1	4,63	411,9	-0,33	0,74
S _{6x8}	H37	-992,8	-991,96	-920,8	-992	593,8	-2760,1	-1813,8	-198	4,43	413,7	1,23	0,22
S _{6x9}	H38	455,3	454,70	307,9	454,7	2246,6	-1098,5	-359,6	1261,9	4,6	415,6	0,57	0,57
S _{6x10}	H39	-228,2	-227,37	-420	-227,4	1560,3	-1984,3	-1027,6	574,7	4,49	411,4	-0,43	0,67
S _{7x8}	H40	1110,2	1109,45	1378,1	1109,5	2705	-484,8	272,7	1911,7	4,74	418,4	-0,33	0,74
S _{7x9}	H41	167,2	167,26	243,8	167,3	1660,2	-1514,9	-626,2	987	4,61	412,7	-1,81	0,07
S _{7x10}	H42	-14,1	-16,36	446,6	-16,4	1831,7	-1590,5	-851,8	769,4	4,66	412,6	-0,39	0,7
S _{8x9}	H43	-798,1	-800,00	-1087,1	-800	685,1	-2618,5	-1589,7	14	4,56	410,3	-1,48	0,14
S _{8x10}	H44	-510,1	-510,96	-988,3	-511	1065,4	-2294,7	-1310	283,3	4,5	410,8	-0,5	0,62
S _{9x10}	H45	440,7	441,06	411,9	441,1	2049,7	-1158,8	-370,7	1254,3	4,91	416,3	-0,83	0,41



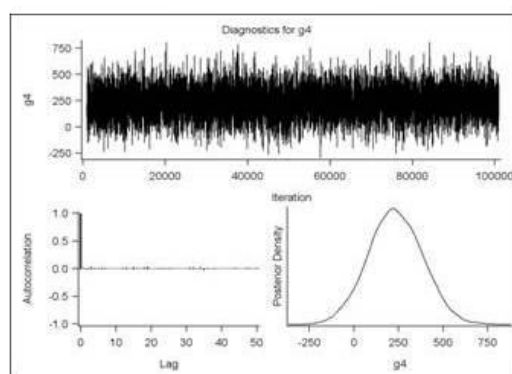
Linhagem 1



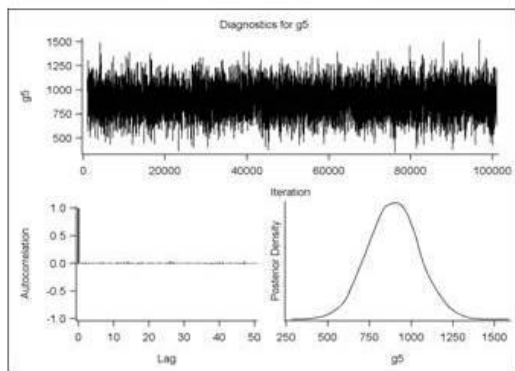
Linhagem 2



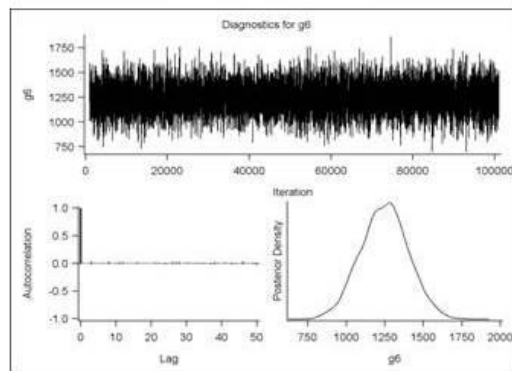
Linhagem 3



Linhagem 4



Linhagem 5



Linhagem 6

Figura 3. Estimativas da capacidade geral de combinação "a posteriori" dos genótipos avaliados: linhagens 1, 2, 3, 4, 5 e 6.

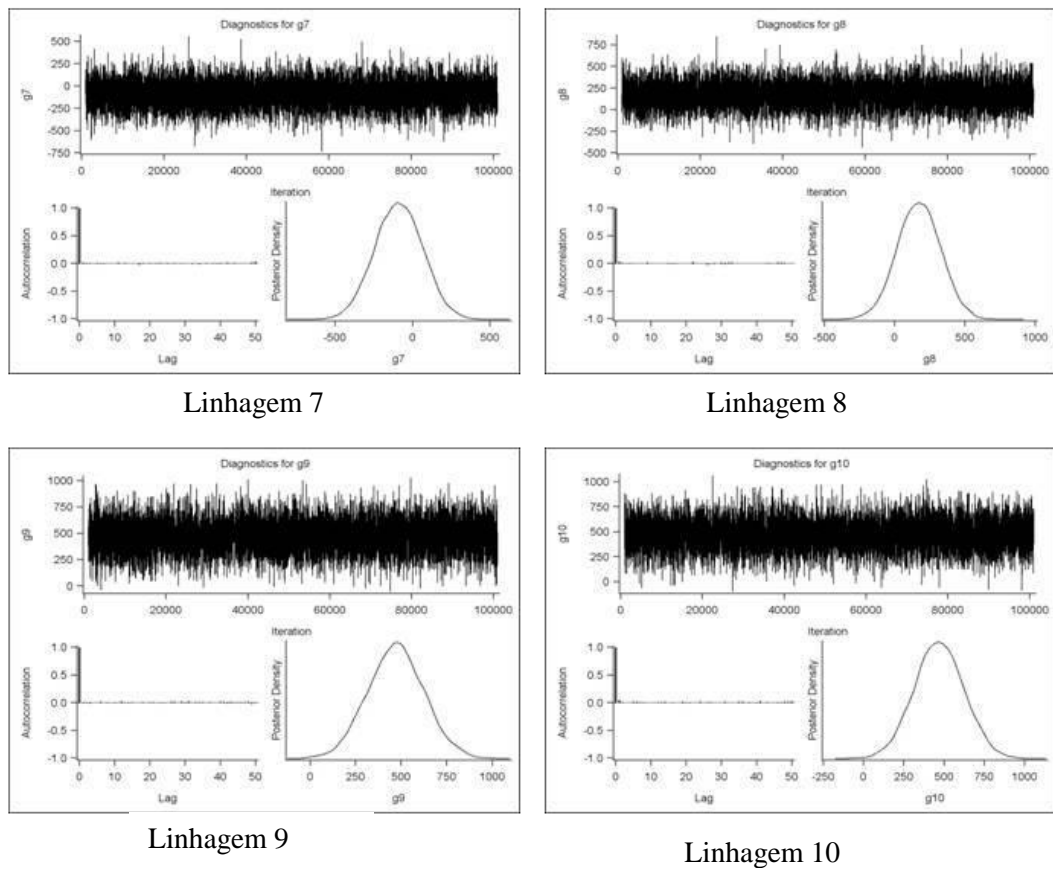


Figura 4. Estimativas da capacidade geral de combinação *a posteriori* dos genótipos avaliados: linhagens 7, 8, 9 e 10.

Na Tabela 3, são apresentadas as estimativas *a posteriori* e regiões de credibilidade Bayesiano (HPD; 95% de probabilidade) para os efeitos genéticos representados pela capacidade específica de combinação (\hat{S}_{ij}) do rendimento de grãos, para os 45 híbridos simples de milho.

A melhor combinação entre as linhagens foi o cruzamento 2x6 (híbrido 13), que apresentou estimativa *a posteriori* de 2615,0 e intervalo de credibilidade (HPD) de 2574,4-2652,8, não compreendendo o valor zero entre os limites de credibilidade, indicando a superioridade genética.

As informações sobre a convergência, como a análise visual dos traços durante a simulação, também foram consistentes com a estacionariedade; não houve auto-correlação das cadeias MCMC durante o processo de simulação estocástica para CEC, garantindo a estimativa dentro do espaço parâmetro.

A segunda melhor estimativa de CEC obtida após a simulação MCMC foi para combinação 1x8, com estimativa *a posteriori* de 2160,3 e intervalo de credibilidade bayesiano de 1290,8 a 2938,3. Os valores da moda e mediana da distribuição marginal foram de 1321,4 e 2163,9. A região de credibilidade da CEC do cruzamento 2x6 (híbrido 13) inclui o valor da média *a posteriori* do cruzamento 1x8, indicando a superioridade genética de acordo com sua capacidade específica de combinação.

No presente estudo, as estimativas de capacidade específica de combinação para as combinações híbridas 2x10; 3x10; 7x8; 3x5; 3x8; 1x4; 2x9 e 1x7 apresentam estimativas *a posteriori* positivas e estatisticamente diferentes de zero de acordo com as regiões de credibilidade (Tabela 3).

As diferenças significativas encontradas nos intervalos de credibilidade para os efeitos para capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC) indicam a existência de variabilidade originada de efeitos gênicos aditivos e não aditivos (LEMOS et al., 2002). Estes resultados estão em concordância com os encontrados por Parentoni et al. (1991) e Lemos et al. (2002), em trabalhos com cruzamentos dialélicos entre dez linhagens de milho.

A característica mais importante para a indústria do milho é sem dúvida o rendimento de grãos. Portanto, para o caráter, a linhagem 6, é promissora do ponto de vista da CGC, sendo a mais favorável em aumentar o caráter. Para ganhos genéticos de longo prazo no programa de melhoramento, além da linhagem 6, as linhagens 5, 9 e 10, seriam também adequadas devido a sua alto desempenho em rendimento de grãos.

De acordo com diversos autores (SPRAGUE & TATUM, 1942; SCAPIM et al., 2002; LEMOS et al., 2002), baixas estimativas de CGC indicam genótipos com combinações que não diferem muito da média de todos os cruzamentos no sistema dialélico, enquanto altos valores (com sinal positivo ou negativo) indicam genótipos melhores ou piores que os restantes, com os quais se compara, e estes valores indicam, portanto, a importância dos genes de efeitos predominantemente aditivos.

Poucos trabalhos têm sido relatados quanto à análise dialélica em espécies vegetais usando uma abordagem Bayesiana. No presente estudo, a implementação do modelo Bayesiano, implementada pelo algoritmo de Metropolis-Hastings, foi útil para identificar diferenças estatísticas dos parâmetros estimados da CEC e CGC, com base nas regiões de credibilidade. Esta seria a principal vantagem da implementação desta metodologia. Neste contexto, as vantagens confirmadas neste estudo estão em concordância com Resende (2002): A análise Bayesiana propicia resultados adicionais aos obtidos pelo método clássico,

destacando-se os intervalos de credibilidade Bayesianos para as estimativas paramétricas; tais estimativas obtidas usando os algoritmos MCMC podem ser mais precisas que as estimativas clássicas; e a análise Bayesiana é uma técnica elegante e flexível que permite a estimação simultânea dos parâmetros genéticos de maneira precisa, incluindo amostras de tamanho finito. Entretanto, Resende (2002) afirma que, para a obtenção de uma apropriada precisão do método Bayesiano, é recomendável que o pesquisador tenha uma quantidade suficiente de informação experimental *a priori*. No presente estudo, este fator foi considerado por meio das estimativas clássicas do modelo de Griffing (1956) método 4.

4 CONCLUSÕES

A análise Bayesiana fornece resultados adicionais aos obtidos pelo método clássico, destacando-se os intervalos de credibilidade para as estimativas da capacidade geral e específica de combinação.

Neste contexto, as linhagens 6, 5, 9 e 10 foram as que apresentaram maiores valores para capacidade geral de combinação para rendimento de grãos, podendo-se inferir, portanto, que são promissoras para a utilização no programa de melhoramento genético. Os híbridos que mais se destacaram foram as combinações 2x6 e 1x8, para a característica rendimento de grãos.

5 REFERÊNCIAS

- BERNARDO, R. genetic models for predicting maize single-cross performance in unbalanced yield trial data. **Crop Science**, v. 35, n. 1, p. 141-147, 1995.
- BESAG J.; KOOPERBERG C. L. On conditional and intrinsic autoregressions. **Biometrika**. Oxford, v. 82, p. 733-746, 1995.
- BROOKS, S. P.; ROBERTS G. O. Diagnosing convergence of Markov chain Monte Carlo algorithms. **Statistics and Computing**, London, v. 8, p. 319-335, 1998.
- CHEN, M. H.; SHAO, Q. M. Monte Carlo estimation of bayesian credible and HPD Intervals. **Journal of Computational and Graphical Statistics**. Alexandria, v. 8, p. 69-92, 1999.
- CRUZ, C. D. **Programa Genes: versão Windows: aplicativo computacional em genética e estatística**. Viçosa: UFV, 2006.
- CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2. ed. Viçosa: UFV, 1994. 390p.
- EHLERS, R. S.; BROOKS, S.P. Adaptive proposal construction for reversible jump MCMC. **Scandinavian Journal of Statistics**, Nordic statistical associations, v. 35, p. 677-690, 2008.
- EMBRAPA. Centro Nacional e Pesquisa em Solos. **Sistema Brasileiro de Classificação de Solos**. Brasília: Embrapa-SPI; Rio de Janeiro: Embrapa-Solos. 2006. 306p.
- IAPAR-Instituto Agrônômico do Paraná. **Cartas climáticas básicas do Estado do Paraná**. Londrina: Iapar, 1987.
- GAMMERMAN, D. **Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. Chapman and Hall, London, 1997. 344p.
- GARDNER, C. O.; EBERHART, S. A. Analysis and interpretation of variety cross diallel and realated populations. **Biometrics**, Washington. v. 22, p. 439-52, 1966.
- GILKS W. R.; RICHARDSON, S.; SPIEGELHALTER, D.J. Markov Chain Monte Carlo in Practice. London: **Chapman & Hall/CRC Interdisciplinary Statistics**, 1996. 512p.
- GRIFFING, J.B. Concept of general and specific combining ability in relation to diallel systems. **Australian Journal of Biological Science**, East Melbourne, v. 9, p. 463-493, 1956.
- HAYMAN, B.I. The analysis of variance of diallel crosses. **Biometrics**, Arlington, v. 10, p. 235-244, 1954.
- HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph : University of Guelph, 1984. 462p.

LEMOS, M. A.; GAMA, E. E. G.; MENEZES, D.; SANTOS, V. F.; TABOSA, J.N. Avaliação de dez linhagens e seus híbridos de milho superdoce em um dialelo completo. **Horticultura Brasileira**, Brasília, v. 20, n. 2, p. 167-170, 2002.

METROPOLIS, N.; ROSENBLUTH, A. W.; MARSHALL, N. R.; TELLER, A. H. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, Melville, v, 21, p. 1087–1092, 1953.

PARENTONI, S. N.; GAMA, E. E. G.; REIFSCHNEIDER, F. J. B.; GUIMARÃES, P.E.O. Avaliação da capacidade combinatória de dez linhagens de milho doce. **Horticultura Brasileira**, Brasília. v. 9, n.2, p.71-73, 1991.

RESENDE, M. D. V. Software Selegen–REML/BLUP. **Embrapa Florestas, Documentos 77**, Colombo. 2002. 67p.

