

Modelos mistos para seleção de genótipos superiores e de futuros genitores de cana-de-açúcar

Edjane Gonçalves de Freitas¹

Maria Marta Pastina²

Rodrigo Gazaffi¹

Luciana Rossini Pinto³

Mauro Alexandre Xavier³

Marcos Guimarães de Andrade Landell³

Antonio Augusto Franco Garcia¹

1 Introdução

A evidente expansão do setor sucroalcooleiro do Brasil tem enfrentado vários desafios, e o melhoramento genético tem sido essencial para assegurar o sucesso do setor. Todos os anos, desenvolve e libera novas variedades com potencial produtivo superior às variedades comerciais. A introdução no mercado de materiais competitivos encara o desafio de atender a crescente demanda nacional e internacional no que concerne à produção de açúcar e álcool. Cresce a perspectiva de produção do etanol de segunda geração (oriundo da palha e/ou do bagaço do cana) em escala industrial e paralelamente, aumenta-se o investimento em pesquisas. Nesse cenário positivo, o melhoramento genético busca aprimorar novas metodologias de análise a fim de otimizar o processo de obtenção e seleção de genótipos superiores, na expectativa de atender a exigência do mercado de açúcar e biocombustível, através da oferta de materiais genéticos com elevados índices de rendimento industrial.

A criação e a avaliação de variedades melhoradas é uma das mais antigas atividades da pesquisa agrícola com grande impacto sobre a produção mundial de alimentos, especialmente nos últimos 100 anos. Especificamente no caso da cana-de-açúcar, na década de 1970 intensificaram-se as pesquisas de melhoramento genético no Brasil, e surgiram as primeiras variedades liberadas que alimentaram o mercado de açúcar e álcool. Historicamente, o emprego do conhecimento de genética quantitativa foi marcante no melhoramento de cana, destacando-se os estudos de fatores genéticos relacionados especialmente ao potencial produtivo e ao entendimento e exploração da variabilidade genética. Borlaug (2001) apud Smith, Cullis e Thompson (2005),

¹LCE - ESALQ/USP. e-mail: edjane.gf@gmail.com

²EMBRAPA

³IAC

argumenta que é importante reconhecer que os métodos convencionais de melhoramento continuam a fazer contribuições significantes para produção de alimentos, e, nesse sentido, é essencial que métodos estatísticos usados para projetar e analisar dados experimentais sejam tão preciso, eficiente e informativo quanto possível.

A aplicação de ANAVA, seguida decomposição da variância fenotípica em componentes ambiental e genético, permitiu estimar componentes de variância, a herdabilidade dos caracteres de interesse e o ganho genético, parâmetros importantes recorridos na decisão de seleção. As atuais variedades de cana-de-açúcar que sustentam o setor sucroalcooleiro são resultados do melhoramento clássico, e refletem a contribuição histórica da genética quantitativa, muito embora baseada em modelos estatísticos com fortes restrições.

Nos programas de melhoramento de cana-de-açúcar, todos os anos são instalados vários experimentos com o objetivo de selecionar genótipos superiores, provenientes de hibridações entre genitores previamente selecionados (MARIOTTI, 1973; MATSUOKA et al., 1999). São vários genótipos candidatos, tornando a seleção mais difícil, e exigindo do melhorista o uso de ferramentas de análise apropriadas para aumentar a eficiência do processo de seleção e garantir o sucesso do programa de melhoramento. O emprego de métodos de genética quantitativa, junto com métodos estatísticos adequados, pode resultar em melhores estimativas de componentes variância genéticos e residuais, e conseqüentemente em predições mais acuradas do valor genotípico.

O modelo tradicional (ANAVA), todos os efeitos são fixos, exceto o erro experimental, que é sempre aleatório. Essa abordagem deu grandes contribuições ao melhoramento genético, entretanto, é limitada devido à pressuposição de independência dos erros, que no contexto dos experimentos de melhoramento genético é inadequada. Outra limitação dessa técnica ocorre quando se tem desbalanceamento de dados, seja por perda de parcelas ou quando o delineamento não é ortogonal. Há também a situação em que cada parcela fornece vários dados em diferentes locais e anos (colheitas), e são, por isso, correlacionados.

A eficiência de seleção num programa de melhoramento pode ser aumentada com o emprego de modelos estatísticos que representem mais fielmente a natureza dos dados. Isso é conseguido com a modelagem de estruturas de variância-covariâncias (VCOV) genética dos indivíduos nos diversos locais e colheitas, modelando a interação genótipo \times ambiente. As informações de covariância quando modelada adequadamente aumenta o poder preditivo do modelo, e leva à obtenção de estimativas acuradas do valor genotípico (VG), aumentando a eficiência de seleção.

A metodologia de modelos mistos é uma alternativa aos modelos de ANAVA. Permite-se modelar efeitos fixos e aleatórios além do erro experimental, é flexível, no sentido de permitir estruturar a informação de covariância, e tem sido aplicado com sucesso no melhoramento animal e de plantas. Peculiaridades, tais como a modelagem de informação de VCOV genética

entre os efeitos do modelo e a incorporação de parentesco genético, despertam o interesse de pesquisadores, e a aplicação dessa abordagem tem aumentado, especialmente em pesquisas de Seleção Assistida por Marcadores (MAS) e Seleção Genômica Ampla (GWS)(CROSSA et al., 2006, 2007; BOER et al., 2007; BAUER; LÉON, 2008; LORENZANA; BERNARDO, 2009; SCHULZ-STREECK; PIEPHO, 2010; ZHANG et al., 2010; BURGUEÑO et al., 2012).

No desenvolvimento de novas variedades de cana-de-açúcar, como também a liberação e a recomendação para plantio comercial requerem que etapas de seleção sejam praticadas entre um grande número de genótipos candidatos. É comum a obtenção de banco de dados extremamente desbalanceados em função de sucessivo descarte de genótipos ocorrido durante o processo de seleção e/ou implantação de ensaios com apenas alguns genótipos comuns. Nesses casos, a análise de variância conjunta tradicional torna-se impraticável. A análise conjunta dos experimentos com predição do VG dos genótipos em experimentação é importante para estabelecer, durante o processo de seleção, comparações entre indivíduos em função da superioridade genética, e com efeito, auxiliar na decisão de seleção. Assim, o modelo misto juntamente com a metodologia REML/BLUP (do inglês, *Restricted Maximum Likelihood/Best Linear Unbiased Prediction* - REML/BLUP) pode fornecer predições mais acuradas do VG, medida usada como referência no ranqueamento e seleção de genótipos superiores destinados à liberação comercial. Para o melhoramento de cana-de-açúcar, tal abordagem é uma boa alternativa que pode aumentar a eficiência de seleção.

Os experimentos com cana-de-açúcar normalmente são conduzidos em vários locais e anos (colheitas), chamados de METs (do inglês, *multi-environment trial* - MET), estratégia que visa aumentar a eficiência de seleção, contornando o efeito ambiental sobre os genótipos. Entretanto, modelos de predição mais realistas são necessários, que incluam uma adequada estrutura de variância-covariância para efeitos de interação.

Diante do exposto, considerando as características dos ensaios com cana-de-açúcar, o uso dos modelos mistos com a modelagem da matriz de variância-covariância para os efeitos de interação é uma estratégia interessante e poderá fornecer predições mais acuradas do VG. Tal modelo poderá auxiliar os melhoristas na tomada de decisão durante processo de seleção e recomendação de novas variedades para plantio comercial.

Nesse contexto, o objetivo desse trabalho foi determinar um modelo para predição do valor genotípico de genótipos de cana-de-açúcar utilizando a abordagem de modelo misto, modelando adequadamente a matriz de variância-covariância genética para interação $G \times L \times C$. Tal modelo buscará o aumento da eficiência de seleção e poderá auxiliar na seleção de genótipos superiores.

2 Material e métodos

2.1 Rede de experimentos

O banco de dados utilizado para análise foi gentilmente fornecido pelo o programa de melhoramento genético do Centro da Cana-de-açúcar do Instituto Agronômico de Campinas - IAC. As informações são de um período de doze anos, referentes aos anos de 1997 a 2001 e de 2003 a 2009. A rede experimental analisada é constituída por 74 genótipos, sendo que nem todos os indivíduos estão presentes em todos experimentos. A frequência de um genótipo em experimentos é variável, há genótipos que foram avaliados mais vezes, os quais foram submetidos a um maior número de ambientes diferentes, enquanto outros foram menos. Um total de 44 experimentos foram avaliados. O modelo será ajustado para o caráter TPH (toneladas de pol por hectare), uma medida importante para o melhoramento, que mensura conjuntamente a quantidade de sacarose produzida por uma tonelada de cana colhida em um hectare.

