

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Análise de experimentos em látice quadrado no melhoramento vegetal
utilizando modelos mistos**

Viviane Panariello Paulenas

Dissertação apresentada para obtenção do título de
Mestra em Ciências. Área de concentração: Estatística
e Experimentação Agronômica

**Piracicaba
2016**

Viviane Panariello Paulenas
Engenheira Agrônoma

Análise de experimentos em látice quadrado no melhoramento vegetal utilizando modelos mistos

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof^a. Dr^a. **SÔNIA MARIA DE STEFANO PIEDADE**

Dissertação apresentada para obtenção do título de Mestra em Ciências. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba
2016

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Paulenas, Viviane Panariello

Análise de experimentos em látice quadrado no melhoramento vegetal utilizando modelos mistos / Viviane Panariello Paulenas. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.

77 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Experimentos em látice 2. Modelos mistos 3. Máxima verossimilhança restrita
4. EBLUP 5. *Zea mays* L. 6. Correlação I. Título

CDD 633.15
P326a

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

À minha mãe, **Eliana**, exemplo de carinho,
perseverança e competência. Ao meu pai, **Bruno**,
por me ensinar o valor da integridade, da humildade e do
esforço. A ambos pelo amor incondicional, por acreditarem
sempre em meus sonhos e me apoiarem em todas as decisões.

Com Amor Dedicado

AGRADECIMENTOS

A Deus pelo dom da vida, por me iluminar e pela oportunidade de concluir mais esta nova etapa.

Aos meus pais, Bruno e Eliana, às minhas irmãs Mariana e Paula, à minha querida avó Carmen Aparecida (Zita), aos meus padrinhos Angela e Sidney e a toda a minha família, pelo incentivo, pelos abraços calorosos, pelo amor e pela compreensão nos momentos de ausência. Ao meu avô Paulo (*in memoriam*), que infelizmente não pode acompanhar esta conquista até o fim, mas que ficaria muito feliz em vê-la concluída.

À Escola Superior de Agricultura “Luiz de Queiroz”, seus professores e funcionários, pela estrutura e oportunidade oferecidas e por tantos anos de acolhimento.

À querida Professora Sônia Maria De Stefano Piedade pela orientação, confiança, pela amizade e por sempre irradiar seu amor pela nossa Gloriosa ESALQ por onde quer que passe.

À querida Professora Renata Alcarde Sermarini pelas sugestões, pelos esclarecimentos prestados e principalmente pela atenção.

Aos Professores do Programa de Pós-Graduação em Agronomia - Estatística e Experimentação Agronômica, por compartilharem seus conhecimentos.

Ao Prof. Cláudio Lopes de Souza Júnior do Departamento de Genética pelas sugestões e pela concessão dos dados experimentais sem os quais eu não teria a oportunidade da aplicação. À doutoranda Melina Andrade pela intermediação.

Às secretárias Solange Sabadin e Luciane Brajão por todas as informações e esclarecimentos oferecidos, além de todo apoio e atenção.

Aos funcionários do Departamento de Ciências Exatas, Mayara, Eduardo, Jorge e Rosni pelo suporte e à Tina, por manter a ordem e a limpeza dos ambientes de estudo.

Aos colegas de turma César Jotta pelo papo descontraído, ao Christopher Silva de Pádua por me ajudar a desvendar os segredos do \LaTeX , R em geral, pela amizade e companhia nas gordices, à Gislaine Pereira pelas conversas e incentivo, à Luiza Barros pelo incentivo na fase final da dissertação, à Maíra Faretto pela amizade e por me ajudar a entender vários conceitos estatísticos e por último, mas não menos importante, meu amigo Vinícius Menarin (Polacão) pelas valiosas ajudas em R, \LaTeX , Excel e pela amizade.

À querida Alessandra dos Santos pela sugestão do tema do trabalho, pela amizade e competência. Mesmo estando na Austrália, não deixou de me auxiliar na realização deste trabalho. À Natália Martins que me deu todo o apoio para que eu regressasse aos estudos, à Natalie Rondinel pelo material e valiosa ajuda.

Ao amigo de longa data Victor Augusto Forti por me mostrar os encantos da Estatística na Agronomia.