SCAPIM, C. A.; PACHECO, C. A. P.; TONET, A.; BRACCINI, A. D. L.; PINTO, R.J.B. Análise dialélica e heterose de populações de milho-pipoca. **Bragantia**, Campinas, v. 61, n. 3, p. 219-230, 2002.

SPRAGUE, G. F., TATUM, L. A. General vs. specific combining ability in single crosses of corn. **Journal American Society of Agronomy**, Madison, v. 34, p. 923-932, 1942.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica no fitomelhoramento**. Ribeirão Preto : Sociedade Brasileira de Genética, 1992. 496p.

CAPÍTULO 3
AVALIAÇÃO GENOTÍPICA DE HÍBRIDOS SIMPLES DE MILHO EM
MULTILOCAIS POR MEIO DE MODELOS MISTOS

RESUMO

Um ponto fundamental no processo de seleção são os experimentos voltados ao estudo de interação genótipos x ambientes, isto é, à resposta diferenciada de cada genótipo em diferentes ambientes. Com base nisso, o objetivo desse trabalho foi avaliar o desempenho, a adaptabilidade e a estabilidade de 43 híbridos simples de milho em ensaios de rede, pela metodologia de modelos lineares mistos. Os ensaios foram realizados em oito ambientes no período de dois anos. As parcelas consistiram de duas linhas de 5 m, espaçamento entre as linhas de 0,80 m e duas repetições em blocos. Para as análises de adaptabilidade e estabilidade, foram utilizados os dados de rendimento de grãos, seguindo o modelo 114 do software SELEGEN-REML/BLUP. De acordo com os valores genotípicos, os dez melhores híbridos, considerando os efeitos interação de ambiente e a média genotípica, foram: 14, 32, 13, 4, 34, 6, 33, 22, 10 e 37. Conclui-se que a metodologia de modelos lineares mistos é alternativa útil e adequada para a avaliação genotípica de híbridos simples de milho em multilocais.

Palavras-chave: REML, BLUP, MHPRVG.

GENOTYPIC ASSESSMENT OF SIMPLE HYBRIDS OF CORN IN MULTISITE VIA MIXED MODELS

ABSTRACT

A fundamental point in selection process is the experiments towards the study on genotypes vs environments interaction, that is, the different responses of each genotype in different environments. On this basis, the aim of this study was to assess the performance, adaptability and stability of single-cross maize hybrids through network testing competition, by using the methodology of mixed linear models. The trials were carried out in eight environments, during two years. All field plots consisted of two lines of 5m – 0.80m apart from each other – and a double block. For analysis of adaptability and stability, data on grain yield was used, according to model 114 of the software SELEGEN-REML/BLUP. Based on genotypic values, the ten best hybrids – considering interaction and environments effects and the genotypic mean – were 14, 32, 13, 4, 34, 6, 33, 22, 10 and 37. In conclusion, the linear models methodology is a useful and adequate alternative for genotypic assessment of single-cross maize hybrids at multisite.

Keywords: Restricted Maximum Likelihood (REML), Best Linear Unbiased Prediction (BLUP), Harmonic Mean of Relative Performance of Genotypic Value (MHPRVG).

1 INTRODUÇÃO

A cultura do milho (*Zea mays*) é uma das mais importantes no setor agrícola brasileiro. A vasta utilização dos grãos a coloca como um dos principais cereais, cuja utilização vai desde a produção destinada para alimentação humana e animal a fontes de energia renovável.

Nos programas de melhoramento de milho, as técnicas de avaliação genética exercem papel fundamental, pois permitem a predição dos valores genéticos aditivos e genotípicos dos candidatos à seleção. Tais técnicas são relevantes tanto para o melhoramento intrapopulacional quanto interpopulacional, visando à utilização de híbridos heteróticos (RESENDE, 2002). O maior problema no melhoramento genético vegetal em geral está na predição dos valores genéticos dos diversos genótipos avaliados. Tal predição requer que o conhecimento dos componentes de variância sejam conhecidos ou precisamente estimados (RESENDE, 2004).

Avanços nos métodos genéticos de avaliação têm sido desenvolvidos desde a metade do século XX, com o intuito de solucionar o problema da predição dos valores genéticos (HENDERSON, 1988). Do ponto de vista frequentista, o estimador de variância REML (*Restricted Maximum Likelihood* - Máxima Verossimilhança Restrita), seguido do preditor de valores genéticos BLUP (*Best Linear Unbiased Predictor* - Melhor Preditor Linear Não-Viesado) constitui o método mais eficiente para a identificação de indivíduos superiores, o que permite alcançar o máximo ganho genético na avaliação dos genótipos.

Estes métodos são associados a um modelo linear misto, isto é, modelo que contém efeitos fixos, além da média geral, e efeitos aleatórios, além do erro (RESENDE, 2004).

No melhoramento clássico, na avaliação genética de indivíduos, admite-se que os genótipos sejam de efeitos fixos, o que pode gerar distorções nas avaliações e inferir de forma direta na obtenção das estimativas dos valores genéticos (HENDERSON, 1974; 1975; 1984).

A metodologia de modelos lineares mistos foi desenvolvida por Henderson (1949) para minimizar as distorções na análise clássica, em que genótipo é tido como fixo. Na abordagem de modelos lineares mistos os efeitos de genótipo podem ser classificados como aleatórios ou fixos.

Para efeitos aleatórios, os valores genéticos de cada indivíduo são preditos por meio do preditor BLUP. Para efeitos fixos, são estimados os valores fenotípicos via os estimadores

BLUE (*Best Linear Unbiased Estimator* - Melhor Estimador Linear Não-Viesado) (HENDERSON, 1974; 1975; 1984).

Indiferente dos preditores “Best”, é necessário obter estimativas exatas dos componentes de variância. Os procedimentos usuais vão desde os métodos dos mínimos quadrados até o método da máxima verossimilhança restrita (REML) (SEARLE et al., 1992). O método REML possui propriedades estatísticas superiores quando comparadas aos demais métodos de estimação dos parâmetros genéticos (RESENDE et al., 2008).

Um ponto fundamental no processo de seleção de cultivares são os experimentos em que são feitos estudos de interação genótipos x ambientes, isto é, a resposta diferenciada de cada genótipo em diferentes ambientes. Estudos de interação genótipos x ambientes são importantes para programas de melhoramento, pois tornam possível a seleção de cultivares com diferentes comportamentos. Para identificar o comportamento de cada genótipo em estudo, em diferentes ambientes, devem ser utilizadas técnicas de adaptabilidade e estabilidade. Tais informações fornecem subsídios para identificar cultivares com comportamento previsível e que sejam responsivos de acordo com as mais diferentes variações ambientais (DUARTE & VENCOVSKY, 1999).

São várias as metodologias desenvolvidas para estudo de adaptabilidade e estabilidade de cultivares de milho. Entre elas, destacam-se metodologias baseadas em análise de variância, regressão, métodos não paramétricos e multivariados (RESENDE, 2007).

Recente metodologia para estudo de adaptabilidade e estabilidade denominada MHPRVG (Média Harmônica da Performance Relativa do Valor Genotípico) foi proposta por Resende (2007). A metodologia MHPRVG é capaz de selecionar três atributos simultaneamente: adaptabilidade, estabilidade e produtividade.

O objetivo do presente trabalho foi avaliar o desempenho, a adaptabilidade e a estabilidade de híbridos simples de milho em ensaios de rede, utilizando a metodologia de modelos lineares mistos, via estimador REML, para os componentes de variância, e preditor BLUP, para os valores genéticos.

2 MATERIAL E MÉTODOS

2.1 Locais de avaliação

O presente trabalho utilizou dados de híbridos simples de milho disponibilizados da rede de ensaios da empresa Syngenta divisão Seeds em oito locais: 1 - Campo Mourão-PR, 2 - Palmeira das Missões-RS, 3 - Toledo-PR 4 - Cruz Alta-RS, 5 - Três Barras-PR, 6 - Capão Bonito-SP, 7 - Itapeva-SP, 8 - Passo Fundo-RS nos anos de 2009 e 2010.

2.2 Genótipos avaliados

Foram avaliados 43 híbridos de estágio avançado de avaliação, obtidos do cruzamento de linhagens endogâmicas.

2.3 Modelo de avaliação genética

Foi utilizado o seguinte modelo de avaliação genética:

$$y = Xf + Zg + Oa + ti + Wt + e$$

em que:

y é o vetor de observações fenotípicas;

f é o vetor dos efeitos das combinações repetição-local-ano (assumidos como fixos) somados à média geral; g é o vetor dos efeitos genotípicos (assumidos como aleatórios);

a é vetor dos efeitos da interação de genótipos com anos (aleatórios); i é o vetor dos efeitos da interação genótipos x locais (aleatórios);

t é o vetor dos efeitos da interação tripla genótipos x locais x anos (assumidos como aleatórios);

e é o vetor de erros ou resíduos (aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

Esse modelo é equivalente ao utilizado tradicionalmente para a análise envolvendo a avaliação de genótipos em vários locais e anos de plantio, dado pelo denominado modelo 144 do aplicativo computacional Selegen REML-BLUP (Delineamento em Blocos Completos com Interação Tripla e Estabilidade e Adaptabilidade – Método MHPRVG: Modelo 114) (RESENDE, 2002b; 2007b).

Modelo 114: estabilidade, adaptabilidade e produtividade com locais e anos: blocos, genótipos, várias plantas por parcela, vários locais.

Este modelo considera os três efeitos principais (genótipos, locais e anos), bem como suas interações. Adicionalmente, fornece medidas de estabilidade e adaptabilidade dos genótipos ao longo dos ambientes (locais e anos).

$$Y_{ijkn} = \mu + g_i + b_{j/k/n} + a_k + l_n + g_{ik} + gl_{in} + al_{kn} + gal_{ikn} + gbal_{ij/k/n}$$

em que:

μ é o efeito da média geral;

g_i é o efeito do genótipo i;

$b_{j/k/n}$ é o efeito do bloco j dentro do ano k dentro do local n;

l_n é o efeito do local n;

a_k é o efeito do ano de plantio k;

gl_{in} é o efeito da interação genótipos x locais;

ga_{ik} é o efeito da interação genótipos x anos de semeadura;

al_{kn} é o efeito da interação locais x anos de semeadura;

gal_{ikn} é o efeito da interação genótipos x anos x locais,

$gbal_{ij/k/n}$ é o erro ou resíduo aleatório.

3 RESULTADOS E DISCUSSÃO

As estimativas dos componentes de variância e parâmetros genéticos via Máxima Verossimilhança Restrita (REML), para o experimento envolvendo híbridos e locais são apresentados na Tabela 4.

Para o caráter rendimento de grãos, a herdabilidade no sentido amplo foi de $0,22 \pm 0,0295$, valor considerado baixo, pois o mesmo se encontra sem efeitos de interação condizente com o caráter avaliado (RESENDE, 2004).

Os coeficientes de determinação das interações genótipo x ano (c^2_{ga}), genótipo x local (c^2_{gl}) e genótipo x ano x local (c^2_{gla}) indicam quanto cada componente contribuiu para a variação fenotípica total (BORGES et al., 2010). Para os coeficientes de determinação das interações c^2_{ga} , c^2_{gl} e c^2_{gla} , os valores foram 0,04; 0,04 e 0,16, respectivamente. Trabalhos realizados por Carbonell et al. (2007) em feijão demonstraram coeficientes de determinação para 15 locais em dois anos, com magnitudes entre 0,165 e 0,325, semelhantes às obtidas neste trabalho para interação tripla.

Estes valores se levados a percentual representam a proporção explicada pela interação. Logo, os coeficientes de determinação indicam que os efeitos de interação tiveram pouca importância na variância fenotípica (RESENDE, 2007).

Os valores de correlações genotípicas por meio do ambiente foram considerados de magnitude alta para todas as interações (r_{gl} (0,86); r_{ga} (0,83); r_{gl_a} (0,88); r_{ga_l} (0,85) r_{gl_ma} (0,90); r_{ga_ml} (0,85); r_{gla} (0,48)) exceto para a interação tripla que foi da ordem de 0,48 (baixa magnitude). Esta correlação indica que o desempenho dos genótipos por meio dos ambientes não foi rigorosamente a mesma. Como observado, a maioria dos valores de correlação estão na faixa de 0,8 (exceto r_{gla}). Segundo Cruz & Castoldi (1991), correlações próximas a este valor indicam predominância de interação simples, ou seja, a causa dessa interação é a diferença do comportamento de genótipos em ambientes diferentes.

Resultados de baixa magnitude para a interação r_{gla} também foi observada por Borges (2009) em arroz. Segundo este autor, valores de magnitude baixos para r_{gla} indicam nível moderado de interação complexa, corroborando com a premissa de que os genótipos não apresentam mesmo comportamento nos diferentes locais avaliados.

A correlação genotípica por meio dos locais, em um dado ano (r_{gl_a}) no valor de 0,88, indica que os genótipos tiveram comportamentos semelhantes nos locais avaliados. Da mesma forma, a correlação genotípica por meio dos anos, em um dado local (r_{ga_l}) no valor

de 0,85, indica que o comportamento dos genótipos teve pouca influencia pelo efeito de ano. Dados semelhantes foram observados por Borges (2009), que encontrou valores de 0,739 e 0,759, respectivamente para r_{gl_a} e r_{ga_l} , na cultura do arroz.

Tabela 4. Estimativas dos componentes de variância (REML) e de parâmetros genéticos para rendimento de grãos kg ha⁻¹.

Estimativas	Valores
$\hat{\sigma}_g^2$	505251,26
$\hat{\sigma}_{ga}^2$	100589,53
$\hat{\sigma}_{gl}^2$	83735,26
$\hat{\sigma}_{gla}^2$	363310,57
$\hat{\sigma}_e^2$	1204386,48
\hat{h}_g^2	0,22±0,0295
c_{ga}^2	0,04
c_{gl}^2	0,04
c_{gla}^2	0,16
r_{gl}	0,86
r_{ga}	0,83
r_{gl_a}	0,88
r_{ga_l}	0,85
r_{gla}	0,48
Média geral	11525,23

$\hat{\sigma}_g^2$: variância genotípica; $\hat{\sigma}_{ga}^2$: var. da interação genótipo x ano; $\hat{\sigma}_{gl}^2$: var. da interação genótipo x local; $\hat{\sigma}_{gla}^2$: var. da interação genótipos x locais x anos; $\hat{\sigma}_e^2$: var. residual. \hat{h}_g^2 : herdabilidade no sentido amplo; c: são os coeficientes de determinação dos efeitos da interação genótipos x anos, interação genótipos x locais, e da interação genótipos x locais x anos. r_{gl} : são as correlações genotípicas por meio dos locais, válida para qualquer ano; r_{ga} : correlação genotípica dos materiais genéticos por meio dos anos, válida para qualquer local; r_{gl_a} : correlação genotípica dos materiais genéticos por meio dos locais, em um dado ano; r_{ga_l} : correlação genotípica dos materiais genéticos por meio dos anos, em um dado local; e r_{gla} : correlação genotípica dos materiais genéticos por meio dos locais e anos, respectivamente.