2.2 Desenvolvimento do Modelo Misto

As análises dos experimentos foram realizadas utilizando-se o *software* GenStat 14^o (PAYNE et al., 2009). A princípio buscou-se ajustar o modelo misto adequado que represente o mais fielmente o padrão natural de resposta dos dados. Isso foi feito comparando-se diferentes estruturas de VCOV para os efeitos genéticos de local e colheita (Tabela 1). O modelo matemático descrito aqui foi uma adaptação do modelo usado por Pastina (2010) para análise de MET, de modo que sua descrição seguiu a mesma notação. A notação foi originalmente descrita por Eckermann et al. (2001), Verbyla et al. (2003) e Boer et al. (2007). O modelo estatístico, no qual o sublinhado identifica a variável aleatória, pode ser escrito como:

$$\underline{y}_{ijk r} = \mu + L_j + \underline{C}_k + \underline{LC}_{jk} + \underline{G}_{ijk} + \underline{\varepsilon}_{ijk r} \quad (1)$$

em que $\underline{y}_{ijk r}$ é a resposta fenotípica do i -ésimo genótipo na r -ésima repetição do j -ésimo local e k -ésima colheita; μ é a média geral; L_j é o efeito de local; \underline{C}_k o efeito de colheita; \underline{LC}_{jk} é o efeito da interação local e colheita; \underline{G}_{ijk} é o efeito genético do genótipo i no local j e colheita k ; e $\underline{\varepsilon}_{ijk r}$ os efeitos não genético. Nesse estudo, todos os indivíduos (clones e variedades) foram designados como genótipos devido as testemunhas serem diferentes entre os experimentos.

O efeito genético aleatório \underline{G}_{ijk} foi aqui representado por uma notação em minúscula, \underline{g}_{ijk} , com vetor $\underline{\mathbf{g}} = (g_{111}, \dots, g_{IJK})$ que possui distribuição normal multivariada com média zero e

matriz de VCOV \mathbf{G} , escrito na forma de $\mathbf{g} \sim N(0, \mathbf{G})$. E o termo $\underline{\varepsilon}_{ijk_r}$, foi modelado como:

$$\underline{\varepsilon}_{ijk_r} = b_{jkr} + \eta_{ijk_r} \quad (2)$$

em que b_{jkr} é o efeito do bloco r no local j e colheita k ; η_{ijk_r} é o erro experimental, assumido que $\eta_{ijk_r} \sim N(0, \mathbf{R})$, com $\mathbf{R} = \sigma^2 \mathbf{I}$, considerada assim para todos os modelos avaliados.

Todas as interação duplas e triplas entre os efeitos foram também incluídas no modelo, mas por questão de simplificação, não foram demonstrados aqui.

A rotina de análise foi procedida em duas etapas. Na primeira foram ajustados 20 modelos (1 - 20) utilizando a metodologia de modelos mistos descrita por Henderson (1986) e McCulloch e Searle (2001) (Tabela 1). Nesta, cada modelo foi avaliado utilizando-se simultaneamente duas diferentes estruturas de VCOV genética, sendo uma para local ($\mathbf{G}_{J \times J}^L$) e outra para colheita ($\mathbf{G}_{K \times K}^C$). Assumiu-se ausência de parentesco genético entre os indivíduos, dessa forma, a relação entre os genótipos foi modelada por uma matriz identidade ($\mathbf{I}_{I \times I}^G$). As predições do valor genotípico média de TPH (VG) para genótipo teve matriz de VCOV genética dada por $\mathbf{G}_{I \times I} = \mathbf{I} \sigma_g^2$, em que σ_g^2 é o componente de variância genética. Já na segunda etapa foram ajustados mais 20 modelos (21 - 40). Nestes, foi incluída a informação de parentesco através da matriz \mathbf{A} . Assim, a predição do valor genético médio (VA) de TPH para cada genótipo, teve matriz de VCOV genética dada por $\mathbf{G}_{I \times I} = \mathbf{A} \sigma_a^2$, em que σ_a^2 é o componente de variância aditiva (FALCONER e MACKAY, 1996; LYNCH; WALSH, 1998; PIEPHO et al., 2008).

Em experimentos com cana-de-açúcar, conforme ressaltado anteriormente, os dados são oriundos de diversos locais e colheitas. Assim a matriz de VCOV genética global \mathbf{G} foi modelada através do produto direto entre matrizes de VCOV de componentes do modelo estatístico que contêm informação de correlação genética. Nesse contexto, para os 20 primeiros modelos foi assumido ausência de parentesco genético, e a matriz de VCOV \mathbf{G} foi obtida da seguinte forma:

$$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G \quad (3)$$

e para os modelos com a informação de parentesco (21 a 40), foi dada por:

$$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G \quad (4)$$

em que, \mathbf{G}^L , \mathbf{G}^C e \mathbf{A}^G são as matrizes variância-covariância genética para locais, colheitas e de parentesco genético, respectivamente. Essas duas condições (3 e 4) foram verificadas a fim de conhecer a contribuição dada com a inclusão da genealogia. Uma vez que tais modelos são mais realistas, espera-se um maior poder preditivo, e portanto o aumento da eficiência de seleção (Tabela 1).

Os 40 modelos foram examinados e comparados através dos critérios AIC e BIC. Segundo esses critérios o modelo que melhor se ajusta aos dados é aquele que tem o menor valor dessas

estatísticas. Após selecionado o “melhor” modelo, foram obtidas as predições do valor genotípico médio (VG) e/ou do valor genético médio (VA) de TPH, e, em seguida, foi definido o ranqueamento dos genótipos.

Como a análise tradicional de ANAVA conjunta é inviável, para efeito de comparação, estimou-se a média aritmética de TPH. Esse tipo de análise, é o procedimento estatístico comumente usado por melhorista de cana para avaliar o potencial produtivo médio dos genótipos nos grupos de experimentos, especialmente nos casos em que nem todos os genótipos são avaliados em todos experimentos. A análise conjunta via ANAVA é dificultada devido o grande desbalanceamento existente. Para o conjunto de dados analisado aqui, a extensão do desbalanceamento tornou-se maior devido existir diferentes genótipos, números de repetições e também diferentes números de colheitas. Retrata-se então, uma situação frequentemente enfrentada pelos programas de melhoramento. Neste caso, o procedimento clássico de análise de variância conjunta tornou-se inviável em razão das pressuposições existentes. Em geral, para contornar esse impasse, uma alternativa que tem sido utilizada, na maioria da vezes, é simplesmente o cálculo da média aritmética simples para cada indivíduo. Entretanto, a média de cada genótipo é calculada com números diferentes de observações devido ao grande desbalanceamento dos dados. Assim, análogo aos programas de melhoramento, foi calculado a média aritmética (tratada aqui como Modelo Tradicional), e, em seguida, foi estabelecido o ranqueamento dos genótipos.

Em relação a obtenção das estimativas do modelo, os efeitos fixos foram estimados pelo *Best Linear Unbiased Estimator* (BLUE), e os aleatórios por *Best Linear Unbiased Predictor* (BLUP), já implementados no *software* Genstat. O processo de estimação ocorreu via Máxima Verossimilhança e Máxima Verossimilhança Restrita (REML) (PATTERSON e THOMPSON, 1971).

3 Resultados e discussões

De acordo com o critério AIC e BIC, em geral os modelos com a inclusão da informação de parentesco (21 - 40), tendem a ser melhores quando comparados aos correspondentes (1 - 20) sem esta informação. Esse resultado confirma que a informação de parentesco contribuiu para melhoria do modelo, tornando-o mais realista. Uma explicação é que os genótipos compartilham locos que são idênticos por descendência, e através de cruzamentos, essa herança genética é transmitida para novas gerações. Finalmente há formação de uma grande “rede de relacionamento” com indivíduos de diferentes graus de parentesco. Com essas considerações, modelos que assumem independência entre os genótipos, não desfruta das relações existentes. Em contra partida, modelos que incluem o parentesco oferecem aos genótipos a oportunidade absorver a informação da genealogia e mesmo os genótipos com pouca informação genealógica são beneficiados.