Aos funcionários da Divisão de Biblioteca e Documentação da ESALQ pela estrutura e por todo o apoio nas buscas bibliográficas em especial à Thais Cristiane Campos de Moraes pelos sorrisos amistosos e pela constante disposição em me ajudar.

Aos amigos do G9, André Colaço, Camila Ortiz, Diego Chiou, Erik Barreto Jr., Heitor Arakawa, Larissa Pagliuca, Leonardo Bordin e Mônica Georgino pelos churrascos, momentos de descontração e incentivo constante.

Às meninas do F5, Alessandra e Érica da Paz, Marisol Bandeira, Mônica Georgino e Natália Lovadino pelos bons momentos juntas e pelo carinho.

Ao Alcides Júnior que esteve presente uma boa parte do caminho mas que nunca deixou de me incentivar.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pela concessão da bolsa de estudos.

O meu MUITO OBRIGADA, vocês são parte disso!

“Queridos jovens: se queremos que nossa vida tenha realmente sentido e plenitude, digo a cada um e a cada uma de vocês: ‘bote fé’ e a vida terá um sabor novo, terá uma bússola que indica a direção; ‘bote esperança’ e todos os seus dias serão iluminados e o seu horizonte já não será escuro, mas luminoso; ‘bote amor’ e a sua existência será como uma casa construída sobre a rocha, o seu caminho será alegre, porque encontrará muitos amigos que caminham com você.

‘Bote fé’, ‘bote esperança’, ‘bote amor’! ”

Papa Francisco

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE ILUSTRAÇÕES	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Melhoramento genético vegetal da cultura do milho	21
2.2 Delineamentos em blocos incompletos	22
2.3 Delineamentos látices quadrados ou reticulado quadrado	23
2.3.1 Látice quadrado balanceado	24
2.3.2 Látice quadrado parcialmente balanceado	24
2.3.2.1 Construção do delineamento látice quadrado	25
2.3.3 Análises estatísticas convencionais para delineamentos em látice quadrado	25
2.3.4 Análises individuais e conjuntas	26
2.4 Modelos lineares mistos	27
2.4.1 Estimação dos componentes de variância	31
2.4.1.1 Método dos mínimos quadrados - ANOVA	33
2.4.1.2 Método da Máxima Verossimilhança	33
2.4.1.3 Método da Máxima Verossimilhança Restrita	34
2.4.1.4 Predição do efeitos aleatórios	35
2.4.2 Seleção de modelos	36
2.4.2.1 Teste da Razão de Verossimilhança	36
2.5 Herdabilidade	36
3 MATERIAL E MÉTODOS	39
3.1 Material	39
3.2 Métodos	40
3.3 Especificação dos modelos	41
3.3.1 Análise convencional como látice quadrado	42
3.3.1.1 Análises individuais	42
3.3.1.2 Análises conjuntas	42
3.3.2 Análise considerando repetição como um bloco completo	43
3.3.2.1 Análises individuais	43
3.3.2.2 Análise conjunta	44
4 RESULTADOS E DISCUSSÃO	45
4.1 Análise descritiva	45
4.2 Coeficiente de determinação fenotípico entre valores observados e preditos	46

4.3	Estimação dos componentes de variância	46
4.3.1	Considerando a estrutura do látice quadrado	47
4.3.2	Considerando cada repetição como um bloco completo	48
4.4	Diagnósticos	49
4.5	Ordenação e seleção das progênies obtidos a partir da predição dos efeitos aleatórios	56
4.5.1	Correlação entre os EBLUPs obtidos por ambiente nas diferentes estruturas de análise	59
5	CONSIDERAÇÕES FINAIS	63
	REFERÊNCIAS	65
	ANEXOS	71

RESUMO

Análise de experimentos em látice quadrado no melhoramento vegetal utilizando modelos mistos