Na Tabela 5, são apresentados os valores da análise de deviance para as fontes de variação do modelo. Os efeitos de genótipos, das interações genótipos x anos e genótipos x locais x anos, assim como seus componentes de variação e coeficientes de determinação foram significativos.

Para o efeito de interação genótipos x locais, assim como para seus componentes de variação e coeficientes de determinação, a Análise de Deviance não foi significativa em nível de 5%, mas foi significativa em nível de 10%.

Em resumo, a análise de deviance mostra a existência de variabilidade nos genótipos, e comportamentos distintos de genótipos dentro de anos e locais, de acordo com a interação tripla.

Tabela 5. Anodev (Análise de Deviance) componentes de variância, coeficientes de determinação referentes à análise conjunta (todos os locais e genótipos).

Efeito	Deviance	LRT(χ^2)	Componente de Variância	Coeficiente de determinação
Genótipos	31212,78	160,6**	$\hat{\sigma}_g^2$: 505251,25	\hat{h}_g^2 0,224
Genótipos x Locais	31055,01	2,8*	$\hat{\sigma}_{gl}^2$: 83735,257	c_{gl}^2 0,037
Genótipos x Anos	31063,34	11,2**	$\hat{\sigma}_{ga}^2$: 100589,53	c_{ga}^2 0,045
Genótipos x Locais x Anos	31109,64	57,5**	$\hat{\sigma}_{gla}^2$: 363310,56	c_{gla}^2 0,161
Resíduo			1204386,48	0,757
Modelo Completo	31052,17			$c_{total}^2=1$

Deviance do modelo ajustado sem os efeitos correspondentes. Significativo pelo teste de qui -quadrado a 10% (2,71) **, 5% (3,84) e a 1% (6,63) *.

3.1. Seleção genotípica em todos os locais

A predição dos valores genotípicos dos híbridos encontra-se na Tabela 6, sendo estes valores referentes ao valor genotípico predito (u + g) sem efeito de interação genótipo ambiente. O valor genotípico médio (u + g + gem) são resultados que contabilizam a interação dos genótipos em diferentes ambientes.

Foram selecionados os dez primeiros genótipos. O décimo na classificação foi o genótipo 37. O valor genotípico predito foi de 12052,7, com ganho de 642,5. A nova média foi 12167,8 e seu valor genotípico médio em todos os ambientes foi de 12063,6 (Tabela 6).

O genótipo que obteve a melhor classificação foi o 14. O valor genotípico predito foi de 12353,3 (u+g) e o ganho deste genótipo foi de 828,1; logo, a nova média foi de 12353,3 e o valor genotípico médio 12370,5.

Com base em valores genotípicos, o genótipo 29 foi o que apresentou a pior classificação entre todos os genótipos avaliados. O valor genotípico predito foi de 9619,6, o

ganho com a seleção foi nulo e sua nova média e valor genotípico médio em todos os ambientes foram de 11525,2 e 9580,1, respectivamente,.

É importante ressaltar a importância do valor genotípico predito, pois com base nele é possível sugerir cultivares em ambientes nos quais não foram testadas, visto que, este valor é livre de interação com ambiente. Logo, de acordo com Resende (2002), os valores genotípicos médios são adequados apenas para recomendação de cultivares nos locais avaliados.

Como era de se esperar as magnitudes do critério u+g+gem foram superiores às do critério u+g, justamente pela capitalização a interação média.

Tabela 6. Valores de efeitos genótipos, valores genotípicos preditos (u+g), ganho com a nova média, médias melhoradas, e valores genotípicos médios nos vários ambientes u + g + gem.

Ord.	Genótipo	g	u + g	Ganho	Nova Média	u+g+gem*
1	14	828,1	12353,3	828,1	12353,3	12370,5
2	32	734,6	12259,8	781,3	12306,6	12275,0
3	13	719,7	12244,9	760,8	12286,0	12259,8
4	4	670,4	12195,6	738,2	12263,4	12209,5
5	34	660,4	12185,7	722,6	12247,9	12199,3
6	6	609,1	12134,3	703,7	12228,9	12146,9
7	33	601,4	12126,6	689,1	12214,3	12139,1
8	22	542,3	12067,6	670,7	12196,0	12078,8
9	10	532,0	12057,2	655,3	12180,6	12068,2
10	37	527,5	12052,7	642,5	12167,8	12063,6
11	17	469,5	11994,8	626,8	12152,0	12004,5
12	19	421,9	11947,2	609,7	12135,0	11955,9
13	18	403,1	11928,4	593,8	12119,1	11936,7
14	11	360,1	11885,3	577,1	12102,4	11892,7
15	16	323,6	11848,8	560,2	12085,5	11855,5
16	26	312,5	11837,7	544,8	12070,0	11844,2
17	12	304,1	11829,3	530,6	12055,8	11835,6
18	20	297,6	11822,8	517,7	12042,9	11829,0
19	41	281,7	11806,9	505,2	12030,5	11812,8
20	9	276,1	11801,3	493,8	12019,0	11807,0

Tabela 6, continuação...

21	15	222,1	11747,4	480,8	12006,1	11752,0
22	38	193,4	11718,7	467,8	11993,0	11722,7
23	43	184,8	11710,1	455,5	11980,7	11713,9
24	42	182,9	11708,2	444,1	11969,4	11712,0
25	21	76,7	11602,0	429,4	11954,7	11603,5
26	2	4,6	11529,9	413,1	11938,3	11530,0
27	40	-53,2	11472,0	395,8	11921,0	11470,9
28	28	-127,1	11398,2	377,1	11902,4	11395,5
29	23	-207,3	11317,9	357,0	11882,2	11313,6
30	3	-250,5	11274,7	336,7	11862,0	11269,5
31	24	-258,7	11266,5	317,5	11842,8	11261,1
32	35	-277,3	11247,9	298,9	11824,2	11242,2
33	8	-358,3	11167,0	279,0	11804,3	11159,5
34	27	-369,6	11155,6	260,0	11785,2	11148,0
35	25	-384,7	11140,5	241,5	11766,8	11132,5
36	7	-425,8	11099,5	223,0	11748,2	11090,6
37	31	-461,3	11063,9	204,5	11729,7	11054,3
38	39	-653,4	10871,8	181,9	11707,2	10858,2
39	36	-875,1	10650,1	154,8	11680,0	10632,0
40	5	-1117,8	10407,4	123,0	11648,2	10384,2
41	1	-1217,8	10307,4	90,3	11615,5	10282,2
42	30	-1796,6	9728,6	45,4	11570,6	9691,4
43	29	-1905,7	9619,6	0,0	11525,2	9580,1

* interação genótipo x ambiente para os 8 locais, 43 genótipos e 2 anos. Seleção de Genótipos - Todos Locais

3.2 Avaliações genotípicas com base em cada local

Nas Tabelas 7, 8, 9 e 10 a seguir, são apresentados os valores obtidos dos 43 genótipos nos oito locais nos dois anos avaliados. Nelas, podem-se observar os efeitos genotípicos (g + ge) e valores genotípicos preditos para cada local (u + g + ge).

Tabela 7. Ordenamento dos genótipos avaliados nos locais 1 e 2, no qual são apresentados os efeitos genotípicos (g +ge), os valores genotípicos preditos para os respectivos locais (u+g+ge).

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
1	1	4	849,0	11720,3	849,0	11720,3	2	1	14	1041,6	13045,2	1041,6	13045,2
1	2	14	795,3	11666,6	822,2	11693,4	2	2	13	849,2	12852,8	945,4	12949,0
1	3	32	780,7	11652,0	808,4	11679,6	2	3	32	665,2	12668,8	852,0	12855,6
1	4	13	681,4	11552,6	776,6	11647,9	2	4	10	632,5	12636,1	797,1	12800,7
1	5	33	674,6	11545,8	756,2	11627,4	2	5	37	626,1	12629,7	762,9	12766,5
1	6	37	668,7	11539,9	741,6	11612,9	2	6	34	618,7	12622,3	738,9	12742,5
1	7	34	655,5	11526,8	729,3	11600,6	2	7	33	595,8	12599,4	718,4	12722,0
1	8	6	653,2	11524,5	719,8	11591,1	2	8	19	540,6	12544,2	696,2	12699,8
1	9	17	599,9	11471,1	706,5	11577,7	2	9	6	482,0	12485,6	672,4	12676,0
1	10	10	596,0	11467,3	695,4	11566,7	2	10	11	467,2	12470,8	651,9	12655,5
1	11	22	508,9	11380,1	678,5	11549,7	2	11	4	450,9	12454,5	633,6	12637,2
1	12	9	459,2	11330,4	660,2	11531,4	2	12	22	448,1	12451,7	618,1	12621,7
1	13	19	432,8	11304,1	642,7	11514,0	2	13	17	446,3	12449,9	604,9	12608,5
1	14	18	419,6	11290,8	626,8	11498,0	2	14	18	365,5	12369,1	587,8	12591,4
1	15	41	418,9	11290,1	612,9	11484,2	2	15	16	319,1	12322,8	569,9	12573,5
1	16	11	368,7	11239,9	597,6	11468,9	2	16	20	319,1	12322,7	554,2	12557,8
1	17	16	357,2	11228,5	583,5	11454,7	2	17	38	278,6	12282,2	538,0	12541,6
1	18	12	322,2	11193,5	569,0	11440,2	2	18	15	263,2	12266,8	522,7	12526,4
1	19	38	248,9	11120,2	552,1	11423,4	2	19	12	243,8	12247,4	508,1	12511,7
1	20	20	244,5	11115,7	536,8	11408,0	2	20	43	238,3	12242,0	494,6	12498,2
1	21	26	243,6	11114,9	522,8	11394,0	2	21	26	236,3	12239,9	482,3	12485,9
1	22	15	186,2	11057,5	507,5	11378,7	2	22	9	225,1	12228,7	470,6	12474,2
1	23	42	163,6	11034,9	492,6	11363,8	2	23	42	153,9	12157,5	456,8	12460,4

Tabela 7, continuação...

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
1	24	21	125,3	10996,6	477,3	11348,5	2	24	41	78,3	12081,9	441,0	12444,7
1	25	40	70,6	10941,9	461,0	11332,2	2	25	2	57,1	12060,7	425,7	12429,3
1	26	2	14,4	10885,7	443,8	11315,1	2	26	21	46,5	12050,2	411,1	12414,7
1	27	43	-17,1	10854,1	426,7	11298,0	2	27	40	-100,7	11902,9	392,2	12395,8
1	28	35	-149,6	10721,7	406,2	11277,4	2	28	24	-138,9	11864,7	373,2	12376,8
1	29	28	-157,5	10713,8	386,7	11258,0	2	29	28	-167,4	11836,2	354,5	12358,2
1	30	24	-260,5	10610,8	365,1	11236,4	2	30	23	-206,5	11797,1	335,8	12339,4
1	31	3	-262,8	10608,5	344,9	11216,1	2	31	35	-283,7	11719,9	315,9	12319,5
1	32	23	-294,4	10576,8	324,9	11196,2	2	32	27	-300,2	11703,4	296,6	12300,2
1	33	27	-380,0	10491,2	303,6	11174,8	2	33	31	-346,9	11656,7	277,1	12280,7
1	34	7	-384,6	10486,6	283,3	11154,6	2	34	8	-370,6	11633,0	258,1	12261,7
1	35	8	-499,2	10372,1	261,0	11132,2	2	35	25	-428,1	11575,5	238,4	12242,1
1	36	25	-584,4	10286,9	237,5	11108,7	2	36	3	-457,1	11546,6	219,1	12222,7
1	37	39	-649,5	10221,7	213,5	11084,7	2	37	7	-649,4	11354,2	195,7	12199,3
1	38	31	-650,1	10221,1	190,8	11062,0	2	38	39	-670,6	11333,0	172,9	12176,5
1	39	36	-806,1	10065,1	165,2	11036,5	2	39	36	-791,7	11211,9	148,1	12151,7
1	40	5	-1139,4	9731,8	132,6	11003,8	2	40	5	-1034,4	10969,2	118,6	12122,2
1	41	1	-1289,9	9581,3	97,9	10969,1	2	41	1	-1363,8	10639,8	82,4	12086,0
1	42	30	-1960,8	8910,4	48,9	10920,1	2	42	30	-1645,3	10358,3	41,3	12044,9
1	43	29	-2053,2	8818,0	0,0	10871,2	2	43	29	-1733,5	10270,1	0,0	12003,6

Tabela 8. Ordenamento dos genótipos avaliados nos locais 3e 4 onde são apresentados os efeitos genotípicos (g + ge), os valores genotípicos preditos para os respectivos locais (u+g+ge).

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
3	1	14	856,1	13373,0	856,1	13373,0	4	1	32	966,0	13810,5	966,0	13810,5
3	2	4	847,3	13364,2	851,7	13368,6	4	2	14	904,9	13749,4	935,4	13779,9
3	3	32	837,6	13354,5	847,0	13363,9	4	3	4	894,6	13739,1	921,8	13766,3
3	4	6	830,1	13347,0	842,7	13359,7	4	4	33	814,3	13658,8	895,0	13739,5
3	5	13	759,7	13276,6	826,1	13343,1	4	5	6	756,4	13600,9	867,3	13711,8
3	6	17	712,8	13229,7	807,2	13324,2	4	6	13	719,7	13564,2	842,7	13687,2
3	7	34	684,8	13201,7	789,8	13306,7	4	7	34	684,5	13529,0	820,1	13664,6
3	8	33	657,6	13174,5	773,2	13290,2	4	8	10	591,6	13436,2	791,5	13636,0
3	9	37	606,1	13123,0	754,7	13271,6	4	9	22	479,8	13324,3	756,9	13601,4
3	10	22	534,9	13051,8	732,7	13249,6	4	10	26	466,6	13311,1	727,8	13572,4
3	11	19	482,1	12999,0	709,9	13226,8	4	11	18	439,8	13284,3	701,7	13546,2
3	12	10	435,8	12952,7	687,1	13204,0	4	12	12	433,3	13277,9	679,3	13523,8
3	13	9	392,2	12909,2	664,4	13181,3	4	13	19	402,5	13247,0	658,0	13502,5
3	14	20	380,6	12897,5	644,1	13161,0	4	14	17	379,1	13223,6	638,1	13482,6
3	15	18	377,4	12894,3	626,3	13143,3	4	15	11	376,7	13221,2	620,7	13465,2
3	16	41	337,0	12853,9	608,2	13125,2	4	16	38	341,1	13185,6	603,2	13447,7
3	17	12	335,8	12852,7	592,2	13109,2	4	17	42	331,2	13175,7	587,2	13431,7
3	18	16	287,8	12804,7	575,3	13092,2	4	18	15	330,5	13175,0	572,9	13417,4
3	19	26	283,0	12799,9	559,9	13076,9	4	19	43	322,4	13166,9	559,7	13404,2
3	20	11	251,2	12768,1	544,5	13061,4	4	20	41	240,9	13085,4	543,8	13388,3
3	21	15	199,5	12716,4	528,1	13045,0	4	21	16	232,5	13077,0	529,0	13373,5
3	22	43	174,1	12691,0	512,0	13028,9	4	22	20	164,8	13009,3	512,4	13356,9

Tabela 8, continuação...