Adicionalmente, a informação de genealogia no modelo dá uma expectativa, em termos de desempenho produtivo, do valor genético que o indivíduo tem e transmite para sua descendência. Para tanto, os efeitos genéticos aditivo (valor genético) dos genitores transmitido para progênie são incluídos no modelo via matriz de parentesco. Assim, o modelo resultante fornece predições do valor genético. No melhoramento de cana, as predições geradas por esse modelo mensura o quanto de variância genética aditiva um genótipo poderá transmitir a sua descendência, portanto permite ao melhorista inferir sobre potencial do indivíduo como genitor.

É interessante destacar o Modelo 1 que assumi a estrutura de independência (ID) para modelar o efeito genético nos diferentes locais, colheitas. Esse resultado sugere que, mesmo quando se tem grandes desbalanceamento dos dados, em detrimento da abordagem de ANAVA, é possível realizar a análise conjunta de grupos de experimentos, sem a necessidade de retirar indivíduos que estão presentes apenas em alguns experimentos mas não em todos.

Vale ressaltar que modelar estruturas de VCOV para efeitos relacionados não é uma obrigação, trata-se apenas de uma sutileza estatística para obter modelos mais realistas com potencial preditivo na expectativa que tais modelos possam maximizar a acurácia das predições. Nesse contexto, uso do modelos misto sem modelar estrutura de VCOV, em primeiro momento, pode e deve ser incorporado na rotina de análise dos programas de melhoramento genético de cana-de-açúcar. Certamente a análise conjunta grupos de experimentos pode ser realizada com vantagens, utilizando abordagem de modelos misto, e com o uso de modelo simples semelhante ao Modelo 1. Uma vez que o Modelo 1 é similar a abordagem tradicional de ANAVA que assume homogeneidade de variância para local e colheita, além de ausência de correlação genética entre os genótipos, portanto, não requer maiores conhecimento estatístico do melhorista.

A seleção do “melhor” modelo pode ser baseada em dois critérios de informação: o AIC e o BIC. Assim, em relação a decisão de seleção, para a maioria dos modelos, os dois critérios foram discordantes. Como exemplo tem-se o Modelo 19 (FA1 \otimes UNST \otimes ID, com 102 parâmetros) que apresentou o menor valor de AIC (41426,59) e o Modelo 29 (ID \otimes AR1 \otimes Kinship, com 2 parâmetros) o menor valor de BIC (41885,83), ambos foram destacados com negrito (Tabela 2). Uma justificativa para esse resultado deve-se ao grande número de experimentos analisado. O banco é composto de 44 experimentos, aqui também referidos como locais. A matriz de VCOV genética para local tem dimensão 44×44 , ou seja, é bastante grande. Os modelos de alternativos de matriz de VCOV para local são bastante parametrizados, conseqüentemente foram mais penalizado pelo critério BIC. Deve-se ressaltar que o tamanho da amostra eficaz para utilizar no cálculo em não claramente definido no âmbito de modelos mistos (PAULER, 1998). Verifica-se que em geral, o critério de informação BIC aplica uma maior penalidade para a matriz de local, priorizando o uso de estrutura mais simples, porém menos realista, como a ID ou DIAG. Esse resultado evidencia que para esse grupo de experimentos, o uso desse critério não faz sentido, uma vez que, a presença de heterogeneidade de variância e correlação genética entre locais é esperada. Esse argumento sugere que para este

caso, particularmente, o AIC poderá ser o melhor critério de seleção. Aplica menor penalidade, e portanto permite o uso de estruturas mais complexas. Proporcionando o ajuste de modelos mais realista para local e colheita.

Diante do exposto, para seleção do melhor modelo considerou-se apenas o critério AIC. Assim, de acordo com AIC verificou-se que o a estrutura FA1 foi a melhor para local. Outros estudos tem recomendado a estrutura Fator analítico para modelar a VCOV genética de local (THOMPSON et al., 2003; BOER et al., 2007; SMITH et al., 2007; MEYER, 2009), portanto, esse resultado era esperado. Já a matriz UNST foi a melhor para colheita. A matriz de VCOV genética para colheita tem dimensão 5×5 , requerendo a estimação de no máximo 15 parâmetros. É uma matriz pequena e o ajuste do modelo UNST foi rápido. Esse tipo de matriz captura toda a variação de forma geral, porém requer estimação do número máximo de parâmetros, o que a torna indesejável. Assim, para colheita optou-se por outro modelo mais parcimonioso, definido como AR1 (autoregressiva heterogênea de primeira ordem). A estrutura AR1 captura a informação de correlação entre sucessivas colheitas, e explica de forma sistemática a dependência temporal existente. Nesse contexto, o conhecimento do padrão de resposta dos dados deve ser considerado durante a escolha do modelo adequado. A produtividade da cana-de-açúcar tende a reduzir ao longo das colheitas, e diante desse comportamento, espera-se, por exemplo, uma maior correlação entre colheitas mais próximas (corte 1 e 2), e menor entre as mais distante (corte 1 e 3), em função de alterações fisiológicos e genéticas. Nesse sentido, o Modelo 11 (FA1 \otimes AR1 \otimes ID, com 89 parâmetros), o segundo melhor de acordo com AIC, pode ser considerando o melhor modelo. É compatível os padrões de respostas observado no campo, na avaliação de grupos de experimentos de cana-de-açúcar. Como esperado, assume a matriz FA1 para local e AR1 para colheita. Acredita-se que a estrutura AR1 não foi indicada pelo AIC como a melhor, devido esse critério não penalizar fortemente estruturas mais parametrizadas, pelo menos para colheita, ao contrário do BIC, observa-se uma certa prioridade por estruturas mais parametrizadas. Embora para o AIC a penalização tende a ser mais forte quando a matriz de VCOV tem muitos parâmetros a serem estimados, como a estrutura de VCOV para local, por exemplo. Ela invariavelmente foi penalizada fortemente por ambos critérios, o que resultou numa série de modelos que não rodou e foram identificados por NR.

O Modelo 11, no geral é o melhor modelo. É um modelo realista que assume estrutura adequadamente de VCOV para explicar as interações existentes entre genótipos, locais e corte. Este modelo poderá ser utilizado para predição do valor genotípico de genótipos de cana-de-açúcar em programas de melhoramentos. Os valores genotípicos quando ranqueados, poderão auxiliar o melhorista na identificação e seleção de genótipos superiores e de futuras variedades para recomendação comercial.

Vale salientar que um aspecto intrigante do Modelo 11 é que, embora tenha sido considerado o melhor modelo, ele não inclui a informação de parentesco genético. E como já mencionado, de acordo com AIC, os modelos são “melhorados” quando a informação de parentesco

é incorporada. Verifica-se então, que o Modelo 31 correspondente ao 11 com o parentesco (FA1 \otimes AR1 \otimes Kinship, com 89 parâmetros) não rodou devido a limitações computacionais (Computador Dell Studio XPS 8100 Intel[®] CoreTM i7-860, 17, 2.80GHz 16GB utilizando o Windows 7 Home Premium e o GenStat 14^o). Logo de início houve problemas com alocação de memória o que inviabilizou o início do processo de iterativo. Mas quando colocados lado a lado os modelos sem parentesco (1 - 20) e com parentesco (21 - 40), verifica-se que sempre o modelo correspondente com o parentesco é melhor. Portanto, as evidências sugerem que se o Modelo 31 tivesse rodado, provavelmente ele seria melhor modelo. Embora não haja resultados, esse modelo seria mais adequado. O mesmo admite estruturas de VCOV genética (para local e colheitas) que explicam mais adequadamente a natureza das correlações existentes. E adicionalmente, incorpora a informação de parentesco genético. Em resumo, trata-se de um modelo mais realista, e portanto, espera-se que possua maior poder preditivo e que seja capaz de gerar predições mais acuradas, o que resultaria numa maior eficiência de seleção.

As comparações entre os modelos são mostradas também nas Figuras 1-3. Foi estabelecido um ponto de corte que corresponde a seleção dos 10 genótipos superiores (uma intensidade de seleção 13,51%) para ambos modelos. Determinou-se o coeficiente de correlação *Spearman* (ρ) para cada comparação, afim de medir a intensidade da associação entre os modelos em função das coincidências de ranqueamento. A Figura 1 mostra a comparação entre o Modelo 1 e o Tradicional. Os valores genotípico preditos obtidos com o Modelo 1 e as médias estimadas de TPH com o Modelo Tradicional foram correlacionados significativamente ($\rho = 0,72$, p-valor < 0,001). No ponto de corte, nota-se que existem cinco genótipos (50%) que seriam selecionados com o Modelo 1, mas não com o modelo Tradicional. Isso reforça a vantagem de usar modelos misto em análise de grupos de experimentos de cana-de-açúcar.