Experimentos conduzidos no delineamento látice ou reticulado são bastante comuns no melhoramento genético vegetal em que diversos materiais genéticos são comparados, principalmente nas etapas iniciais do programa, visando explorar com maior intensidade a variabilidade genética disponível. Em situações de restrições espaciais e financeiras estes delineamentos se destacam por permitir a comparação de todas as progênies em teste estando ou não instaladas no mesmo bloco. O objetivo do trabalho foi a avaliação de testes de progênies de milho (*Zea mays* L.), em diferentes ambientes para o caráter produção de grãos em t.ha⁻¹. Duzentas e cinquenta e seis progênies foram instaladas em 4 estações experimentais do município de Piracicaba em diferentes anos agrícolas. Os dados de produção de grãos obtidos pelos diferentes ambientes foram analisados de forma individual e conjunta, a fim de verificar presença da interação genótipo × ambiente. O delineamento usado foi, portanto, o látice quadrado 16 × 16, com duas repetições em cada local. Duas abordagens experimentais foram confrontadas, considerando a estrutura de blocos incompletos parcialmente balanceados do látice e a outra em que cada repetição do látice foi analisada como se fosse um bloco completo. Uma maneira de se analisar estruturas experimentais como esta é utilizando modelos mistos, por meio da inclusão de fatores de efeito aleatório e, fazendo o uso da máxima verossimilhança restrita (REML) para estimar os componentes de variância associados a tais fatores com um menor viés. Além dos componentes de variância, os EBLUPs (melhores preditores lineares não viesados empíricos) também foram calculados e a partir deles foi verificada a correlação entre os diferentes ambientes, e a porcentagem de progênies selecionadas comparando-se os resultados obtidos pelas duas abordagens do conjunto de dados. Análises estatísticas foram implementadas utilizando o software gratuito R, com o pacote estatístico lme4.

Palavras-chave: Experimentos em látice; Modelos mistos; Máxima verossimilhança restrita; EBLUP; *Zea mays* L.; Correlação

ABSTRACT

Analysis of experiments in square lattice in plant breeding using mixed models

Experiments conducted in the lattice design are quite common in plant breeding in which several genetic materials are compared, especially in the early stages of the program, aiming to explore more intensively the genetic variability available. In situations of space and financial constraints these designs stand out for allowing the comparison of all progenies being tested whether or not installed in the same block. The aim of the study was the evaluation of maize (*Zea mays* L.) progeny tests in different environments for grain yield in $t.ha^{-1}$. Two hundred and fifty six progenies were tested in four experimental stations in the city of Piracicaba, in different agricultural years. Grain production data obtained by different environments were analyzed individually and jointly in order to verify the presence of genotype \times environment interaction. Therefore, the square lattice design with dimension 16×16 was used with two replications in each location. Two experimental approaches were compared, considering the partially balanced incomplete block structure of the lattice and the other in each repetition of the lattice was analyzed as if it were a complete block. One way to analyze experimental structures like this is with the use of mixed models, by adding random effect factors, and by making use of the restricted maximum likelihood (REML) for estimating the variance components associated with such factors with less bias. Besides the variance components, EBLUPs (empirical best linear predictor unbiased) were also calculated and from them was checked the correlation between the different environments, and the percentage of selected progenies comparing the results obtained by the two assembly approaches data. Statistical analyzes were implemented for the open-source software R, using the statistical package lme4.

Keywords: Square lattice design; Mixed models; Restricted maximum likelihood (REML); EBLUP; *Zea mays* L.; Correlation

LISTA DE ILUSTRAÇÕES

Figura 1 – Cruzamento das linhagens para obtenção das progênies	39
Figura 2 – Resíduos condicionais das análises individuais pelo látice	50
Figura 3 – Resíduos condicionais das análises individuais por blocos completos	51
Figura 4 – arrumar legenda	52
Figura 5 – arrumar legenda	53
Figura 6 – HNP das análises individuais pelo látice	54
Figura 7 – HNP das análises individuais por blocos completos	55
Figura 8 – Correlação de Spearman das análises individuais pelo látice	60
Figura 9 – Correlação de Spearman da análise conjunta pelo látice	60
Figura 10 – Correlação de Spearman das análises individuais por blocos completos	62
Figura 11 – Correlação de Spearman da análise conjunta por blocos completos	62