Local	Ord.	Genótipo	g+Ge	u+g+ge	Ganho	Nova Média	Local	Ord	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
3	23	21	128,7	12645,6	495,3	13012,2	4	23	9	14,5	12859,1	490,8	13335,3
3	24	40	88,7	12605,7	478,4	12995,3	4	24	37	-13,5	12831,0	469,8	13314,3
3	25	42	70,0	12587,0	462,0	12979,0	4	25	2	-65,3	12779,2	448,4	13292,9
3	26	38	12,5	12529,4	444,7	12961,7	4	26	21	-108,6	12735,9	426,9	13271,4
3	27	28	-14,2	12502,7	427,7	12944,7	4	27	23	-151,2	12693,3	405,5	13250,0
3	28	2	-29,3	12487,6	411,4	12928,4	4	28	28	-196,6	12647,9	384,0	13228,5
3	29	3	-215,0	12302,0	389,8	12906,8	4	29	3	-224,7	12619,8	363,0	13207,5
3	30	35	-231,3	12285,6	369,1	12886,0	4	30	8	-249,9	12594,6	342,6	13187,1
3	31	24	-235,1	12281,8	349,6	12866,6	4	31	35	-262,5	12582,0	323,1	13167,6
3	32	23	-332,2	12184,7	328,3	12845,2	4	32	24	-273,3	12571,2	304,4	13149,0
3	33	7	-433,3	12083,6	305,2	12822,2	4	33	25	-295,9	12548,7	286,3	13130,8
3	34	27	-443,7	12073,2	283,2	12800,1	4	34	7	-448,3	12396,2	264,7	13109,2
3	35	8	-465,7	12051,3	261,8	12778,7	4	35	39	-473,1	12371,4	243,6	13088,1
3	36	31	-472,7	12044,2	241,4	12758,3	4	36	40	-484,0	12360,5	223,4	13067,9
3	37	25	-479,7	12037,2	221,9	12738,9	4	37	27	-490,1	12354,4	204,1	13048,6
3	38	39	-823,6	11693,4	194,4	12711,3	4	38	31	-565,3	12279,2	183,8	13028,3
3	39	36	-1032,6	11484,3	162,9	12679,9	4	39	36	-1009,6	11834,9	153,2	12997,7
3	40	5	-1226,3	11290,6	128,2	12645,1	4	40	5	-1061,9	11782,6	122,9	12967,4
3	41	1	-1346,0	11170,9	92,3	12609,2	4	41	1	-1452,3	11392,2	84,4	12928,9
3	42	30	-1790,7	10726,2	47,4	12564,4	4	42	29	-1706,5	11138,0	41,8	12886,3
3	43	29	-1991,7	10525,3	0,0	12516,9	4	43	30	-1755,2	11089,3	0,0	12844,5

Tabela 9. Ordenamento dos genótipos avaliados nos locais 5 e 6, onde são apresentados os efeitos genotípicos (g +ge), os valores genotípicos preditos para os respectivos locais (u+g+ge).

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
5	1	14	852,0	12694,9	852,0	12694,9	6	1	13	792,5	11457,7	792,5	11457,7
5	2	13	803,6	12646,5	827,8	12670,7	6	2	10	755,0	11420,3	773,8	11439,0
5	3	32	767,6	12610,5	807,8	12650,6	6	3	34	726,1	11391,4	757,9	11423,1
5	4	37	713,3	12556,2	784,1	12627,0	6	4	14	711,3	11376,6	746,2	11411,5
5	5	34	680,8	12523,7	763,5	12606,3	6	5	11	610,1	11275,4	719,0	11384,3
5	6	22	617,5	12460,4	739,1	12582,0	6	6	32	574,0	11239,3	694,8	11360,1
5	7	6	526,0	12368,9	708,7	12551,6	6	7	6	569,5	11234,8	676,9	11342,2
5	8	4	523,9	12366,8	685,6	12528,5	6	8	17	559,4	11224,6	662,2	11327,5
5	9	33	514,0	12356,9	666,5	12509,4	6	9	37	558,1	11223,4	650,7	11315,9
5	10	10	423,7	12266,6	642,2	12485,1	6	10	33	550,0	11215,3	640,6	11305,9
5	11	17	408,9	12251,8	621,0	12463,9	6	11	4	531,7	11196,9	630,7	11296,0
5	12	41	396,9	12239,8	602,4	12445,2	6	12	22	522,3	11187,5	621,7	11286,9
5	13	19	363,5	12206,4	584,0	12426,9	6	13	18	473,4	11138,7	610,3	11275,5
5	14	9	363,3	12206,2	568,2	12411,1	6	14	16	389,3	11054,5	594,5	11259,7
5	15	18	359,8	12202,7	554,3	12397,2	6	15	20	344,2	11009,5	577,8	11243,1
5	16	16	343,6	12186,5	541,2	12384,0	6	16	26	336,6	11001,8	562,7	11228,0
5	17	12	317,2	12160,0	528,0	12370,9	6	17	19	327,0	10992,3	548,9	11214,1
5	18	20	292,0	12134,9	514,9	12357,8	6	18	38	324,3	10989,5	536,4	11201,6
5	19	11	288,4	12131,3	503,0	12345,8	6	19	43	277,4	10942,7	522,7	11188,0
5	20	38	234,5	12077,3	489,5	12332,4	6	20	42	269,7	10934,9	510,1	11175,4
5	21	15	221,6	12064,4	476,8	12319,6	6	21	12	259,1	10924,3	498,1	11163,4
5	22	26	205,7	12048,6	464,5	12307,3	6	22	41	204,5	10869,8	484,8	11150,1
5	23	43	170,4	12013,3	451,7	12294,5	6	23	9	203,4	10868,7	472,6	11137,8

Tabela 9, continuação...

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
5	24	42	159,5	12002,4	439,5	12282,4	6	24	40	195,8	10861,0	461,0	11126,3
5	25	40	108,9	11951,8	426,3	12269,1	6	25	15	194,2	10859,5	450,4	11115,6
5	26	21	79,6	11922,5	412,9	12255,8	6	26	21	41,7	10707,0	434,6	11099,9
5	27	2	30,7	11873,6	398,8	12241,7	6	27	2	-46,0	10619,3	416,8	11082,1
5	28	23	-135,6	11707,3	379,7	12222,6	6	28	28	-195,4	10469,9	395,0	11060,2
5	29	28	-146,4	11696,4	361,6	12204,4	6	29	23	-199,3	10465,9	374,5	11039,7
5	30	24	-173,4	11669,5	343,7	12186,6	6	30	35	-216,1	10449,2	354,8	11020,1
5	31	35	-232,2	11610,6	325,1	12168,0	6	31	24	-305,5	10359,8	333,5	10998,8
5	32	3	-282,6	11560,3	306,2	12149,0	6	32	8	-340,5	10324,7	312,4	10977,7
5	33	7	-391,3	11451,6	285,0	12127,9	6	33	31	-344,3	10320,9	292,5	10957,8
5	34	25	-396,8	11446,1	265,0	12107,8	6	34	7	-382,0	10283,3	272,7	10938,0
5	35	8	-405,2	11437,7	245,8	12088,7	6	35	27	-414,5	10250,8	253,1	10918,3
5	36	31	-442,5	11400,4	226,7	12069,6	6	36	25	-438,6	10226,6	233,8	10899,1
5	37	27	-448,7	11394,2	208,4	12051,3	6	37	3	-448,6	10216,7	215,4	10880,7
5	38	39	-710,8	11132,1	184,3	12027,1	6	38	39	-746,3	9919,0	190,1	10855,4
5	39	36	-835,4	11007,5	158,1	12001,0	6	39	1	-957,0	9708,2	160,7	10825,9
5	40	5	-1028,5	10814,3	128,4	11971,3	6	40	36	-1007,3	9658,0	131,5	10796,7
5	41	1	-1284,4	10558,5	94,0	11936,9	6	41	5	-1286,1	9379,1	96,9	10762,2
5	42	30	-1861,6	9981,3	47,4	11890,3	6	42	30	-1938,7	8726,6	48,4	10713,7
5	43	29	-1991,8	9851,1	0,0	11842,9	6	43	29	-2034,4	8630,9	0,0	10665,3

Tabela 10. Ordenamento dos genótipos avaliados nos locais 7 e 8, onde são apresentados os efeitos genotípicos (g +ge), os valores genotípicos preditos para os respectivos locais (u+g+ge).

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
7	1	32	865,7	11939,2	865,7	11939,2	8	1	14	834,3	11218,1	834,3	11218,1
7	2	4	805,3	11878,9	835,5	11909,1	8	2	37	646,7	11030,5	740,5	11124,3
7	3	14	766,7	11840,3	812,6	11886,2	8	3	6	635,6	11019,4	705,5	11089,3
7	4	22	765,7	11839,3	800,8	11874,4	8	4	13	620,1	11003,9	684,2	11068,0
7	5	34	746,4	11820,0	790,0	11863,5	8	5	34	596,0	10979,8	666,6	11050,4
7	6	33	664,4	11738,0	769,0	11842,6	8	6	4	571,4	10955,2	650,7	11034,5
7	7	13	650,5	11724,1	752,1	11825,7	8	7	22	551,4	10935,2	636,5	11020,3
7	8	19	558,0	11631,6	727,8	11801,4	8	8	32	541,5	10925,3	624,6	11008,4
7	9	26	533,1	11606,6	706,2	11779,8	8	9	33	440,2	10824,0	604,1	10987,9
7	10	6	520,7	11594,3	687,6	11761,2	8	10	9	433,3	10817,1	587,1	10970,9
7	11	10	511,2	11584,8	671,6	11745,2	8	11	20	411,8	10795,6	571,1	10954,9
7	12	37	501,8	11575,4	657,5	11731,0	8	12	17	402,5	10786,3	557,1	10940,9
7	13	18	478,3	11551,9	643,7	11717,3	8	13	10	398,2	10782,0	544,8	10928,6
7	14	16	418,8	11492,3	627,6	11701,2	8	14	18	378,2	10762,0	532,9	10916,7
7	15	12	340,4	11414,0	608,5	11682,0	8	15	19	338,8	10722,6	520,0	10903,8
7	16	17	325,4	11399,0	590,8	11664,4	8	16	41	328,2	10712,0	508,0	10891,8
7	17	42	310,6	11384,2	574,3	11647,9	8	17	11	304,4	10688,2	496,0	10879,8
7	18	41	295,7	11369,3	558,8	11632,4	8	18	16	294,0	10677,8	484,8	10868,6
7	19	43	281,5	11355,1	544,2	11617,8	8	19	38	262,1	10645,9	473,1	10856,9
7	20	11	273,4	11347,0	530,7	11604,3	8	20	26	247,0	10630,8	461,8	10845,6
7	21	20	273,3	11346,9	518,4	11592,0	8	21	12	231,4	10615,2	450,8	10834,6
7	22	15	206,9	11280,5	504,3	11577,9	8	22	15	211,8	10595,6	439,9	10823,7
7	23	9	163,4	11237,0	489,4	11563,0	8	23	21	196,7	10580,5	429,4	10813,2

Tabela 10, continuação...

Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média	Local	Ord.	Genótipo	g+ge	u+g+ge	Ganho	Nova Média
7	24	21	116,5	11190,1	473,9	11547,5	8	24	43	62,2	10446,1	414,1	10797,9
7	25	2	54,2	11127,8	457,1	11530,7	8	25	42	35,3	10419,1	398,9	10782,7
7	26	3	0,8	11074,4	439,6	11513,2	8	26	2	21,9	10405,7	384,4	10768,2
7	27	28	-112,1	10961,5	419,1	11492,7	8	27	28	-47,9	10335,9	368,4	10752,2
7	28	38	-122,4	10951,2	399,8	11473,4	8	28	40	-84,3	10299,5	352,2	10736,0
7	29	23	-161,9	10911,7	380,4	11454,0	8	29	3	-156,0	10227,8	334,7	10718,5
7	30	27	-220,6	10853,0	360,4	11434,0	8	30	24	-180,5	10203,3	317,5	10701,3
7	31	40	-229,5	10844,1	341,4	11415,0	8	31	23	-211,6	10172,2	300,5	10684,3
7	32	25	-230,8	10842,8	323,5	11397,1	8	32	7	-214,5	10169,3	284,4	10668,2
7	33	8	-338,6	10735,0	303,4	11377,0	8	33	35	-254,8	10129,0	268,0	10651,8
7	34	31	-401,9	10671,7	282,7	11356,3	8	34	8	-256,0	10127,8	252,6	10636,4
7	35	24	-545,6	10528,0	259,0	11332,6	8	35	25	-287,2	10096,6	237,2	10621,0
7	36	7	-573,3	10500,3	235,9	11309,5	8	36	27	-320,1	10063,7	221,7	10605,5
7	37	39	-580,0	10493,6	213,8	11287,4	8	37	31	-543,4	9840,4	201,0	10584,8
7	38	35	-634,2	10439,4	191,5	11265,1	8	38	39	-682,0	9701,8	177,8	10561,6
7	39	36	-840,2	10233,4	165,1	11238,7	8	39	36	-823,4	9560,4	152,1	10535,9
7	40	1	-1142,0	9931,6	132,4	11206,0	8	40	5	-949,5	9434,3	124,6	10508,4
7	41	5	-1401,8	9671,8	95,0	11168,6	8	41	1	-1109,1	9274,7	94,5	10478,3
7	42	30	-1879,0	9194,6	48,0	11121,6	8	42	30	-1839,4	8544,4	48,5	10432,3
7	43	29	-2014,9	9058,7	0,0	11073,6	8	43	29	-2035,2	8348,6	0,0	10383,8

Com base nestas informações foi possível verificar que, para o local 1, as dez melhores classificações foram obtidas pelos genótipos 4, 14, 32, 13, 33, 37, 34, 6, 17 e 10; os efeitos genotípicos foram de 849,0; 795,3; 780,7; 681,4; 674,6; 668,7; 655,5; 653,2; 599,9 e 596,0, respectivamente. Estes valores correspondem aos valores genotípicos capitalizados do efeito de cada genótipo em relação a cada local, isto é, sobre efeito da interação os valores genótipos preditos para cada local dos dez genótipos selecionados foram de 11720,3; 11666,6; 11652,0; 11552,6; 11545,8; 11539,9; 11526,8; 11524,5; 11471,1 e 11467,3, respectivamente, (Tabela 7).

No local 2, os dez genótipos mais bem selecionados foram: 14, 13, 32, 10, 37, 34, 33, 19, 6 e 11. Estes obtiveram efeitos genotípicos de 1041,6; 849,2; 665,2; 632,5; 626,1; 618,7; 595,8; 540,6; 482,0 e 467,2, respectivamente. Isto é, o efeito do valor genotípico mais o efeito do ambiente. Os valores genotípicos preditos foram, respectivamente, 13045,2; 12852,8; 12668,8; 12636,1; 12629,7; 2622,3; 12599,4; 12544,2; 12485,6 e 12470,8 (Tabela 7).