A Figura 2 mostra a comparação entre o Modelo 11 e o Tradicional. Nesta os valores genotípicos preditos com o Modelo 11 foram correlacionado significativamente com as médias estimadas para TPH com o modelo Tradicional ($\rho = 0,63$, p-valor < 0,001). Há coincidência de ordenamento dos genótipos, porém o melhoramento de cana requer maior eficiência para seleção dos melhores indivíduos, nesse contexto essa associação não é classificada como alta. A associação existente não é suficiente para declarar o Modelo Tradicional uma boa opção para o melhoramento. Há ganho de eficiência quando se utiliza o Modelo 11, e esse ganho não é refletido com o Modelo Tradicional, portanto, o Modelo Tradicional não substitui com eficiência o 11. No ponto de corte, verifica-se que existem seis genótipos (60%) selecionados com o Modelo 11, mas que não foram selecionados com o Modelo Tradicional. Isso ressalta a importância de utilizar modelos mais realista, este possui maior capacidade preditiva e aumenta a eficiência de seleção.

Já a Figura 3 tem-se a comparação do Modelo 1 e 11. A correlação entre os valores genotípicos preditos dos dois modelos foi alta e significativa ($\rho = 0,87$, p-valor < 0,001). Isso sugere que existe concordância de ordenamento entre os dois modelos. E quando se observa o ponto

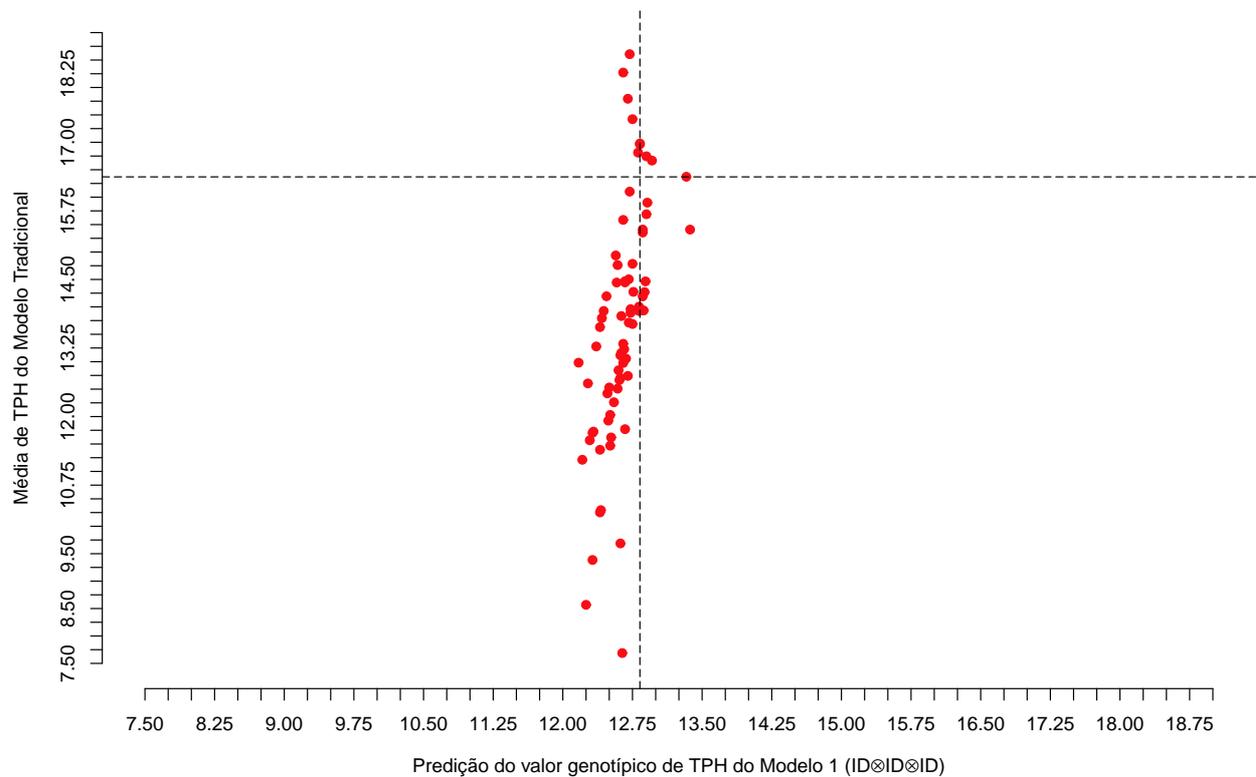


Figura 1: Gráfico das predições do valor genético de TPH do Modelo 1 contra as médias estimadas de TPH do Modelo Tradicional para 74 genótipos ($\rho = 0,72$, $p\text{-valor} < 0,001$), ambos modelos considera a informação dos 44 experimentos. As linhas tracejadas na horizontal e vertical indicam pontos de corte para a seleção dos 10 genótipos superiores de cana-de-açúcar

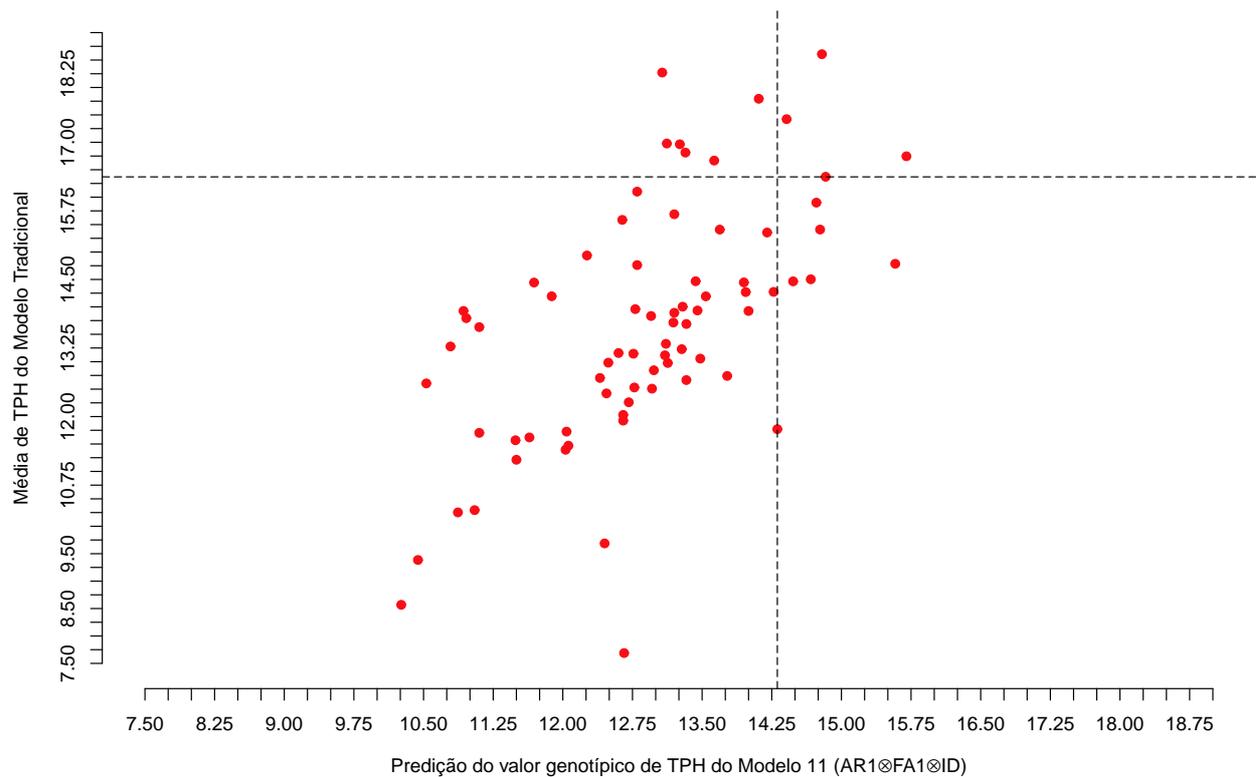


Figura 2: Gráfico das previsões do valor genético de TPH do Modelo 11 contra as médias estimadas de TPH do Modelo Tradicional para 74 genótipos ($\rho = 0,63$, $p\text{-valor} < 0,001$), ambos modelos considera a informação dos 44 experimentos. As linhas tracejadas na horizontal e vertical indicam pontos de corte para a seleção dos 10 genótipos superiores de cana-de-açúcar