LISTA DE TABELAS

Tabela 1 – Estações experimentais e anos em que os experimentos foram instalados . . .	39
Tabela 2 – Análise exploratória da produtividade em t.ha ⁻¹ por ambiente	45
Tabela 3 – Coeficiente de determinação fenotípico entre os valores observados e os preditos pelos modelos nas análises individuais e conjuntas	46
Tabela 4 – Componentes de variância das análises individuais - látice	47
Tabela 5 – Estimativa dos componentes de variância da análise conjunta - látice	47
Tabela 6 – Componentes de variância das análises individuais - blocos completos	48
Tabela 7 – Estimativa dos componentes de variância da análise conjunta - blocos completos	48
Tabela 8 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas nas análises individuais e conjunta como látice	57
Tabela 9 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas nas análises individuais e conjunta como blocos casualizados completos	58
Tabela 10 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas pelas duas abordagens de análise individuais	59

1 INTRODUÇÃO

Na experimentação agronômica, quando pretende-se controlar a heterogeneidade ambiental, o delineamento em blocos casualizados completos é geralmente adotado. No entanto, quando o número de tratamentos a serem avaliados, ou o tamanho das parcelas é muito grande, os blocos completos podem tornar-se heterogêneos, e deste modo, este tipo de delineamento perde sua eficiência, pois a pressuposição de homogeneidade dentro dos blocos nem sempre é atendida. Uma alternativa são os delineamentos em blocos incompletos, dentre eles, os látices quadrados, conhecidos também por reticulados quadrados.

Nos programas de melhoramento genético vegetal, especialmente do milho, estes delineamentos têm-se destacado por serem delineamentos eficientes para testar grande número de tratamentos (híbridos, linhagens, progênies etc.), principalmente nas etapas iniciais do programa, visando explorar com maior intensidade a variabilidade genética disponível. Nesse caso, em geral, o interesse é tratar o efeito de tratamento como aleatório. Considerar blocos como fixos ou aleatórios, no entanto, acarreta certa polêmica, mas segundo (HINKELMANN; KEMPTHORNE, 2005), em ensaios de campo, há sempre uma justificativa clara de casualização para se tomar os blocos como aleatórios. É de grande interesse para os melhoristas a simplificação dos métodos analíticos já existentes, sem perda da precisão nas estimativas dos componentes de variância.

Recentemente a abordagem dos modelos mistos tornou-se popular pela facilidade com que dados incompletos podem ser trabalhados, pela capacidade de modelar adequadamente a variação do erro dentro do experimento e pela capacidade de lidar com diferentes combinações de efeitos que podem ser fixos ou aleatórios. Quando a quantidade de material genético é grande, como no caso do melhoramento do milho, em geral, o interesse é tratar o efeito de tratamento como aleatório. Considerar blocos como fixos ou aleatórios, no entanto, acarreta certa polêmica (TOMÉ; BUENO-FILHO; FERREIRA, 2002), mas segundo Hinkelmann e Kempthorne (2005), em ensaios de campo, há sempre uma justificativa clara de casualização para se tomar os blocos como aleatórios.

O uso do método da máxima verossimilhança restrita (REML) na abordagem dos modelos mistos modificou a estimação dos componentes de variância e de parâmetros genéticos (SEARLE; CASELLA; MCCULLOCH, 2006) quando comparado com o método dos mínimos quadrados. A predição de valores genéticos usando o BLUP (melhor preditor linear não viesado) assume que os componentes de variância são conhecidos. Por meio do método REML, os componentes de variância podem ser estimados diretamente do conjunto de dados, assim como as variâncias dos efeitos assumidos como aleatórios (FRITSCHÉ-NETO, 2008).

Quando o intuito do melhorista é identificar os genótipos mais promissores para as diferentes condições avaliadas, é recomendado que o experimento seja conduzido em ampla

faixa de condições ambientais. Isto acontece devido à possibilidade de se analisar a interação dos genótipos com os ambientes que, ao ser quantificada corretamente, pode ser revertida em favor da maximização das produtividades da cultura (LEITE, 2011).

Uma outra abordagem possível quando se instala um experimento em látice quadrado é a possibilidade de realizar a análise em blocos casualizados, na qual cada repetição do látice é considerada com um bloco casualizado completo. Essa modalidade de análise já foi testada por Paterniani (1968), Silva (1997), Rios (1997) e Cecon (1992), apresentando precisão satisfatória nos testes estatísticos e nas estimativas dos componentes de variância.