Para o local 3, os melhores genótipos foram 14, 4, 32, 6, 13, 17, 34, 33, 37 e 22. Estes apresentaram valores genotípicos: 856,1; 847,3; 837,6; 830,1; 759,7; 712,8; 684,8; 657,6; 606,1 e 534,9, respectivamente. Os valores genotípicos preditos foram: 13373,0; 13364,2; 13354,5; 13347,0; 13276,6; 13229,7; 13201,7; 13174,5; 13123,0 e 13051,8, respectivamente (Tabela 8).

No local 4, os genótipos 32, 14, 4, 33, 6, 13, 34, 10, 22 e 26 foram os que apresentaram melhor classificação, com valores genotípicos preditos de 13810,5; 13749,4; 13739,1; 13658,8; 13600,9; 13564,2; 13529,0 e 13436,2, respectivamente. Os valores genotípicos no ambiente foram 966,0; 904,9; 894,6; 814,3; 756,4; 719,7; 684,5 e 591,6 (Tabela 8).

De acordo com os valores genotípicos, os genótipos mais bem classificados no ambiente 5 foram: 14, 13, 32, 37, 34, 22, 6, 4, 33 e 10, com os seguintes valores genotípicos: 852,0; 803,6; 767,6; 713,3; 680,8; 617,5; 526,0; 523,9; 514,0 e 423,7, respectivamente. Os valores genotípicos preditos foram 12694,9; 12646,5; 12610,5; 12556,2; 12523,7; 12460,4; 12368,9; 12366,8; 12356,9 e 12266,6, respectivamente (Tabela 9).

No local 6, os dez genótipos mais bem classificados foram 13, 10, 34, 14, 11, 32, 6, 17, 37 e 33, com os valores genotípicos correspondentes de 11457,7; 1420,3; 1391,4; 1376,6; 11275,4; 11239,3; 11234,8; 11224,6; 11223,4 e 11215,3. Os efeitos genotípicos observados foram 792,5; 755,0; 726,1; 711,3; 610,1; 574,0; 569,5; 559,4; 558,1 e 550,0, respectivamente (Tabela 9).

Para o local 7, os genótipos que apresentaram melhor ranqueamento foram 32, 4, 14, 22, 34, 33, 13, 19, 26 e 6, com efeitos genotípicos na seguinte ordem: 865,7; 805,3; 766,7; 765,7; 746,4; 664,4; 650,5; 558,0; 533,1 e 520,7. Os valores genotípicos preditos foram, respectivamente, 11939,2; 11878,9; 11840,3; 11839,3; 11820,0; 11738,0; 11724,1; 11631,6; 11606,6 e 11594,3 (Tabela 10).

No local 8, o ranqueamento se deu na seguinte ordem: 14, 37, 6, 13, 34, 4, 22, 32, 33 e 9. Estes apresentaram efeitos genotípicos de 834,3; 646,7; 635,6; 620,1; 596,0; 571,4; 551,4; 541,5; 440,2 e 433,3. Os valores genotípicos preditos para estes genótipos foram 11218,1; 11030,5; 11019,4; 11003,9; 10979,8; 10955,2; 10935,2; 10925,3; 10824,0 e 10817,1, respectivamente (Tabela 10).

Considerando todos os locais avaliados, o genótipo que mais apareceu em primeiro lugar no ranqueamento foi o genótipo 14, seguido pelo genótipo 32, 4, e 13.

O genótipo 14 apresentou ampla adaptabilidade, ou seja, teve destaque em todos os locais avaliados. Outros genótipos que também apareceram entre os dez melhores genótipos para os oito locais avaliados foram os genótipos 32, 13, 33, 34 e 6, seguidos dos genótipos 4 e 37 que se destacaram em seis locais e os genótipos 10 e 22 que apareceram em cinco locais. De acordo com Maia et al., (2009), esses genótipos que aparecem classificados com os mais produtivos em todos os ambientes não sofrem influência significativa do ambiente, conseqüentemente, contribuem pouco para a interação genótipo x ambiente.

3.3 Análise de estabilidade dos genótipos avaliados por meio da Média Harmônica dos Valores Genotípicos

A Média Harmônica dos Valores Genotípicos (MHVG) consolida, em uma única informação, rendimento de grãos e estabilidade. Como característica, a MHVG penaliza aqueles genótipos que possuem maior instabilidade em diversos locais (RESENDE, 2002). Com base nestas informações os genótipos que mais associaram produtividade com estabilidade foram os híbridos 14, 32, 13, 4, 34, 6, 33, 22, 37 e 10, como observado na Tabela 11. Os valores de MHVG foram, respectivamente, 12304,8; 12202,4; 12196,4; 12142,2; 12140,4; 12081,5; 12070,6; 12022,8; 12016,3 e 12009,7 em ordem decrescente.

Os genótipos que mais apareceram entre os dez primeiros colocados para a análise dos valores genotípicos foram também os que apresentaram maior estabilidade associada à produtividade. Estes resultados estão em concordância com o encontrado por Bastos et al., (2007) com clones de cana-de-açúcar. Esses autores afirmaram que o método MHVG é uma

opção segura para a análise de estabilidade e produtividade. No entanto, Zeni Neto et al., (2008) demonstraram que este método não pode ser considerado como seguro quando se analisam experimentos desbalanceados, ou seja, quando todos os genótipos não estão presentes em todos os ambientes avaliados.

3.4 Análise de adaptabilidade por meio de performance relativa dos valores genotípicos.

Para o estudo da adaptabilidade, utilizou-se a performance relativa dos valores genotípicos (PRVG). A característica de adaptabilidade é de grande valor e procurada pelos melhoristas vegetais para a avaliação e obtenção de novos cultivares (BORGES et al., 2010). De acordo com Resende (2002), tal metodologia contabiliza o efeito de resposta de cada genótipo ao estímulo dos ambientes em que foram avaliados, seja ele um ambiente bom ou ruim.

De acordo com os valores de PRVG, os dez genótipos mais adaptados foram: 14, 32, 13, 4, 34, 6, 33, 22, 37 e 10. Estes genótipos apresentaram as seguintes performances: 1,073; 1,065; 1,064; 1,059; 1,059; 1,054; 1,053; 1,048; 1,048 e 1,047 (Tabela 11).

Por meio dos valores de performances relativas, foi possível obter a média de cada genótipo, capitalizado pela interação. Para isso, foi necessário multiplicar o valor de PRVG pela média geral ($PRVG * Média Geral$) (Tabela 11).

Como observado na Tabela 11, o mesmo ordenamento dos dez genótipos mais produtivos para o critério de estabilidade (MHVG) foi observado para adaptabilidade (PRVG).

Em termos de adaptabilidade, a performance relativa dos valores genéticos nos ambientes é uma medida simples e efetiva no contexto dos modelos mistos, pois fornece os resultados em termos de componentes de médias e preserva a escala de avaliação do caráter. Neste caso, os valores genotípicos preditos (dados originais) são expressos como uma proporção da média geral em cada local e, em seguida, o valor médio dessa proporção nos ambientes é obtido (ZENI NETO et al., 2008; VERARDI et al., 2009).

3.5 Análise de adaptabilidade e estabilidade por meio da Média Harmônica da Performance Relativa dos Valores Genotípicos (MHPRVG).

A análise da adaptabilidade e estabilidade por meio da performance relativa dos valores genotípicos preditos tem como principal característica a avaliação dos genótipos

utilizando de uma única informação, na qual agrupa valores de adaptabilidade e estabilidade para rendimento de grãos (RESENDE, 2004). Tal procedimento ainda tem como respaldo a abordagem dos genótipos tidos como aleatórios, utilizando modelos mistos, via estimador de máxima verossimilhança restrita REML e preditor Blup (melhor preditor linear não-tendencioso).

Com base nestas informações são apresentados na Tabela 11 os valores de MHPRVG, referentes aos genótipos. Os genótipos mais estáveis, mais adaptados e com maior produtividade foram: 14, 32, 13, 4, 34, 6, 33, 22, 10 e 37. Estes apresentaram valores de MHPRVG que foram, respectivamente: 1,073; 1,065; 1,064; 1,059; 1,059; 1,054; 1,053; 1,048; 1,047 e 1,047.

Por meio dos valores de MHPRVG, foi possível obter o valor médio genotípico de cada genótipo no ambiente avaliado, incluindo a penalização pela estabilidade e capitalizando a adaptabilidade (CORBONELL et al., 2007). Para isso foi necessário multiplicar o valor de MHPRVG pela média geral (MHPRVG * MG) (Tabela 11).

A seleção simultânea da variável produtividade e os parâmetros de estabilidade e adaptabilidade apresentaram os mesmos genótipos entre os dez melhores, observados para a seleção individual de estabilidade e adaptabilidade (Tabela 11). O ordenamento da classificação dos dez melhores genótipos avaliados foi 100% coincidente com os parâmetros de MHVG e PRVG e diferiram apenas quando alternando os genótipos entre a 9ª e 10ª posição para MHPRVG.

O genótipo 14 foi superior aos demais pela MHPRVG para estabilidade, adaptabilidade e produtividade, confirmando os resultados observados para MHVG e PRVG. Vale ressaltar que os dez melhores genótipos ordenados pela MHVG, PRVG e MHPRVG foram coincidentes com os dez genótipos que mais apareceram na análise dos valores genotípicos (Tabela 6), confirmando que os genótipos mais adaptados e estáveis também são os mais produtivos. Este resultado está em concordância com Scapim et al. (2000), os quais afirmam que a maior estabilidade está associada à maior produtividade.

Tabela 11. Ordenamento dos genótipos avaliados nos locais 1 e 2, nas quais são apresentados a estabilidade dos valores genotípicos (MHVG), adaptabilidade dos valores genotípicos (PRVG), valores genotípicos médios capitalizado pela interação (PRVG*MG), estabilidade e adaptabilidade de valores genotípicos (MHPRVG) e valores genotípicos médios nos locais (MHPRVG*MG) para o caráter rendimento de grãos de híbridos de milho avaliados no período de 2009 e 2010.

Estabilidade de Valores Genéticos (MHVG)			Adaptabilidade de Valores Genéticos (PRVG)				Estabilidade e Adaptabilidade de Valores Genéticos (MHPRVG)			
Ord.	Gen.	MHVG	Ord.	Gen.	PRVG	PRVG*MG	Ord.	Gen.	MHPRVG	MHPRVG*MG
1	14	12304,8	1	14	1,073	12370,9	1	14	1,073	12370,4
2	32	12202,4	2	32	1,065	12271,9	2	32	1,065	12271,0
3	13	12196,4	3	13	1,064	12261,1	3	13	1,064	12260,7
4	4	12142,2	4	4	1,059	12209,3	4	4	1,059	12207,2
5	34	12140,4	5	34	1,059	12202,6	5	34	1,059	12202,3
6	6	12081,5	6	6	1,054	12146,9	6	6	1,054	12146,0
7	33	12070,6	7	33	1,053	12137,5	7	33	1,053	12136,9
8	22	12022,8	8	22	1,048	12083,6	8	22	1,048	12082,6
9	37	12016,3	9	37	1,048	12074,4	9	10	1,047	12070,5
10	10	12009,7	10	10	1,047	12071,9	10	37	1,047	12070,1
11	17	11942,9	11	17	1,042	12006,3	11	17	1,042	12005,1
12	19	11891,8	12	19	1,037	11956,1	12	19	1,037	11955,6
13	18	11879,0	13	18	1,036	11939,9	13	18	1,036	11939,6
14	11	11835,7	14	11	1,032	11896,6	14	11	1,032	11895,3
15	16	11800,1	15	16	1,029	11859,7	15	16	1,029	11859,3
16	26	11781,4	16	26	1,028	11844,9	16	26	1,028	11843,8
17	20	11773,1	17	12	1,027	11834,4	17	12	1,027	11834,2
18	12	11769,9	18	20	1,027	11832,9	18	20	1,027	11832,3
19	41	11754,9	19	41	1,025	11815,8	19	41	1,025	11814,8
20	9	11753,7	20	9	1,025	11812,9	20	9	1,025	11811,0

Tabela 11, continuação...

Ord.	Gen.	MHVG	Ord.	Gen.	PRVG	PRVG*MG	Ord.	Gen.	MHPRVG	MHPRVG*MG
21	15	11687,4	21	15	1,020	11751,0	21	15	1,020	11750,9
22	38	11661,1	22	38	1,017	11724,1	22	38	1,017	11722,0
23	42	11648,1	23	42	1,016	11711,7	23	42	1,016	11710,9
24	43	11645,1	24	43	1,016	11711,2	24	43	1,016	11710,2
25	21	11549,5	25	21	1,007	11607,8	25	21	1,007	11607,2
26	2	11470,1	26	2	1,001	11530,9	26	2	1,001	11530,8
27	40	11417,4	27	40	0,996	11476,6	27	40	0,996	11473,1
28	28	11333,7	28	28	0,989	11395,1	28	28	0,989	11394,8
29	23	11250,5	29	23	0,982	11312,4	29	23	0,982	11312,0
30	3	11206,8	30	3	0,978	11269,1	30	3	0,978	11267,3
31	24	11193,4	31	24	0,977	11257,9	31	24	0,977	11256,5
32	35	11176,1	32	35	0,975	11240,0	32	35	0,975	11238,0
33	8	11096,8	33	8	0,968	11158,1	33	8	0,968	11157,5
34	27	11090,2	34	27	0,967	11149,1	34	27	0,967	11148,6
35	25	11068,3	35	25	0,966	11130,6	35	25	0,966	11129,4
36	7	11034,5	36	7	0,963	11092,6	36	7	0,962	11091,4
37	31	10988,6	37	31	0,959	11051,6	37	31	0,959	11050,5
38	39	10789,9	38	39	0,942	10853,5	38	39	0,942	10852,3
39	36	10570,5	39	36	0,922	10630,1	39	36	0,922	10629,3
40	5	10311,4	40	5	0,900	10376,9	40	5	0,900	10373,8
41	1	10231,8	41	1	0,892	10284,8	41	1	0,892	10284,0
42	30	9606,7	42	30	0,840	9676,9	42	30	0,839	9672,4
43	29	9488,6	43	29	0,830	9562,9	43	29	0,829	9556,4

3.6 Acurácia na seleção de genótipos

A acurácia seletiva $r_{\hat{a}a}$ reflete a precisão experimental e leva em conta a variação genética e residual associada aos genótipos. Tal procedimento se trata da correlação entre o valor genotípico verdadeiro e aqueles obtidos com base nos experimentos de acurácia expressa em porcentagem. Isso tem variação de 0 a 1 e possui as seguintes classes: muito alta, maior ou igual a 0,9; alta referente ao intervalo $0,7 \leq r_{\hat{a}a} < 0,9$; moderada, $0,5 \leq r_{\hat{a}a} < 0,7$, e classe baixa, com acurácia menor que 0,5. Logo, os valores de acurácia e os intervalos de confiança (LIIC e LSIC) dos valores genéticos preditos estão representados na Tabela 12. Os valores da acurácia de seleção foram de mesma ordem 0,89 para todos os genótipos avaliados, enquadrando-os na classe alta de acurácia, a qual reflete uma situação favorável para a seleção. De acordo com Resende (2007), valores acima de 0,70 são recomendados para programas de melhoramento na prática da seleção.