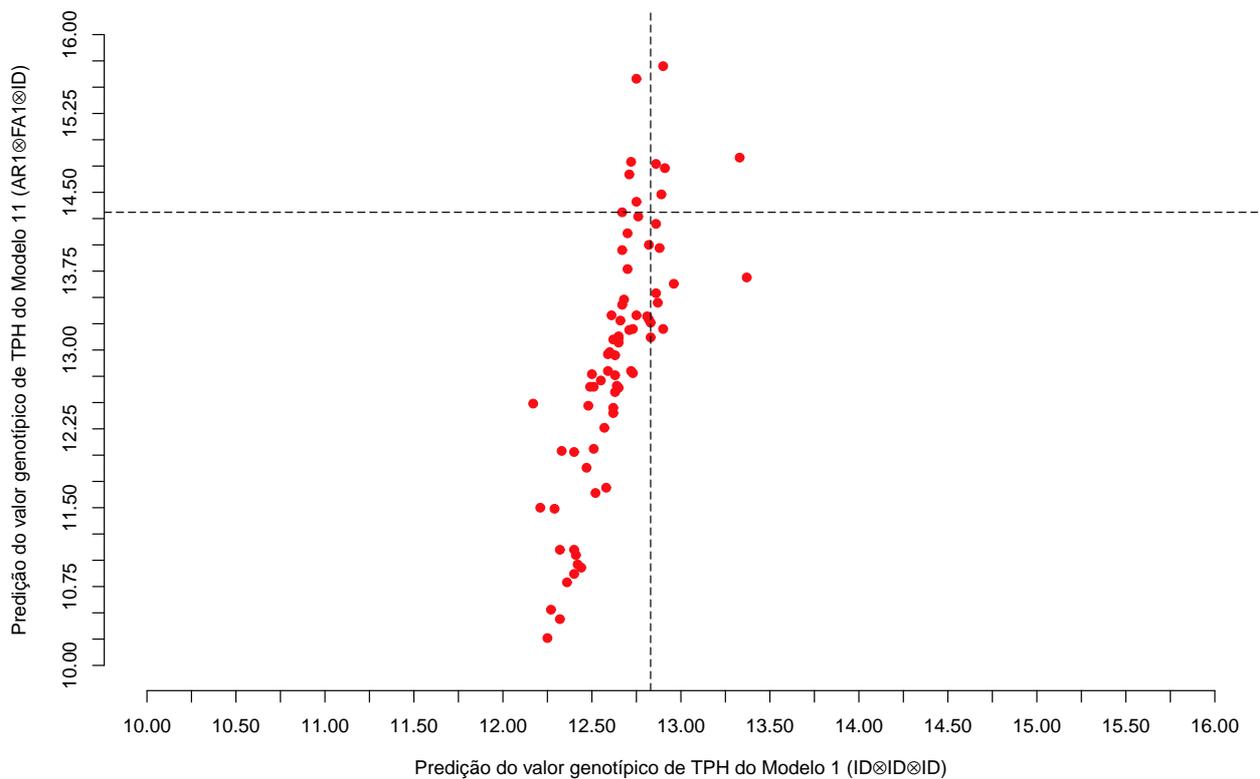


Figura 3: Gráfico das predições do valor genético de TPH do Modelo 1 contra Modelo 11 para 74 genótipos ($\rho = 0,87$, $p\text{-valor} < 0,001$), ambos modelos considera a informação dos 44 experimentos. As linhas tracejadas na horizontal e vertical indicam pontos de corte para a seleção dos 10 genótipos superiores de cana-de-açúcar

de corte, existem cinco genótipos (50%) que foram selecionados com o Modelo 11, mas não foram com o Modelo 1. O argumento a esse resultado é que quando a matriz de VCOV genética é modelada adequadamente, as correlações genéticas entre locais e entre corte são consideradas, o modelo torna-se mais realista e aumenta sua capacidade preditiva.

Ainda nas Figuras 1-3 notam-se que as médias estimadas observadas tiveram maior amplitude de variação do que os valores preditos. Isso é justificado pela propriedade de encolhimento do preditor, que é desejável pois penaliza observações discrepantes.

4 Conclusões

O modelo misto ajustado aqui certamente poderá aumentar a eficiência do processo de seleção de genótipos superiores de cana-de-açúcar, e consequentemente ajudará na identificação, recomendação e liberação de novas variedades em programas de melhoramento.

Referências

5 Bibliografia

- AKAIKE, H. A new look at the statistical model identification. **The Institute of Electrical and Electronics Engineers Transaction and Automatic Control**, Notre Dame, v. 19, p. 716-723, 1974.
- ALLARD, R.W. **Principles of plant breeding**. 2th. ed. New York: John Wiley & Sons, 1999. 254 p.
- ATKIN, F.C.; DIETERS, M.J.; STRINGER, J.K. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. **Theoretical and Applied Genetics**, New York, v. 119, p. 555-565, 2009.
- BALZARINI, M. Applications of mixed models in plant breeding. In: KANG, M.S. **Quantitative genetics, genomics and plant breeding**. New York: CABI Publishing, 2001. p. 353-363.
- BAUER, A.M.; LÉON, J. Multiple-trait breeding values for parental selection in self-pollinating crops. **Theoretical and Applied Genetics**, New York, v. 116, p. 235-242, 2008.
- BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information

from related hybrids. **Crop Science**, Madison, v. 34, p. 20-25, 1994.

BERNARDO, R.; Yu J.M. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, Madison, v. 47, p. 1082-1090, 2007.

BERNARDO, R. **Breeding for quantitative traits in plants**. 2nd.Ed. Minnesota: Stemma Press, 2010, 400p. BOER, M. P.; WRIGHT, D.; FENG, L.; PODLICH, D. W.; LUO, L.; COOPER, M.; EEUWIJK, F.A. van A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. **Genetics**, Bethesda, v. 177, p. 1801-1813, 2007.

BOHN, M.; GROH, S.; KHAIRALLAH, M.M.; HOISINGTON, D.A.; UTZ, H.F.; MELCHINGER, A.E. Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. **Theoretical and Applied Genetics**, New York, v. 103, p. 1059-1067, 2001.

BONNETT, D.G.; REBETZKE, G.J.; SPIELMEYER, W. Strategies for efficient implementation of molecular markers in wheat breeding. **Molecular Breeding**, Berlin, v. 15, p. 75-85, 2005.

BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. **Psychometrika**, Colorado, v. 52, p. 345-370, 1987.

BRESEGHELLO, F.; SORRELLS, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics**, Bethesda, v. 172, p. 1165-1177, 2006.

BROWN, H.; PRESCOTT, R. **Applied Mixed Models in Medicine**. New York: John Wiley & Sons, 2006. 478 p.

BRUMMER E.C. Applying genomics to alfalfa breeding programs. **Crop Science**, Madison, v. 44, p. 1904-1907, 2004.

BURGUEÑO, J.; CROSSA, J.; CORNELIUS, P.L.; TRETOWAN, R.; MCLAREN, G. et al. Modeling additive \times environment and additive \times additive \times environment using genetic covariances of relatives of wheat genotypes. **Crop Science**, Madison, v. 43, p. 311-320, 2007.

BURGUEÑO, J.; CROSSA, J.; CORNELIUS, P.L.; YANG, R.C. Using factor analytic models for joining environments and genotypes without crossover genotype environment interaction. **Crop Science**, Madison, v. 48, p. 1291-1305, 2008.

BURGUEÑO, J.; CROSSA, J.; COTES, J.M.; VICENTE, F.S.; DAS, B. Prediction assessment of linear mixed models for multienvironment trials. **Crop Science**, Madison, v. 51, p. 944-954, 2011.

BURGUEÑO, J.; DE LOS CAMPOS, G.; WEIGEL, K.; CROSSA, J. Genomic prediction of breeding values when modeling genotype environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, p. 707-719, 2012.

BURNHAM, K.P.; ANDERSON, D.R. Multimodel inference understanding AIC and BIC in model selection. **Sociological methods & research**, London, v. 33, p. 261-304, 2004.

CARLSON, J.P.; CHRISTIAN, L.L.; ROTHSCCHILD, M.F. et al. An evaluation of four procedures to rank centrally tested boars. **Journal of Animal Science**, Albany, v. 59, p. 934-940, 1984.

CHRISTENSEN, R.; PEARSON, L. M. Johnson, W. Case-Deletion Diagnostics for Mixed Models. **Technometrics**, Alexandria, v. 34, p. 38-45, 1992.

COLLARD, B.C.; MACKILL, D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. **Philosophical Transactions of the Royal Society B: Biological Sciences**, London, v. 363, p. 557-572, 2008.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira - Cana-de-açúcar: Safra 2012/2013, Segundo levantamento** - Agosto de 2012. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 09 out. 2012.

COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, New York, v. 88, p. 561-572, 1994.

CROSSA, J. Statistical analyses of multilocation trials. **Advances in agronomy**, Mexico, v. 44, p. 55-85, 1990.

CROSSA, J.; FRANCO, J. Statistical methods for classifying genotypes. **Euphytica**, Wageningen, v. 137, p. 19-37, 2004.

CROSSA, J.; YANG, R.C.; CORNELIUS, P.L. Studying crossover genotype \times environment interaction using linear-bilinear models and mixed models. **Journal of agricultural, biological, and environmental statistics**, Alexandria, v. 9, p. 362-380, 2004.

CROSSA, J.; BURGUEÑO, J.; CORNELIUS, P.L.; MCLAREN, G.; TRETOWAN, R.; KRISHNAMACHARI, A. Modeling genotype \times environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, Madison, v. 46, p. 1722-1733, 2006.

CROSSA, J.; BURGUEÑO, J.; DREISIGACKER, S.; VARGAS, M.; HERRERA-FOESSEL, S.A.; LILLEMO, M.; SINGH, R.P.; TRETOWAN, R.; WARBURTON, M.; FRANCO, J.; REYNOLDS, M.; CROUCH, J. H.; ORTIZ, R. Association analysis of historical bread wheat

germplasm using additive genetic covariance of relatives and population structure. **Genetics**, Bethesda, v. 177, p. 1889-1913, 2007.

CROSSA, J.; DE LOS CAMPOS, G.; PÉREZ, P.; GIANOLA, D.; BURGUEÑO, J.; ARAUS, J.L.; MAKUMBI, D.; SINGH, R.P.; DREISIGACKER, S.; YAN, J.; ARIEF, V.; BANZIGER, M.; BRAUN, H.J. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, Bethesda, v.186, p. 713-724, 2010.

DEKKERS J.C.M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, New York, v. 3, p. 22-32, 2002.

DEMPFLE, L. Problems in the Use of the Relationship Matrix in Animal Breeding. In: GIANOLA, D; HAMMOND, K. **Advances in statistical methods for genetic improvement of livestock**. New York: Springer, 1990. p. 454-474.

DEMPSTER A.; LAIRD, N.; RUBIN, D. Maximum likelihood from incomplete data via the EM Algorithm. **Journal of the Royal Statistical Society**. Series B, Oxford, v. 39, p. 1-38, 1977.

DIGGLE, P.J. An Approach to the Analysis of Repeated Measurements. **Biometrics**, Washington, v. 44, p. 959-971, 1988.

DUARTE, J.B; VENCOVSKY, R. **Interação genótipos × ambientes: Uma Introdução à Análise AMMI**. Série Monografias. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p.

EBERHART, S.T.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, p. 36-40, 1966.

ECKERMANN, P.J.; VERBYLA, A.P.; CULLIS, B.R.; THOMPSON, R. The abalysis of quantitative traits in wheat mapping populations. **Australian Journal of Agricultural Research**, Melbourne, v. 52, p. 1195-1206, 2001.

FALCONER, D.S. The problem of environment and selection. **American Naturalist**, Chicago, v. 86, p. 293-298, 1952.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th. Ed. Essex, UK: Longman, 1996. 464 p.

FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Collingwood, v. 14, p. 742-754, 1963.

FISHER, R.A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. **Transactions of the Royal Society of Edinburgh**, Edinburgh, v. 52, p. 399-433, 1918.

FISHER, R.A.; MACKENZIE, W.A. Studies in crop variation. II. The manurial response of different potato varieties. **The Journal of Agricultural Science**, Cambridge, v. 13, p. 311-320, 1923.

FLINT-GARCIA, S.A.; DARRAH, L.L.; McMULLEN, M.D.; HIBBARD, B.E. Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. **Theoretical and Applied Genetics**, New York, v. 107, p. 1331-1336, 2003.

FURLANI, R.C.M.; MORAES M.L.T.; RESENDE, M.D.V.; FURLANI, E. Jr.; GONÇALVES, P.; FILHO, W.V.V.; PAIVA, J.R. Estimation of variance components and prediction of breeding values in rubber tree breeding using the REML/BLUP procedure. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, p. 271-276, 2005.

GALWEY, N.W. **Introduction to mixed modelling: beyond regression and analysis of variance**. Chichester: Wiley, 2007. 376 p.

GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v. 44, p. 705-715, 1988.

GAUCH, H.G. **Statistical analysis of regional yield trials: AMMI analysis of factorial designs**. Amsterdam: Elsevier, 1992. 172 p.

GILMOUR, A.R.; GOGEL, B.J.; CULLIS, B.R.; THOMPSON, R. ASReml User Guide Release 3.0, 2009. Hernal Hempstead: VSN International. Disponível em: <<http://www.vsnl.co.uk>>. Acessado em: 01 Jun. 2011.

HENDERSON C.R.; KEMPTHORNE, O.; SEARLE, S.R.; Von KROSIGK, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Washington, v. 13, p. 192-218, 1959.

HENDERSON, C.R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, Washington, v. 31, p. 423-447, 1975.

HENDERSON, C.R. A simple method for computing the inverse of a numerator relationship Matrix Used in prediction of breeding values. **Biometrics**, Washington, v. 32, p. 69-83, 1976.

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

HENDERSON, C.R. Estimation of Variances in Animal Model and Reduced Animal Model for Single Traits and Single Records. **Journal of Animal Science**, Albany, v. 69, p. 1394-1402, 1986.

HOSPITAL, F. Challenges for effective marker-assisted selection in plants. **Genetica**, Gravenhage, v. 136, p. 303-310, 2009.

HUANG, X.; WEI, X.; SANG, T.; ZHAO, Q.; FENG, Q.; ZHAO, Y. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. **Nature Genetics**, New York, v. 42, p. 961-967, 2010.

HUNTER, R.B. ASTA approach on minimum distance. In: WILKINSON, D. (ed.) **44th Annual Corn Sorghum Industrial Research Conference**. Washington: Am Seed Trade Assoc, 1989. p. 193-195.

KELLY, A.M.; SMITH, A.B.; ECCLESTON, J.A.; CULLIS, B.R. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. **Crop Science**, Madison, v. 47, p. 1063-1070, 2007.

KEMPTHORNE, O. **An introduction to genetic statistics**. Iowa: The Iowa State University, 1969. 545 p.

KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, Cambridge, v. 103, p. 123-135, 1984.

KENNEDY, B.W.; MOXLEY, J.E. Comparison of genetic group and relationship methods for mixed model sire evaluation. **Journal of Dairy Science**, Lancaster, v. 58, p. 1507-1514, 1975.

KERR, R.J.; LI, L.; TIER, B.; DUTKOWSKI, G.W.; McRAE, T.A. Use of the numerator relationship matrix in genetic analysis of autopolyploid species. **Theoretical and Applied Genetics**, New York, v. 124, p. 1271-1282, 2012.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, Bethesda, v. 124, p. 743-756, 1990.

LANGE, K. **Mathematical and Statistical Methods for Genetic Analysis**. New York: Springer-Verlag, 1997. 367 p.

LATOUR, D.; LITTELL, R. **Advanced general linear models with an emphasis on mixed models**: course notes. Cary: Statistical Analysis System Institute, 1996. 614 p.

LINDSEY, J.K. **Models for Repeated Measurements**. 2th. ed. Oxford University Press, 1999. 536 p.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980 p.

MACKILL D.J.; NGUYEN H.T; ZHANG, J. Use of molecular markers in plant improvement programs for rainfed lowland rice. **Field Crops Research**, Amsterdam, v. 64, p. 177-185, 1999.

MALÉCOT, G. **The Mathematics of Heredity**. San Francisco: Freeman, 1969. 88 p.

MARGARIDO, G.R.A. **Mapeamento de QTLs em múltiplos caracteres e ambientes**

em cruzamento comercial de cana-de-açúcar usando modelos mistos. 2011. 107 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2011.

MARIOTTI, J.A. Experiencias de seleccion clonal em cana de azúcar en la provincia de Jujuy: Repetibilidad y heredabilidad de caracteres de interesse agronómico. **Revista Agronómica Norte Argentina**, Argentina, v.10, p. 61-73, 1973.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. Ed. **Melhoramento de espécies cultivadas.** Viçosa: UFV, 1999. p. 205-252.