A produção de grãos de milho é o principal caráter de interesse, tanto para melhoristas como para agricultores. Trata-se de um caráter quantitativo, controlado por um grande número de locos (JUGENHEIMER, 1976; PATERNIANI; VIEGAS, 1987), que possuem pequeno efeito individual sobre o fenótipo (COMSTOCK, 1978), fazendo com que sua expressão sofra acentuado efeito do ambiente (MORO *et al.*, 2012).

Os objetivos deste trabalho são, portanto:

- i) fazer uma revisão das possíveis análises de experimentos em látice quadrado e da análise deles via metodologia de modelos mistos com enfoque no melhoramento genético;
- ii) estimar os componentes de variância e, conseqüentemente, predizer os EBLUPs para o caráter produção de grãos via método da máxima verossimilhança restrita por meio de duas abordagens, considerando a estrutura do látice quadrado e tomando-se cada repetição do látice como se fosse um bloco completo;
- iii) comparar a ordenação dos EBLUPs das progênies entre ambientes e pelas duas abordagens com a ajuda de gráficos de correlação e da contagem das melhores progênies selecionadas comparando-se os ambientes entre si.

2 REVISÃO BIBLIOGRÁFICA

2.1 Melhoramento genético vegetal da cultura do milho

O milho (*Zea mays L.*) em função de seu potencial produtivo, composição química e valor nutritivo, constitui-se em um dos mais importantes cereais cultivados e consumidos no mundo. Devido à sua multiplicidade de aplicações, quer na alimentação humana quer na alimentação animal, assume relevante papel socioeconômico, além de ser indispensável matéria-prima impulsionadora de diversificados complexos agroindustriais (FANCELLI; DOURADO-NETO, 2004). É uma espécie da família das gramíneas e é um dos cereais mais cultivados no planeta. A cultura está disseminada numa vasta região do globo, em altitudes que vão desde o nível do mar até 3 mil metros.

Segundo Paterniani e Viegas (1987), a mais antiga espiga de milho foi encontrada no vale do Tehucan, na região onde hoje se localiza o México, datada de 7.000 a.C. O Teosinte ou “alimento dos deuses”, como era chamado pelos maias, deu origem ao milho por meio de um processo de seleção artificial (feito pelo homem). Ao longo do tempo, o homem promoveu uma crescente domesticação do milho por meio da seleção visual no campo, considerando importantes características, tais como produtividade, resistência a doenças e capacidade de adaptação, dentre outras.

A partir do início do século XX, vários programas de melhoramento genético usando bases científicas foram iniciados. O desenvolvimento de linhagens, oriundas do processo de autofecundação das plantas de milho por várias gerações, e do vigor híbrido, ou heterose, foram os responsáveis pelo impulso que o melhoramento genético convencional tomou no início do século passado. Esse conhecimento permitiu que os programas de melhoramento conseguissem introduzir novas características ao milho como resistência a doenças e pragas, maior resposta às práticas de manejo e menor tombamento e quebraimento de plantas. Esse conjunto de melhorias fez com que o milho se adaptasse a diferentes regiões, condições de clima, solo e finalidade de uso.

Os efeitos do melhoramento podem ser vistos também no aumento da produtividade média nacional. De 2622 kg.ha⁻¹ na safra 94/95, a produtividade média nacional saltou para 5396 kg.ha⁻¹ em 2014/2015. São 2774 kg.ha⁻¹ de aumento na produtividade em apenas 20 anos. Considerando que 50% desse aumento foi devido ao melhoramento genético e os outros 50% devido às melhorias nas condições de cultivo, são, em média, 69,3 kg.ha⁻¹.ano⁻¹ de contribuição do melhoramento, de acordo com dados da CONAB (COMPANHIA NACIONAL DE ABASTECIMENTO, 2016). Nos últimos anos, a área de cultivo de milho no Brasil tem sido de, aproximadamente, 10 milhões de hectares.