Desta forma, os genótipos que apresentaram boa performance em estabilidade, adaptabilidade e produtividade, também apresentaram alta acurácia de seleção.

Tabela 12. Valores genéticos e limites para características rendimento de grãos.

Ordem	Genótipo	VG	Acurácia	LIIC	LSIC	Média f^*
1	14	12353,3	0,897	11736,5	12970,1	12531,3
2	32	12259,8	0,897	11643,0	12876,6	12417,6
3	13	12244,9	0,897	11628,1	12861,7	12399,5
4	4	12195,6	0,897	11578,8	12812,4	12339,6
5	34	12185,7	0,897	11568,9	12802,5	12327,6
6	6	12134,3	0,897	11517,5	12751,1	12265,2
7	33	12126,6	0,897	11509,8	12743,4	12255,9
8	22	12067,6	0,897	11450,7	12684,4	12184,1
9	10	12057,2	0,897	11440,4	12674,0	12171,5
10	37	12052,7	0,897	11435,9	12669,5	12166,0
11	17	11994,8	0,897	11378,0	12611,6	12095,7
12	19	11947,2	0,897	11330,3	12564,0	12037,8
13	18	11928,4	0,897	11311,6	12545,2	12015,0
14	11	11885,3	0,897	11268,5	12502,1	11962,6
15	16	11848,8	0,897	11232,0	12465,6	11918,4
16	26	11837,7	0,897	11220,9	12454,5	11904,9

Tabela 12, continuação...

17	12	11829,3	0,897	11212,5	12446,1	11894,7
18	20	11822,8	0,897	11206,0	12439,6	11886,8
19	41	11806,9	0,897	11190,1	12423,7	11867,5
20	9	11801,3	0,897	11184,5	12418,1	11860,6
21	15	11747,4	0,897	11130,6	12364,2	11795,1
22	38	11718,7	0,897	11101,9	12335,5	11760,2
23	43	11710,1	0,897	11093,3	12326,9	11749,8
24	42	11708,2	0,897	11091,4	12325,0	11747,5
25	21	11602,0	0,897	10985,1	12218,8	11618,4
26	2	11529,9	0,897	10913,1	12146,7	11530,9
27	40	11472,0	0,897	10855,2	12088,8	11460,6
28	28	11398,2	0,897	10781,4	12015,0	11370,9
29	23	11317,9	0,897	10701,1	11934,7	11273,4
30	3	11274,7	0,897	10657,9	11891,5	11220,9
31	24	11266,5	0,897	10649,7	11883,3	11210,9
32	35	11247,9	0,897	10631,1	11864,7	11188,3
33	8	11167,0	0,897	10550,2	11783,8	11090,0
34	27	11155,6	0,897	10538,8	11772,4	11076,2
35	25	11140,5	0,897	10523,7	11757,3	11057,9
36	7	11099,5	0,897	10482,7	11716,3	11008,0
37	31	11063,9	0,897	10447,1	11680,7	10964,8
38	39	10871,8	0,897	10255,0	11488,6	10731,4
39	36	10650,1	0,897	10033,3	11266,9	10462,0
40	5	10407,4	0,897	9790,6	11024,2	10167,2
41	1	10307,4	0,897	9690,6	10924,2	10045,7
42	30	9728,6	0,897	9111,8	10345,4	9342,6
43	29	9619,6	0,897	9002,8	10236,4	9210,1

*f: Média fenotípica.

4 CONCLUSÕES

A utilização dos métodos de estudos de estabilidade MHVG, método de estudo de adaptabilidade PRVG e as duas sob o enfoque de adaptabilidade e estabilidade para rendimento de grãos, por meio da MHPRVG, evidenciou ser útil à seleção dos genótipos avaliados.

A acurácia apresentou valores considerados altos, favorecendo a seleção dos genótipos.

Os valores genotípicos e os valores de MHPRVG classificaram os mesmos híbridos em completa concordância.

Com base em valores genotípicos e de MHPRVG, os genótipos que mais se destacaram foram: 14, 32, 13, 4, 34, 6, 33, 22, 10 e 37.

5 REFERÊNCIAS

- BASTOS, I. T.; BARBOSA, M. H. P.; RESENDE, M. D. V.; PETERNELLI, L. A.; SILVEIRA, L. C. I.; DONDA, L. R.; FORTUNATO, A. A.; COSTA, P. M. A.; FIGUEIREDO, I. C. R. Avaliação da interação genótipo x ambiente em cana-de-açúcar via modelos mistos. **Pesquisa Agropecuária Tropical**, Goiânia, v.37, p.195-203, 2007.
- BORGES, V. **Avaliação genotípica de linhagens de arroz de terras altas via metodologia de modelos mistos**. Tese de Doutorado. Lavras: Universidade Federal de Lavras. 2009. 164p.
- BORGES, V.; FERREIRA, V. P.; SOARES, L.; SANTOS, G. M.; SANTOS, A. M. M. Seleção de clones de batata-doce pelo procedimento REML/BLUP. **Acta Scientiarum Agronomy**. Maringá, v. 32, n.4, p.643-649, 2010.
- CARBONELL, S. A. M.; CHIORATO, A. F.; RESENDE, M. D. V. ; DIAS, A. S.; BERALDO, A. L. A.; PERINA, E. F. Estabilidade de cultivares e linhagens de feijoeiro em diferentes ambientes no Estado de São Paulo. **Bragantia**, Campinas, v. 66, n. 2, p. 193-201, 2007.
- CRUZ, C. D.; CASTOLDI, F. L. Decomposição da interação genótipos x ambientes em partes simples e complexa. **Revista Ceres**, Viçosa, v. 38, n. 219, p. 422-430, 1991.
- DUARTE, J. B.; VENCOSKY, R. Interação genótipo x ambiente uma introdução à análise "AMMI". Ribeirão Preto: **Sociedade Brasileira de Genética**, 1999. 60p.
- HENDERSON, C. R. Best Linear Unbiased estimation and Prediction under a selection model. **Biometrics**, CIDADE., v. 31, n.2, p. 423-447, 1975.
- HENDERSON, C. R. Estimates of changes in herd environment. **Journal of Dairy Science**, Champaign, v. 32, p.706, 1949.
- HENDERSON, C. R. **Applications of linear models in animal breeding**. Guelph : University of Guelph, 1984. 462p.
- HENDERSON, C. R. General flexibility of linear models for sire evolution. **Journal of Dairy Sciences**. Champaign, v. 57, n. 8, p. 963-972, 1974.
- HENDERSON, C. R. The use of equivalent linear models in mixed model estimation and prediction. **Communications in Statistics - Theory and Methods**, Philadelphia, vol. 17, n. 4, p. 969-982, 1988.
- MAIA, M. C. C.; RESENDE, M. D. V.; PAIVA, J. R.; CAVALCANTI, J. J. V.; BARROS, L. M. Seleção simultânea para produção, adaptabilidade e estabilidade genotípicas em clones de cajueiro, via modelos mistos. **Pesquisa Agropecuária Tropical**. Goiânia, v. 39, p.43-50, 2009.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Embrapa: Informação Tecnológica, Brasília, 2002. 975p.

RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo**. Colombo: Embrapa Florestas, 2004. 65p.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Embrapa Florestas, Colombo, 2007. 561p.

RESENDE, M. D. V. **SELEGEN–REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Colombo: Embrapa Florestas, 2007a. 361p.

RESENDE, M. D. V. **Software Selegen–REML/BLUP**. Colombo: Embrapa Florestas, Documentos 77. 2002b. 67p.

RESENDE, M. D. V.; RESENDE, R. M. S.; VALE, C. B.; JANK, L. Experimentação e análises estatísticas no melhoramento de forrageiras. In: RESENDE, R. M. S.; VALE, C. B.; JANK, L. **Melhoramento de Forrageiras Tropicais**. Campo Grande: Embrapa Gado de Corte. 2008. p. 195-293.

SEARLE, S. R., CASELLA, G.; MCCULLOCH, C. E. **Variance components**. New York : John Wiley & Sons. 1992. 501p.

SCAPIM, C. A.; OLIVEIRA, V. R.; BRACCINI, A. L. DE.; CRUZ, C. D.; ANDRADE, C. A. B.; VIDIGAL, M. C. G. Yield stability in maize (*Zea mays* L.) and correlations among the parameters of the Eberhart and Russell, Lin and Binns and Huehn models. **Genetics and Molecular Biology**, Ribeirão Preto, v. 23, n. 2, p. 387-393, 2000.

VERARDI, C. K.; RESENDE, M. D. V.; COSTA, R. B.; GONÇALVES, P. S. Adaptabilidade e estabilidade de borracha e seleção em progênies de seringueira. **Pesquisa Agropecuária Brasileira**, Brasília. v. 44, n. 10, p. 1277-1282, 2009.

ZENI-NETO, H.; OLIVEIRA, R. A.; DAROS, E.; BESPALHOK FILHO, J. C.; ZAMBON, J. L. C.; IDO, O. T.; WEBER, H. Seleção para produtividade, estabilidade e adaptabilidade de clones de cana-de-açúcar em três ambientes no Estado do Paraná via modelos mistos. **Scientia Agraria**, Curitiba, v. 9, p. 425-430, 2008.

CAPÍTULO 4

AVALIAÇÃO GENÉTICA DE FAMÍLIAS DE MEIOS-IRMÃOS DE MILHO PIPOCA USANDO A METODOLOGIA BAYESIANA

RESUMO

Populações de milho pipoca de polinização aberta são de grande importância para pequenos agricultores que utilizam material melhorado útil para vários anos agrícolas. O presente trabalho teve por objetivo estimar parâmetros genéticos da capacidade de expansão e rendimento de milho pipoca em um ensaio de 169 famílias de meios-irmãos, utilizando a metodologia Bayesiana como preditor do mérito genético. O método Bayesiano foi implementado por meio do algoritmo de Cadeias de Independência (*Independence Chain*), com informações *a priori* informativas para os componentes de variância familiar e residual (distribuição *a priori* Inversa Gamma). A capacidade de expansão mostrou ser moderadamente herdável com moda *a posteriori* de $h^2=0,34$ e intervalo de credibilidade de 0,22 – 0,44 (90% de probabilidade). O rendimento de grãos apresentou uma herdabilidade (nível de família) de $h^2=0,4$, com intervalo de credibilidade de 0,28 – 0,49. De acordo com as informações obtidas com a população em estudo, é possível inferir que existe variabilidade genética significativa para os ciclos subsequentes de melhoramento e que a metodologia Bayesiana pode auxiliar no processo de seleção de famílias de milho pipoca.

Palavras-chave: *Zea mays* L., capacidade de expansão, rendimento, algoritmo de Cadeias de Independência.

GENETIC EVALUATION OF HALF-SIB FAMILIES OF POPCORN USING THE BAYESIAN METHODOLOGY

ABSTRACT

Open-pollinated populations of popcorn are of great importance for those small farmers who use improved material for several crop years. This study aimed to estimate genetic parameters of popping expansion and yield in a trial with 169 half-sib families, using the Bayesian methodology as a predictor of genetic merit. The Bayesian method was implemented by means of the Independence Chain algorithm with informative prior information about variance components (Inverse Gamma prior distribution). Popping expansion was found to be moderately heritable with posterior mode of $h^2 = 0.34$ and credible interval of 0.22 - 0.44 (90% of probability). Grain yield had a heritability (family level) of $h^2 = 0.4$ with credible interval of 0.28 - 0.49. According to information obtained from the population under study it is possible to infer that there is significant genetic variability for subsequent cycles of popcorn breeding, and that the Bayesian methodology can assist in the selection process of popcorn families.

Keywords: Zea mays L., Popping expansion, Yield, Independence Chain Algorithm.

1 INTRODUÇÃO

No Brasil, existe uma crescente demanda anual de milho pipoca, estimando-se um consumo aproximado de $70.000 \text{ t}\cdot\text{ano}^{-1}$, das quais 71% são importados, principalmente da Argentina e Estados Unidos (SCAPIM et al., 2006). O número reduzido de cultivares é fator limitante na expansão da cultura no Brasil. A cultura do milho pipoca representa uma alternativa para produção de pequenos agricultores, devido ao elevado valor obtido por unidade de produção, além do consumo ser bastante disseminado no Brasil (RODOVALHO et al. 2008).

No melhoramento do milho pipoca devem-se levar em consideração, além da produtividade e caracteres agronômicos, os aspectos relacionados à qualidade da pipoca, como textura e maciez. Ao agricultor interessa produtividade elevada e os demais atributos de boa população de milho normal; ao consumidor, alta capacidade de expansão (CE), que confere à pipoca melhor textura e maciez (SCAPIM et al. 2002).

Populações de milho pipoca de polinização aberta são de grande importância para pequenos agricultores, que podem utilizá-las no decorrer de vários anos agrícolas. Os trabalhos de seleção de famílias têm como objetivo fundamental melhorar características de interesse com o aumento da frequência de alelos favoráveis nos ciclos de melhoramento. As variedades de polinização aberta são por definição populações com intercruzamento de indivíduos entre si (PACHECO et al., 2002).

A idéia central de um processo de seleção é aumentar a frequência dos alelos favoráveis para a característica de interesse. Neste contexto, algumas características que estão sobre processo de seleção são complexas, isto é, são características quantitativas. A principal vantagem deste tipo de população é a estabilidade de produção associada à possibilidade de manutenção das sementes para outros cultivos subsequentes, tornando-se uma opção viável para pequenos agricultores (BISOGNIN et al., 1997).

Projetos visando ao melhoramento intrapopulacional são de grande importância no melhoramento genético de milho (MATTA & VIANNA, 2003). Tal procedimento tem como objetivo o incremento na frequência de alelos favoráveis, ou desfavoráveis, para determinada característica, mantendo a variabilidade genética da população melhorada (DOERKSEN, et al., 2003).

O método de seleção entre e dentro de família de meios-irmãos é um dos métodos mais difundidos em programas de melhoramento no Brasil (VIANA, 2007). Os métodos de

melhoramento intrapopulacionais se mostram eficientes no que diz respeito ao ganho de melhoramento na população *per se* (SOUZA JR, 1993) elevando a frequência de alelos favoráveis para caracteres desejáveis (PACHECO, et al., 1998; VIANA, 2007).

O principal objetivo de todo processo de melhoramento intrapopulacional é obter o maior ganho com a nova população selecionada, porém a seleção não criteriosa pode levar a gerações com altos níveis de endogamia, conduzindo a uma redução no vigor populacional (PATERNIANI & VIEGA, 1987). Os métodos de seleção de meios-irmãos fornecem estimativas de parâmetros genéticos, ganhos de seleção, diferenciais de seleção, dentro outros (HALLAUER & MIRANDA FILHO, 1981; SCAPIM, et al., 2006).