McCULLOCH, C.E.; SEARLE, S.R. **Generalized, linear and mixed models.** New York: John Wiley & Sons, 2001. 325 p.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda,, v. 157, p. 1819-1829, 2001.

MEYER, K. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. **Genetics Selection Evolution**, Les Ulis, v. 41, p. 1-11, 2009.

OAKEY, H.; VERBYLA, A.; PITCHFORD, W.; CULLIS, B.; KUCHEL, H. Joint modelling of additive and non-additive genetic line effects in single field trials. **Theoretical and Applied Genetics**, New York, v. 113, p. 809-819, 2006.

OLIVEIRA, R.A.; DAROS, E.; BESPALHOK FILHO, J.C.; ZAMBON, J.L.C.; IDO, O.T.;

WEBER, H.; RESENDE, M.D.V.; ZENI NETO, H. Seleção de famílias de cana-de-açúcar via modelos mistos. **Scientia Agrária**, Curitiba, v. 9, p. 269-274, 2008.

PASTINA, M.M. **Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar usando a abordagem de modelos mistos.** 2010. 89 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2010.

PASTINA, M.M.; MALOSETTI, M.; GAZAFFI, R.; MOLLINARI, M.; MARGARIDO, G.R.; OLIVEIRA, K.M.; PINTO, L.R.; SOUZA, A.P.; van EEUWIJK, F.A.; GARCIA, A.A.F. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. **Theoretical and Applied Genetics**, New York, v. 124, p. 835-849, 2012.

PATTERSON, H.D.; THOMPSON, R. Recovery of interblock information when block sizes are unequal. **Biometrika**, Cambrigde, v. 31, p. 100-109, 1971.

PATTERSON, H.D.; SILVEY, V.; TALBOT, M.; WEATHERUP, S.T.C. Variability of yields of cereal varieties in UK trials. **The Journal of Agricultural Science**, Cambridge, v. 89, p. 239-245, 1977.

PAULER, D.K. The Schwarz Criterion and Related Methods for Normal Linear Models.

Biometrika, Cambridge, v. 85, p. 13-27, 1998.

PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. GenStat for Windows (14th Edition) Introduction. **VSN International**, Hemel Hempstead, 2011.

PIEPHO, H.P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, p. 761-766, 1997.

PIEPHO, H.P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, New York, v. 97, p. 195-201, 1998.

PIEPHO, H.P.; MOHRING, J.; MELCHINGER, A.E. et al. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, p. 209-228, 2008.

PIEPHO, H.P. Ridge regression and extensions for genomewide selection in maize. **Crop Science**, Madison, v. 49, p. 1165-1176, 2009.

POLLAK, E.J.; UFFORD, G.R.; GROSS, S.J. Comparison of alternative models for within-herd genetic evaluation fo beef cattle. **Journal of Animal Science**, Albany, v.45, p. 1010-1014, 1977.

PURBA, A.R.; FLORI, A.; BAUDOUIN, L.; HAMON, S. Prediction of oil palm (*Elaeis guineensis*, Jacq.) agronomic performances using the best linear unbiased predictor (BLUP). **Theoretical and Applied Genetics**, New York, v. 102, p. 787-792, 2001.

RESENDE, M.D.V. **Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes**. Colombo: Embrapa Florestas. Documentos n. 47, 2000. 102 p.

RESENDE, M. D. V.; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Brazilian Journal of Mathematics and Statistics**, São Paulo, v. 22, p. 31-52, 2004.

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

RESENDE, M.D.V. **Genética Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais**. Colombo: EMBRAPA Florestas, 2008, 330p.

SEARLE, S.R. **Linear models**. New York: John Wiley & Sons, 1971. 532 p.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Washington, v. 57, p. 1138 - 1147, 2001.

SMITH, A.; CULLIS, B.R.; THOMPSON, R. Exploring variety environment data using

random effects models with adjustment for spatial field trends: Part 1. **Theory Quantitative genetics, genomics and plant breeding**. CAB International, 2002. 400 p.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, Cambridge, v. 143, p. 449-462, 2005.

SO, Y.; EDWARDS, J.W. Predictive ability assessment of linear mixed models in multi-environment trials in corn (*Zea mays* L). **Crop Science**, Madison, v. 51, p. 542-552, 2011.

SOUZA, E.; SORRELLS, M.E. Pedigree analysis of North American oat cultivars released from 1951 to 1985. **Crop Science**, Madison, v. 29, p. 595-601, 1989.

SOUZA JÚNIOR, C.L. **Melhoramento de espécies de reprodução vegetativa**. Publicação didática, Departamento de Genética, Piracicaba: Esalq/USP, 1995. 41 p.

THOMPSON, R.; CULLIS, B.R.; SMITH, A.B.; GILMOUR, A.R. A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. **Australian and New Zealand Journal of Statistics**, Oxford, v. 45, p. 445-459, 2003.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer, 2009. 592 p.

VERBYLA, A.; ECKERMAN, P.J.; THOMPSON, R.; CULLIS, B. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. **Australian Journal of Agricultural Research**, Melbourne, v. 54, p. 1395-1408, 2003.

WHITE, T.L.; HODGE, G.R. **Predicting Breeding Values with Applications in Forest Tree Improvement**. Kluwer Academic Publishers, Dordrecht, 1989. 363 p.

WEI, X.M.; BORRALHO, M.G. Genetic gains and levels of relatedness from best linear unbiased prediction selection of *Eucalyptus urophylla* for pulp production in southeastern China. **Canadian journal of forest research**, Ottawa, v. 30, p. 1601-1607, 2000.

WOLFINGER, R.D. Covariance structure selection in general mixed linear models. **Communications in Statistics A, Theory and Methods**, Zug, v. 22, p. 1079-1106, 1993.

YPMA, T.J. Historical development of the Newton-Raphson method. **SIAM review**, Philadelphia v. 37, p. 531-551, 1995.

YU, J.; PRESSOIR, G.; BRIGGS, W.; VROH, Bi I.; YAMASAKI, M.; DOEBLEY, J.F.; McMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, p. 203-208, 2006.

ZHE-ZHANG.; LIU, J.; DING, X.; BIJMA, P.; DE KONING; D.J.; ZHANG, Q. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived

relationship matrix. **PLoS One**, Cambridge, v. 5, p. 1-8, 2010.

ZHIWU-ZHANG.; ERSOZ, E.; LAI, C.Q.; TODHUNTER, R.J.; TIWARI, H.K.; GORE, M.A.; BRADBURY, P.J.; YU, J.; ARNETT, D.K.; ORDOVAS, J.M.; BUCKLER, E.S. Mixed linear model approach adapted for genome-wide association studies. **Nature genetics**, New York, v. 42, p. 355-360, 2010.

Tabela 1: Diferentes modelos analisados de matriz de variância-covariância \mathbf{G}