O milho é uma das culturas mais importantes mundialmente, seja do ponto de vista econômico, seja do ponto de vista social. Destaca-se por ser o grão mais produzido no mundo: segundo informações do UNITED STATES DEPARTMENT OF AGRICULTURE - USDA (2016), a produção mundial deste cereal deverá atingir 989,83 milhões de toneladas na safra agrícola 2015/16. É importante destacar que, entre as safras 2004/05 e 2014/15, a produção mundial de milho registrou um crescimento de 38,4%. Este forte incremento ocorreu para atender à necessidade de abastecer a demanda, que registrou uma expansão de 41%, no mesmo período. O milho também se destaca como o grão mais consumido do mundo: estima-se que, na safra 2014/15, seu consumo alcance 971,2 milhões de toneladas (SOLOGUREN, 2015).

2.2 Delineamentos em blocos incompletos

Um dos princípios básicos da experimentação é agrupar as unidades experimentais em blocos com parcelas homogêneas para aumentar a precisão e para fazer comparações sob condições mais uniformes possíveis. Isto é geralmente verdade, no entanto, a precisão aumenta à medida que o tamanho do bloco diminui, de modo que os blocos pequenos são preferíveis aos grandes (PETERSEN, 1994).

Quando um programa de melhoramento vegetal atinge a fase de produção inicial há geralmente sementes em quantidades adequadas para permitir a repetição dos ensaios. No entanto, o número de novas seleções a serem comparadas é muitas vezes bem grandes. Nesta situação não é comum encontrar parcelas uniformes suficientes para formar blocos com dimensões satisfatoriamente grandes para conter uma repetição completa com todas as novas repetições. Por isso o efeito de blocos é confundido, a não ser que o tamanho do bloco possa ser menor que o número de novas seleções sendo comparadas (YATES, 1936). Para estes ensaios, é necessário um processo de redução do tamanho dos blocos. Isso é possível com uma classe de delineamento chamada de delineamento em blocos incompletos. Como o nome implica, no delineamento em blocos incompletos as parcelas são agrupadas em blocos que não são grandes o bastante pra conter uma repetição completa de todos os tratamentos.

De uma maneira geral, um experimento em blocos incompletos tem v tratamentos, b blocos, r repetições e k parcelas por bloco, com $k < v$ e assumindo nestes casos o índice λ como sendo o número de ocorrências de cada par de tratamentos em blocos, sendo que a igualdade $rv = bk$ deve ser satisfeita (COCHRAN; COX, 1957).

Os delineamentos em blocos incompletos se dividem em duas grandes classes: blocos incompletos balanceados (BIB) e blocos incompletos parcialmente balanceados (BIPB). Nos delineamentos balanceados, todos os pares de tratamentos aparecem juntos em um bloco o mesmo número de vezes, ou seja, apresentam um único λ que é dado por $\lambda = r(k - 1)/(v - 1)$, e todos os pares de tratamentos são comparados com a mesma precisão mesmo que a diferença entre os blocos seja grande. É possível construir um delineamento balanceado para qualquer

número de tratamentos e qualquer tamanho de blocos, com o mesmo de número de parcelas por bloco. Esses dois parâmetros fixam o número mínimo de repetições necessária para o equilíbrio. Esse mínimo é frequentemente grande para o delineamento balanceado ter um significado prático (PETERSEN, 1994).

Os delineamentos balanceados nem sempre são práticos, por isso uma classe de delineamentos chamada blocos incompletos parcialmente balanceados foi desenvolvida por Bose e Nair (1939). Na mais útil delas, alguns pares de tratamentos ocorrem juntos no mesmo bloco λ_1 vezes, e outros pares ocorrem juntos λ_2 vezes, onde λ_1 e λ_2 são números inteiros. Embora os delineamentos parcialmente balanceados sejam mais úteis em termos práticos, eles não são tão desejáveis como os modelos balanceados, por serem estatisticamente difíceis de analisar. As comparações entre alguns pares de tratamentos são feitas com maior precisão do que para outros pares.