No melhoramento genético do milho, a estimação dos componentes de variância e a predição dos valores genéticos são aspectos chaves na avaliação genética de características de importância econômica (SCAPIM et al. 2006), os quais podem ser abordados por método de métodos Bayesianos, incorporando as diferentes variantes dos procedimentos MCMC (Markov Chain Monte Carlo).

Tais procedimentos são utilizados para determinar as distribuições marginais *a posteriori* das estimativas dos parâmetros de interesse. A metodologia Bayesiana é considerada ferramenta importante na avaliação genética, uma vez que leva em conta a incerteza existente sobre todos os parâmetros do modelo (incluindo componentes de variância) (NOGUEIRA et al., 2003).

No contexto da inferência Bayesiana, todos os parâmetros do modelo, sejam fixos (incluindo os componentes de variância) ou aleatórios, são considerados variáveis aleatórias, conforme o conceito subjetivo de probabilidade, pelo qual se quantifica a incerteza existente sobre eles. Utilizando uma generalização do teorema de Bayes, informações *a priori* sobre os parâmetros são utilizadas em associação com os dados amostrais (representados no teorema pela função de verossimilhança), possibilitando uma inferência *a posteriori* sobre aqueles (NOGUEIRA et al., 2003).

O presente trabalho teve por objetivo estimar parâmetros genéticos da capacidade de expansão e rendimento de milho pipoca em um ensaio de 169 famílias de meios-irmãos, utilizando a metodologia Bayesiana como preditor do mérito genético.

2 MATERIAL E MÉTODOS

O material genético utilizado no presente trabalho consistiu de uma população de milho pipoca de sementes brancas UEM-C1. Foram avaliadas 169 famílias de meio-irmãos no delineamento em Lattice 13x13, com três repetições. O experimento foi conduzido na Fazenda Experimental de Iguatemi, pertencente à Universidade Estadual de Maringá, no ano agrícola 2008-2009.

Os caracteres avaliados foram rendimento de grãos em (kg ha⁻¹) e capacidade de expansão (mL g⁻¹). O caráter rendimento de grãos foi corrigido para 13% de umidade e os procedimentos para expansão foram feitos de acordo com Pacheco et al. (1998).

Do ponto de vista clássico de inferência científica, os dados do ensaio estão representados no seguinte modelo linear misto (BUENO FILHO & VENCOVSKY, 2000):

$$y = X\beta + Z\mu + \varepsilon$$

em que:

y é o vetor das observações fenotípicas (capacidade de expansão e rendimento);

X corresponde à matriz de incidência dos coeficientes dos parâmetros de efeito fixo;

β é o vetor de parâmetros de repetições e bloco dentro de repetições;

Z é a matriz de incidência para os efeitos da família;

u corresponde ao vetor dos efeitos familiares e ε o vetor dos resíduos.

O modelo anterior foi sugerido por Bueno Filho & Vencovsky (2000) para análise do experimento em látice no melhoramento, usando modelos mistos; Predição de valores aleatórios de tratamentos; efeitos aleatórios de tratamento e efeitos fixos de repetições e de blocos dentro de repetições (nomeado pelos autores como Modelo BLUP).

Os critérios de informação de Akaike (AIC: Akaike Information Criterion) e Bayesiano de Schwarz (BIC: Bayesian Information Criterion) foram usados para observar a significância do efeito familiar. Adicionalmente, foi utilizado o teste da razão de verossimilhança (LRT: Likelihood Ratio test) para comparar o modelo nulo (sem efeito familiar) e completo; LRT tem distribuição χ^2 , o qual foi usado como critério de significância ($p=0.95$) (RESENDE, 2007). AIC, BIC, LRT e os componentes de variância REML (*Restricted Maximum Likelihood*) foram obtidos usando o procedimento PROC MIXED do SAS (SAS-INSTITUTE, 1992).

Do ponto de vista do modelo bayesiano, os componentes de variância, herdabilidade e predição dos valores genéticos foram obtidos usando o procedimento PROC MIXED, com a opção PRIOR e amostras *a posteriori* de tamanho 50.000, gerada pelo algoritmo de Cadeias de Independência (*Independence Chain*); variante dos métodos de Monte Carlo via Cadeias de Markov (MCMC). A função de densidade conjunta de (β, μ, θ) , sendo θ os parâmetros (componentes de variância) a serem estimados, é dada por:

$$f(\beta, \mu, \theta | y) = f(\beta, \mu | \theta, y) f(\theta | y)$$

Cada distribuição *a priori* dos elementos do modelo Bayesiano é dado por:

$$f(y | \beta, \mu, \sigma_e^2, \sigma_f^2) \sim N(X\beta + Z\mu, R)$$

$$f(\beta) \propto 1 \text{ (Flat prior)}$$

$$f(\mu | \sigma_f^2) \sim N(0, G) \text{ (Normal)}$$

$$f(\sigma_f^2) \sim IG(a, b) \text{ (Inversa Gama)}$$

$$f(\sigma_e^2) \sim IG(c, d) \text{ (Inversa Gama)}$$

A distribuição *a posteriori* conjunta é dada por:

$$f(\beta, \mu, \sigma_e^2, \sigma_f^2 | y) \propto f(\beta) f(\mu | \sigma_f^2) f(\sigma_f^2) f(\sigma_e^2) f(y | \beta, \mu, \sigma_e^2, \sigma_f^2)$$

As seguintes rotinas do SAS correspondem à análise Bayesiana:

1) Numa primeira rodada, a distribuição *a priori* de Jeffreys foi considerada para obter os valores de densidade base dos componentes de variância:

```
PROC MIXED DATA=POPCORN;
CLASS FAMILY REP BLOCK;
MODEL CE = REP BLOCK(REP);
RANDOM FAMILY / S;
PRIOR JEFFREYS / OUT=BAYES1 NSAMPLE=50000;
RUN;
PROC UNIVARIATE DATA=BAYES1; WHERE SAMPLE >5000;
RUN; QUIT;
```

2) Obtendo os valores de densidade base, uma segunda rodada foi feita com as informações *a priori* para os componentes de variância, dada por:

```
DATA PRIOR;  
INPUT DENSITY TYPE $ PARM1 PARM2;  
CARDS;  
1 IG 81.272 2301  
2 IG 152.82 2904  
;  
PROC MIXED DATA=POPCORN;  
CLASS FAMILY REP BLOCK;  
MODEL D = REP BLOCK(REP);  
RANDOM FAMILY / S;  
PRIOR DATA=PRIOR / OUT=BAYES2 NSAMPLE=50000;  
RUN;  
PROC UNIVARIATE DATA=BAYES2; WHERE SAMPLE >5000;  
RUN; QUIT;
```

O coeficiente de correlação de Pearson foi calculado entre os valores genéticos preditos da capacidade de expansão e o rendimento, para determinar a relação genética entre as características, usando o procedimento PROC CORR de SAS.

3 RESULTADOS E DISCUSSÃO

Encontraram-se diferenças significativas entre as famílias para ambas as características, de acordo com os critérios de informação AIC e BIC (Tabela 13); os menores valores dos critérios de informação apontam a seleção do modelo que inclui o efeito familiar. Este resultado foi coincidente com o teste de razão de verossimilhança ($p < 0,05$).

Tabela 13. Critérios de informação de Akaike (AIC) e Bayesiano de Schwarz (BIC), o teste da razão de verossimilhança (LRT) (comparando os modelos: nulo e completo) e estimativas paramétricas usando modelos lineares mistos, via Máxima Verossimilhança Restrita (REML).

Indicadores	Capacidade de Expansão		Rendimento	
	Modelo completo	Modelo nulo	Modelo completo	Modelo nulo
Verossimilhança	-2	Res.		
Log.	2870,7	2879,6	6930,2	6940,9
AIC	2874,7	2881,6	6934,2	6942,9
BIC	2881,0	2885,7	6940,3	6947,0
LRT	8,9*		10,7*	
Estimativas paramétricas				
Variância Familiar	3,209	--	70835	--
Variância Residual	19,009	22,2263	328449	398196
Variância Aditiva	12,838	--	283340	--
h^2	0,336	--	0,393	--

* Modelo completo é estatisticamente significativo de acordo com o teste χ^2 ($p < 0,05$).

As estimativas dos componentes de variância e herdabilidade usando Máxima Verossimilhança Restrita (REML) são apresentadas na Tabela 13. A capacidade de expansão e o rendimento de grãos mostraram ser moderadamente herdável, com $h^2=0,34$ e $h^2=0,39$ respectivamente. Estes resultados são coincidentes com os valores de herdabilidade obtidos na linha Bayesiana (Tabela 14), com moda *a posteriori* de $h^2=0,34$ e intervalo de credibilidade de 0,22 – 0,44 (90% de probabilidade), para a capacidade de expansão, e $h^2=0,4$ com intervalo de credibilidade de 0,28 – 0,49, para o rendimento. Os intervalos de credibilidade Bayesianos dos componentes de variância e herdabilidade incluem o valor da estimativa REML,

confirmando a equivalência entre os procedimentos. As diferenças significativas entre as famílias permite a obtenção de ganhos genéticos que variam entre 4 e 5,2% para a capacidade de expansão, e entre 7,1 e 9,2%, para o rendimento de grãos (Tabela 15).

Tabela 14. Estimativas Bayesianas dos componentes de variância e herdabilidade (nível de famílias) para o rendimento de grãos e capacidade de expansão em milho pipoca.

Estimativas Bayesianas*						
Capacidade de Expansão	Média	Mediana	Moda	DP	LI	LS
Variância Familiar	3,249	3,218	3,056	0,865	1,889	4,721
Variância Residual	19,064	19,024	18,980	1,094	17,347	20,938
Variância Aditiva	12,994	12,870	12,224	3,459	7,555	18,886
h^2	0,334	0,337	0,344	0,066	0,220	0,438
Rendimento						
Variância Familiar	71702	71049	71962	17453	44219	101528
Variância Residual	329333	328582	328852	20084	297954	363756
Variância Aditiva	286806	284195	287848	69811	176876	406112
h^2	0,390	0,394	0,402	0,066	0,277	0,493

* DP: desvio padrão; LI e LS: limites inferior e superior dos intervalos de credibilidade (P=90%).

Analisando as herdabilidades, verifica-se (Tabelas 13 e 14) que a capacidade de expansão apresentou valores similares comparativamente ao rendimento de grãos. O intervalo de credibilidade da herdabilidade da capacidade de expansão inclui as estimativas pontuais da herdabilidade do rendimento. Os resultados da herdabilidade do rendimento e capacidade de expansão são menores que encontrados na literatura por Arnhold et al. (2009a), Daros et al. (2002) e Santos et al. (2008). Santos et al. (2004) estimaram herdabilidade de 0,72 para capacidade de expansão, em um ensaio genético com famílias S2 da população Beija-Flor, Vilarinho et al. (2003) encontraram as herdabilidades, na mesma população, com variação de 0,32 a 0,60 em famílias S₁ e S₂, respectivamente.

A capacidade de expansão apresenta dominância bidirecional (SCAPIM et al., 2002; SCAPIM et al., 2006; SIMON et al., 2004; ARNHOLD et al. 2009b), dificultando a exploração da heterose. De acordo com diversos autores, a herdabilidade da capacidade de expansão varia, em geral, de 70 a 90% (PEREIRA & AMARAL JÚNIOR, 2001; ARNHOLD

et al. 2006) e é uma característica que possui como principal componente de variância genética os efeitos aditivos (SCAPIM et al., 2006; SIMON et al., 2004; ARNHOLD et al. 2009b). Portanto, em geral, segundo ARNHOLD et al. (2009b), a capacidade de expansão é uma característica menos influenciada pelo ambiente e pelos desvios de dominância, quando comparada ao rendimento de grãos, o que pode acarretar maior associação entre desempenho *per se* e em cruzamento.

Na Tabela 15, visualizam-se as estimativas do ganho genético (seleção familiar), considerando duas intensidades de seleção (20,12 e 10,06%), calculadas de acordo com os valores preditos Bayesianos das famílias de milho pipoca. A significativa variabilidade genética entre as famílias de milho pipoca da população UEM-C1 permitiu a obtenção de progressos genéticos para ambas as características.

Tabela 15. Estimativas do ganho genético (seleção familiar) considerando duas intensidades de seleção (20,12 e 10,06%) calculadas de acordo com os valores preditos Bayesianos das famílias de milho pipoca.

	Capacidade de expansão		Rendimento	
	Intensidade de seleção			
	20,12%	10,06%	20,12%	10,06%
MBVS	1,30	1,67	149,8	195,7
NFS	34	17	34	17
Média dos ensaios	32,15	32,15	2121,4	2121,4
GG%	4,0	5,2	7,1	9,2

MBVS: Média dos *breeding values* das famílias selecionadas; NFS: número de famílias selecionadas; GG%: Ganho genético em porcentagem.

Os valores pontuais da média, mediana e moda das distribuições *a posteriori* dos componentes de variância e herdabilidade foram similares, (Tabela 14), evidenciando carência de assimetria. Consideraram-se cadeias com 50.000 amostras aleatórias com período de descarte (*burn-in*) de tamanho 5.000 (Figura 5). Os critérios de convergência das cadeias IC foram confirmados pelo estatístico de Heidelberger & Welch (1983). Os resultados da análise das cadeias estão em concordância com Wolfinger & Kass (2000). No algoritmo IC, a distribuição base para os componentes de variância é considerada exata e, por isso, as amostras MCMC são rejeitadas somente se estas não estão no espaço parâmetro, o qual poderia ocorrer, por exemplo, se um componente de variância fosse estimado com valor

negativo. Desse modo, o algoritmo IC é um caso especial de amostragem de rejeição (TIERNEY, 1994).

Uma vantagem dos procedimentos Bayesianos é a possibilidade de usar os limites de confiança (ou regiões de credibilidade no contexto Bayesiano) como alternativa de estimação dos parâmetros genéticos, os quais são obtidos diretamente da distribuição *a posteriori* (WRIGHT et al., 2000; NOGUEIRA et al., 2003). De acordo com NOGUEIRA et al. (2003), se existe informação *a priori*, a inferência Bayesiana pode determinar intervalos de credibilidade mais estreitos que os intervalos de confiança, além de testes mais poderosos. No entanto, é possível trabalhar com as chamadas distribuições *a priori* não informativas, quando for o caso. A grande vantagem da inferência Bayesiana, no contexto dos modelos mistos, é a relativa facilidade de obtenção de desvios-padrão para os elementos do efeito principal genotípico (GIANOLA & FOULLEY, 1990; NOGUEIRA et al., 2003).