Matrix \mathbf{G}	Modelo	n_{PAR} em \mathbf{G}^\dagger
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$	1) ID \otimes ID \otimes ID	1
	2) DIAG \otimes ID \otimes ID	$(J + 1) - 1$
	3) FA1 \otimes ID \otimes ID	$(2J + 1) - 1$
	4) UNST \otimes ID \otimes ID	$\left[\frac{J(J+1)+2}{2} \right] - 1$
	5) ID \otimes DIAG \otimes ID	$(1 + K) - 1$
	6) DIAG \otimes DIAG \otimes ID	$(J + K) - 1$
	7) FA1 \otimes DIAG \otimes ID	$(2J + K) - 1$
	8) UNST \otimes DIAG \otimes ID	$\left[\frac{J(J+1)+2K}{2} \right] - 1$
	9) ID \otimes AR1 \otimes ID	2
	10) DIAG \otimes AR1 \otimes ID	$(J + 2) - 1$
	11) FA1 \otimes AR1 \otimes ID	$(2J + 2) - 1$
	12) UNST \otimes AR1 \otimes ID	$\left[\frac{J(J+1)+4}{2} \right] - 1$
	13) ID \otimes AR1 _{Het} \otimes ID	$(K + 2) - 1$
	14) DIAG \otimes AR1 _{Het} \otimes ID	$(J + K + 1) - 1$
	15) FA1 \otimes AR1 _{Het} \otimes ID	$(2J + K + 1) - 1$
	16) UNST \otimes AR1 _{Het} \otimes ID	$\left[\frac{J(J+1)+2(K+1)}{2} \right] - 1$
	17) ID \otimes UNST \otimes ID	$\left[\frac{2+K(K+1)}{2} \right] - 1$
	18) DIAG \otimes UNST \otimes ID	$\left[\frac{2J+K(K+1)}{2} \right] - 1$
	19) FA1 \otimes UNST \otimes ID	$\left[\frac{4J+K(K+1)}{2} \right] - 1$
	20) UNST \otimes UNST \otimes ID	$\left[\frac{J(J+1)+K(K+1)}{2} \right] - 1$
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$	21) ID \otimes ID \otimes Kinship	1
	22) DIAG \otimes ID \otimes Kinship	$(J + 1) - 1$
	23) FA1 \otimes ID \otimes Kinship	$(2J + 1) - 1$
	24) UNST \otimes ID \otimes Kinship	$\left[\frac{J(J+1)+2}{2} \right] - 1$
	25) ID \otimes DIAG \otimes Kinship	$(1 + K) - 1$
	26) DIAG \otimes DIAG \otimes Kinship	$(J + K) - 1$
	27) FA1 \otimes DIAG \otimes Kinship	$(2J + K) - 1$
	28) UNST \otimes DIAG \otimes Kinship	$\left[\frac{J(J+1)+2K}{2} \right] - 1$
	29) ID \otimes AR1 \otimes Kinship	2
	30) DIAG \otimes AR1 \otimes Kinship	$(J + 2) - 1$
	31) FA1 \otimes AR1 \otimes Kinship	$(2J + 2) - 1$
	32) UNST \otimes AR1 \otimes Kinship	$\left[\frac{J(J+1)+4}{2} \right] - 1$
	33) ID \otimes AR1 _{Het} \otimes Kinship	$(K + 2) - 1$
	34) DIAG \otimes AR1 _{Het} \otimes Kinship	$(J + K + 1) - 1$
	35) FA1 \otimes AR1 _{Het} \otimes Kinship	$(2J + K + 1) - 1$
	36) UNST \otimes AR1 _{Het} \otimes Kinship	$\left[\frac{J(J+1)+2(K+1)}{2} \right] - 1$
	37) ID \otimes UNST \otimes Kinship	$\left[\frac{2+K(K+1)}{2} \right] - 1$
	38) DIAG \otimes UNST \otimes Kinship	$\left[\frac{2J+K(K+1)}{2} \right] - 1$
	39) FA1 \otimes UNST \otimes Kinship	$\left[\frac{4J+K(K+1)}{2} \right] - 1$
	40) UNST \otimes UNST \otimes Kinship	$\left[\frac{J(J+1)+K(K+1)}{2} \right] - 1$

Modelos (1-20) utilizam o produto direto de matrizes de variância-covariância para experimento e corte, assumindo ausência de parentesco entre os genótipos, \mathbf{A} . Modelos (21-40) utilizam o produto direto de matrizes de variância-covariância para experimento e corte, assumindo parentesco genético entre os genótipos. \dagger : o número de parâmetros para os modelos (1-40), correspondem à soma dos números de parâmetros das matrizes componentes menos o número de restrições para assegurar a identificabilidade do modelo. I é o número de genótipos, J é o número de experimentos (locais) e K o número de cortes.

Tabela 2: Modelos para matriz de variância-covariância \mathbf{G} , e os correspondentes números de parâmetros estimados n_{PAR} , e respectivos valores de AIC (Critério de Informação de Akaike) e BIC (Critério de Informação de Bayseano)

Matrix \mathbf{G}	Modelo	n_{PAR} em \mathbf{G}^\dagger	AIC	BIC
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$	1) ID \otimes ID \otimes ID	$(1 + 1) - 1 = 1$	42473,64	42502,11
	2) DIAG \otimes ID \otimes ID	$(44 + 1) - 1 = 44$	42266,29	42600,77
	3) FA1 \otimes ID \otimes ID	$(88 + 1) - 1 = 88$	41660,41	42308,02
	4) UNST \otimes ID \otimes ID	$(990 + 1) - 1 = 990$	NR	NR
	5) ID \otimes DIAG \otimes ID	$(1 + 5) - 1 = 5$	42448,34	42505,27
	6) DIAG \otimes DIAG \otimes ID	$(44 + 5) - 1 = 48$	42257,09	42620,03
	7) FA1 \otimes DIAG \otimes ID	$(88 + 5) - 1 = 92$	41666,14	42342,22
	8) UNST \otimes DIAG \otimes ID	$(990 + 5) - 1 = 994$	NR	NR
	9) ID \otimes AR1 \otimes ID	$(1 + 2) - 1 = 2$	42298,04	42333,62
	10) DIAG \otimes AR1 \otimes ID	$(44 + 2) - 1 = 45$	41937,97	42279,56
	11) FA1 \otimes AR1 \otimes ID	$(88 + 2) - 1 = 89$	41487,76	42142,49
	12) UNST \otimes AR1 \otimes ID	$(990 + 2) - 1 = 991$	NR	NR
	13) ID \otimes AR1 _{Het} \otimes ID	$(1 + 6) - 1 = 6$	42269,51	42333,56
	14) DIAG \otimes AR1 _{Het} \otimes ID	$(44 + 6) - 1 = 49$	41922,11	42292,17
	15) FA1 \otimes AR1 _{Het} \otimes ID	$(88 + 6 + 1) - 1 = 94$	41489,18	42172,37
	16) UNST \otimes AR1 _{Het} \otimes ID	$(990 + 6) - 1 = 995$	NR	NR
	17) ID \otimes UNST \otimes ID	$(1 + 15) - 1 = 15$	42241,50	42369,60
	18) DIAG \otimes UNST \otimes ID	$(44 + 15) - 1 = 58$	41908,16	42342,27
	19) FA1 \otimes UNST \otimes ID	$(88 + 15) - 1 = 102$	41426,59	42173,83
	20) UNST \otimes UNST \otimes ID	$(990 + 15) - 1 = 1004$	NR	NR
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$	21) ID \otimes ID \otimes Kinship	$(1 + 1) - 1 = 1$	42177,28	42205,74
	22) DIAG \otimes ID \otimes Kinship	$(44 + 1) - 1 = 44$	42057,86	42392,34
	23) FA1 \otimes ID \otimes Kinship	$(88 + 1) - 1 = 88$	NR	NR
	24) UNST \otimes ID \otimes Kinship	$(990 + 1) - 1 = 990$	NR	NR
	25) ID \otimes DIAG \otimes Kinship	$(1 + 5) - 1 = 5$	42164,91	42221,85
	26) DIAG \otimes DIAG \otimes Kinship	$(44 + 5) - 1 = 48$	42049,15	42412,09
	27) FA1 \otimes DIAG \otimes Kinship	$(88 + 5) - 1 = 92$	NR	NR
	28) UNST \otimes DIAG \otimes Kinship	$(990 + 5) - 1 = 994$	NR	NR
	29) ID \otimes AR1 \otimes Kinship	$(1 + 2) - 1 = 2$	41850,25	41885,83
	30) DIAG \otimes AR1 \otimes Kinship	$(44 + 2) - 1 = 45$	41610,50	41952,09
	31) FA1 \otimes AR1 \otimes Kinship	$(88 + 2) - 1 = 89$	NR	NR
	32) UNST \otimes AR1 \otimes Kinship	$(990 + 2) - 1 = 991$	NR	NR
	33) ID \otimes AR1 _{Het} \otimes Kinship	$(1 + 6) - 1 = 6$	41834,69	41898,74
	34) DIAG \otimes AR1 _{Het} \otimes Kinship	$(44 + 6) - 1 = 49$	41592,66	41962,73
	35) FA1 \otimes AR1 _{Het} \otimes Kinship	$(88 + 6) - 1 = 93$	NR	NR
	36) UNST \otimes AR1 _{Het} \otimes Kinship	$(990 + 6) - 1 = 995$	NR	NR
	37) ID \otimes UNST \otimes Kinship	$(1 + 15) - 1 = 15$	41811,31	41939,41
	38) DIAG \otimes UNST \otimes Kinship	$(44 + 15) - 1 = 58$	41582,15	42016,26
	39) FA1 \otimes UNST \otimes Kinship	$(88 + 15) - 1 = 102$	NR	NR
	40) UNST \otimes UNST \otimes Kinship	$(990 + 15) - 1 = 1004$	NR	NR

\mathbf{G} : é a matriz de VCOV genética; ID: Independente; DIAG: Diagonal; FA1: Fator analítico de primeira ordem; AR1: Auto regressiva de primeira ordem; AR1_{Het}: Auto regressiva de primeira ordem (heterogêneo) e UNST: não-estruturado. Os menores valores dos critérios AIC e BIC estão destacados em negrito, indicam o melhor modelo. †: número total da matriz de VCOV do modelo.