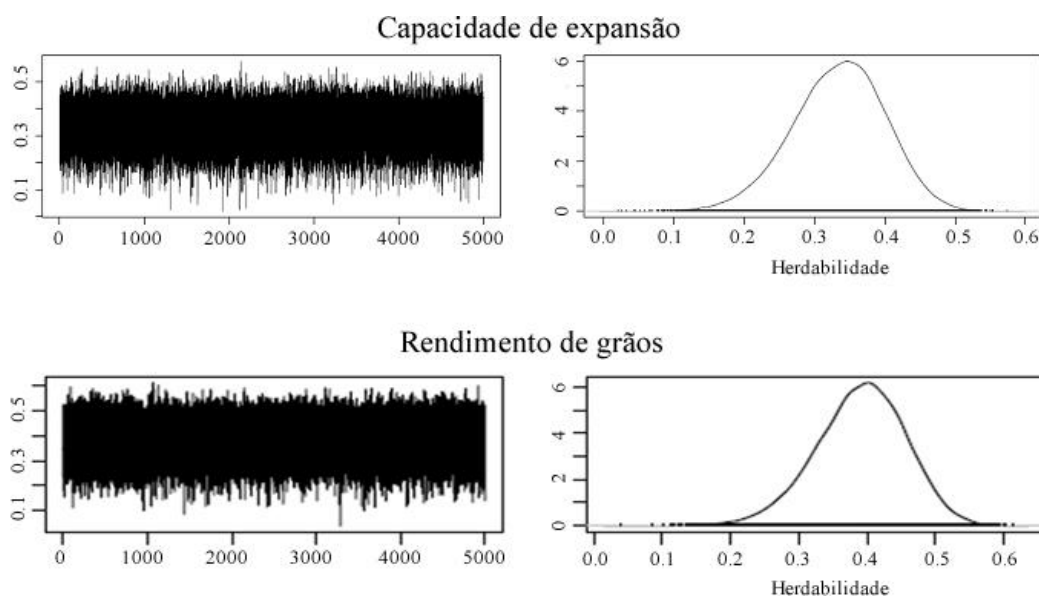


Figura 5. Distribuições marginais *a posteriori* das herdabilidades (nível de famílias), para as características capacidade de expansão e rendimento de grãos, avaliadas em 169 famílias de milho pipoca. À esquerda é mostrado o processo de iteração do algoritmo de cadeias independentes (análise Bayesiana).

Na Figura 6, é apresentado um gráfico com os valores genéticos de famílias, obtidos via predição Bayesiana para a capacidade de expansão e rendimento de grãos, os quais foram confrontados entre si. A linha reta é um indicativo do grau de associação entre os valores genéticos das 169 famílias de meios-irmãos. O coeficiente de correlação de Pearson (calculados entre os valores genéticos das famílias para cada característica) não foi

estatisticamente diferente de zero ($r = -0,103$; $p = 0,2112$), confirmando a falta de associação entre os valores de predição do efeito de famílias de ambas as características.

Uma correlação genética não significativamente diferente de zero, indicaria que a seleção de famílias baseada na capacidade de expansão não teria efeito significativo sobre o rendimento de grãos. Este resultado é amplamente reportado na literatura. Sawazaki (1995), Arnhold et al. (2009a) e Coimbra et al. (2001), por exemplo, confirmaram que a produção de grãos é inversamente correlacionada (associação genética negativa) com a capacidade de expansão. Mora & Arnhold (2006) estimaram coeficiente de correlação de Spearman não significativo em 96 famílias S3 de milho pipoca. Scapim et al. (2002) mencionou que a negativa associação entre a capacidade de expansão e rendimento pode interferir na obtenção de ganhos de seleção simultâneas numa população de melhoramento. Do ponto de vista genético, este fenômeno indica que ambas as características poderiam estar controladas por genes distintos.

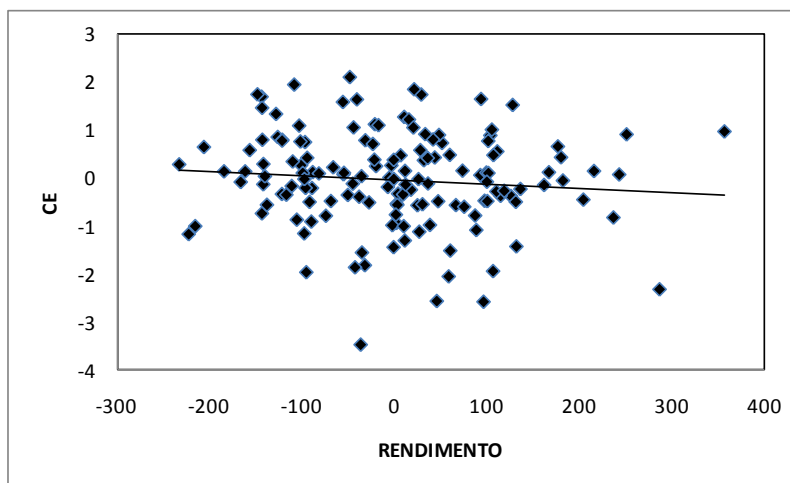


Figura 6. Gráfico de dispersão com os valores genéticos de famílias, obtidos via predição Bayesiana para capacidade de expansão (CE) e rendimento de grãos, avaliados em 169 famílias de meios-irmãos de milho pipoca (população UEM-C₁).

4 CONCLUSÃO

Houve significativa variabilidade genética entre as famílias de milho pipoca da população UEM-C1, o que permitiu a obtenção de progressos genéticos para rendimento de grãos e capacidade de expansão. Porém, a correlação genética foi nula entre ambas as características que limitaria o progresso genético de seleção simultânea.

A avaliação genética, usando abordagem Bayesiana, possibilitou realizar inferências científicas na população de melhoramento, mostrando-se uma ferramenta apreciável para auxiliar no processo de seleção de famílias de milho pipoca e corresponde a um procedimento equivalente a sua contraparte dos modelos lineares mistos clássicos.

5 REFERÊNCIAS

ARNHOLD, E.; MORA, F., SILVA, R. G.; GOOD-GOD, P. I. V.; RODOVALHO, M.A. Evaluation of Top-Cross Popcorn Hybrids Using Mixed Linear Model Methodology. **Chilean Journal of Agricultural Research**, Chillán, v. 69, n. 1, p. 46-53, 2009a.

ARNHOLD, E.; VIANA, J. M. S.; SILVA, R. G. Associação de desempenho entre famílias S3 e seus híbridos topcross de milho-pipoca. **Revista Ciência Agronômica**, Fortaleza, v. 40, n. 3, p. 396-399, 2009b.

BISOGNIN, D. A.; CIPRANDI, O.; COIMBRA, J. L. M.; GUIDOLIN, F. Potencial de variedades de polinização aberta de milho em diferentes condições adversas de ambientes. **Pesquisa Agropecuária Gaúcha**, Porto Alegre, v. 3, n. 1, p. 29-34, 1997.

BUENO FILHO, J. S.; VENCOVSKY, R. Alternativas de análise de ensaios em látice no melhoramento vegetal. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 2, p. 259-269, 2000.

COIMBRA, R. R. G. V.; MIRANDA, J.M.S.; VIANA, C. D. CRUZ. Correlações entre caracteres na população de milho-pipoca DFT-1 Ribeirão. **Revista Ceres**. Viçosa, v. 48, p. n. 1, 427-435, 2001.

DAROS, M.; AMARAL Jr. A. T.; PEREIRA, M. G. Genetic gain for grain yield and popping expansion in full-sib recurrent selection in popcorn. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 2, n. 3, p. 339-344, 2002.

DOERKSEN, T. K.; KANNENBERG, L. W.; LEE, E. A. Effect of recurrent selection on combining ability in maize breeding populations. **Crop Science**, Madison, v. 43, 5, 1652-1658, 2003.

GIANOLA, D.; FOULLEY, J. L. Variance estimation from integrated likelihood (VEIL). **Génétique, Sélection, Évolution**, Versailles, v. 22, p. 403-17, 1990.

HALLAUER, A. R.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. Ames : Iowa State University, 1981. 468p.

HEIDELBERGER, P.; WELCH, R. D. Simulation run length control in the presence of an initial transient. **Operations Research**, Stanford, 31, p. 1109-1144, 1983.

MATTA, F. P.; VIANNA, J. M. S. Eficiências relativas dos processos de seleção entre e dentro de famílias de meios-irmãos em população de milho-pipoca. **Ciência Agrotecnologia**. Lavras, v. 27, n.3, p. 548-556, 2003.

MORA, F.; ARNHOLD, E. Inferencia Bayesiana y metodología de modelos lineales mixtos aplicados al mejoramiento del maíz. **Ciencia e investigación agrária**, Santiago, v. 33, n. 3, 217-223, 2006.

NOGUEIRA D. A.; SÁFADI, T.; BEARZOTI, E.; BUENO-FILHO, J.S.S.. Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência Agrotecnologia**. Lavras, v. 27 n.3 p. 1614-1624, 2003.

PACHECO, C. A. P.; GAMA, E. P.; GUIMARÃES, P. E. O.; SANTOS, M. X.; FERREIRA, A. S Estimativas de parâmetros genéticos nas populações CMS-42 e CMS-43 de milho pipoca. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, 12, p. 1995-2001, 1998.

PACHECO, C. A. P.; SANTOS, M. X.; CRUZ, C. D.; PARENTONI, S. N.; GUIMARÃES, P. E. O.; GAMA, E. E. G.; SILVA, A. E.; CARVALHO, H. W. L.; VIEIRA Jr, P. A. Inbreeding depression of 28 maize elite open pollinated varieties. **Genetics and Molecular Biology**, Ribeirão Preto, v. 25, n.4, 441-448, 2002.

PATERNIANI, E.; VIEGA, G. P. **Melhoramento e produção do milho**. Campinas: Fundação Cargill, 1987. 795p.

PEREIRA, M.G.; AMARAL JÚNIOR, A.T. Estimation of genetic components in popcorn base on the nested design. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 01, n. 01, p. 3-10, 2001.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007. 561p.

RODOVALHO, M. A.; MORA, F.; SANTOS, E. M.; SCAPIM, C. A.; ARNHOLD, E. Herdabilidade de la sobrevivencia en 169 familias de maíz-roseta de granos blancos: un enfoque Bayesiano. **Ciencia e Investigación Agraria**. Santiago, v. 55, n.3, p. 255-260, 2008.

SANTOS, F. S.; AMARAL JÚNIOR, A. T.; FREITAS JÚNIOR, S. P.; RANGEL, R. M.; SCAPIM, C. A.; MORA, F. Genetic gain prediction of the third recurrent selection cycle in a popcorn population. **Acta Scientiarum Agronomy**, Maringá, v. 30, p. 651-655, 2008.

SANTOS, J. F.; VIANA, J. M. S.; VILARINHO, A. A.; CÂMARA, T. M. M. Efficiency of S2 progeny selection strategies in popcorn. **Crop Breeding Applied Biotechnology**. Viçosa, v. 5, p. 183-191, 2004.

SAS-INSTITUTE. Software: **statistical analysis system institute**. Cary: s.n., 1992

SAWAZAKI, E. Melhoramento do milho-pipoca. **Instituto Agronômico de Campinas**, Campinas, São Paulo, Brasil, 1995. 21p.

SCAPIM, C. A.; PACHECO, P. C. A.; TONET, A.; LUCCA e BRACCINI, A.; PINTO, R. J. B. Análise dialética e heterose de populações de milho-pipoca. **Bragantia**, Campinas, v. 61, n. 3, 219-230, 2002.

SCAPIM, C. A.; BRACCINI, A. L.; PINTO, R. J. B.; AMARAL JÚNIOR, A. T.; RODOVALHO, M. A.; SILVA, R. M.; MOTERLE, L. M. Componentes genéticos de médias e depressão por endogamia em populações de milho pipoca. **Ciência Rural**, Santa Maria, v. 36, n. 01, 36-41, 2006

SIMON, G. A.; SCAPIM, C. A.; PACHECO, C. A. P.; PINTO, R. J. B.; BRACCINI, A. L.; TONET, A. Depressão por endogamia em populações de milho-pipoca. **Bragantia**, **Bragantia**, Campinas, v. 63, n. 01, p. 55-62, 2004.

SOUZA JR, C. L. Comparisons of Intra-, Interpopulation, and Modified Recurrent Selection Methods. **Revista Brasileira Genética**. Ribeirão Preto, v. 16, n.1, 91-105, 1993.

TIERNEY, L. Markov chains for exploring posterior distributions. **Annals of Statistics**. v. 22, p. 1701–1786, 1994.

VIANA, J. M. S. Melhoramento intrapopulacional recorrente de milho-pipoca, com famílias de meios-irmãos. **Revista Brasileira de Milho e Sorgo**, Sete Lagoas, v. 6, 199-210, 2007.

VILARINHO, A. A.; VIANA, J. M. S.; SANTOS, J. F.; CÂMARA, T. M. M. E. Eficiência da seleção de progênies S1 e S2 de milho-pipoca, visando à produção de linhagens. **Bragantia**, campinas, v .62 n. 01 p. 9-17, 2003.

WOLFINGER, R. D.; KASS R. E. Non-conjugate Bayesian analysis of variance component models. **Biometrics**, Arlington, v.56 768-774, 2000.

WRIGHT, D. R.; STERN, H. S.; BERGER, P. J. Comparing traditional and Bayesian analyses of selection experiments in animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 5, n. 2, p. 240-256, 2000.

APÊNDICES

CAPÍTULO 3

Análise de convergência das cadeias de independência para Capacidade de Expansão

HEIDELBERGER AND WELCH STATIONARITY AND INTERVAL HALFWIDTH TESTS

=====
=====

Iterations used = 1:50000

Thinning interval = 1

Sample size per chain = 50000

Precision of halfwidth test = 0.1

\$chain1

	Stationarity test	start iteration	p-value
var1	passed	10001	0.315
var2	passed	1	0.355
var3	passed	10001	0.315
var4	passed	1	0.140
var5	passed	1	0.110

Halfwidth Mean Halfwidth

test

var1 passed	3.245	0.007371
var2 passed	19.065	0.009466
var3 passed	12.980	0.029484
var4 passed	0.334	0.000483
var5 passed	0.581	0.001102

Análise de convergência das cadeias de independência para Rendimento

HEIDELBERGER AND WELCH STATIONARITY AND INTERVAL
HALFWIDTH TESTS

=====
=====

Iterations used = 1:50000

Thinning interval = 1

Sample size per chain = 50000

Precision of halfwidth test = 0.1

\$chain1

	Stationarity test	start iteration	p-value
var1	passed	1	0.397
var2	passed	1	0.639
var3	passed	1	0.397

	Halfwidth test	Mean	Halfwidth
var1	passed	71678	167
var2	passed	329339	198
var3	passed	286713	670

HEIDELBERGER AND WELCH STATIONARITY AND INTERVAL
HALFWIDTH TESTS

=====
=====

Iterations used = 1:50000

Thinning interval = 1

Sample size per chain = 50000

Precision of halfwidth test = 0.1

\$chain1

	Stationarity test	start iteration	p-value
var1	passed	1	0.504
var2	passed	1	0.473

	Halfwidth test	Mean	Halfwidth
var1	passed	0.390	0.00067