Cochran e Cox (1957) classificam os BIB e BIPB em cinco tipos:

- Tipo I - quando os blocos podem ser arranjos em repetições de tratamentos;
- Tipo II - quando os blocos não podem ser arranjos em repetições, mas podem ser agrupados formando grupos de repetições;
- Tipo III - quando os blocos não podem ser agrupados formando repetições e nem mesmo em grupos de repetições de tratamentos;
- Tipo IV - quando o número de tratamentos é igual ao número de blocos que possuem um número de unidades experimentais menor que o número de tratamentos;
- Tipo V - quando o número total de parcelas é pequeno em relação ao número de tratamentos e de blocos, acarretando um pequeno número de graus de liberdade para o resíduo.

2.3 Delineamentos látices quadrados ou reticulado quadrado

O delineamento látice quadrado é o tipo de bloco incompleto mais utilizado pelos melhoristas de plantas. É sempre do tipo I, isto é, os blocos incompletos são combinados em grupos formando repetições completas isoladamente e tanto podem ser balanceados quanto parcialmente balanceados e apresentam a análise de variância intrablocos, bem como a análise com recuperação da informação interblocos. Esta recuperação de informação implica assumir que os blocos agora são aleatórios, com média 0 e variância igual a σ_b^2 , e que existem modificações nos procedimentos estudados anteriormente. A primeira modificação é o ajuste na soma de quadrados de blocos para os efeitos de tratamentos e na soma de quadrado de tratamentos para contemplar essa recuperação de informação (ZIMMERMANN, 2014).

Nesses delineamentos o número de tratamentos deve ser um quadrado perfeito. O número de blocos e o número de parcelas em cada bloco é a raiz quadrada do número de tratamentos. Os

delineamentos diferem uns dos outros principalmente no número de repetições exigidas pelos delineamentos. Quando um delineamento em blocos incompletos é constituído de duas ou mais repetições completas, foi demonstrado que o delineamento pode ser analisado como se fosse um bloco casualizado comum. Consequentemente a separação em blocos pode ser ignorada sem destruir a validade da análise mesmo que a precisão possa ser reduzida.

Existem vários tipos de delineamentos reticulados. Estes incluem, em geral, reticulados balanceados e reticulados parcialmente balanceados. Nos reticulados balanceados cada tratamento ocorre junto no mesmo bloco com qualquer outro tratamento o mesmo número de vezes. No entanto, se houver k blocos, deve haver $(k + 1)$ repetições para ser balanceado. Isso está além dos recursos da maioria dos programas de melhoramento, devido a restrições espaciais e financeiras. Quando, ao contrário, o número r de repetições ortogonais é inferior a $(k + 1)$, há pares de tratamentos que ocorrem no mesmo bloco uma só vez, são os primeiros associados e há outros em que isso não ocorre em nenhum bloco, e estes são os segundos associados.

Nos delineamentos látices parcialmente balanceados o número de repetições é reduzido, mas nem todos os tratamentos ocorrem juntos em um mesmo bloco. Isto leva a uma diferença na precisão com que algumas comparações são feitas em relação a outras comparações. No entanto, a redução do número de repetições faz com que o experimento seja possível para a maioria dos programas de melhoramento genético, proporcionando a possibilidade de aumento da precisão associada com blocos menores (PETERSEN, 1994).

2.3.1 Látice quadrado balanceado

Dos delineamentos em blocos incompletos os chamados balanceados são os mais eficientes. O balanceamento está no fato de que quaisquer dois tratamentos devem ocorrer juntos em algum bloco do experimento. Ramalho, Ferreira e Oliveira (2005) definiram látice quadrado balanceado como um delineamento em que $v = k^2$ tratamentos são dispostos em b blocos de k parcelas, $k < v$, sendo que quaisquer dois tratamentos ocorrem juntos em um bloco do experimento. A condição de balanceamento implica na igualdade $r = k + 1$. Verifica-se, portanto, que os Látices Quadrados para serem balanceados exigem $(k + 1)$ repetições. Esses delineamentos só existem nos casos em que k é um número primo ou uma potência de primo. Assim, não existem látices balanceados para 36, 100 ou 144 tratamentos, por exemplo.

2.3.2 Látice quadrado parcialmente balanceado

Látices quadrados parcialmente balanceados são semelhantes a látices quadrados balanceados, mas são mais flexíveis no que diz respeito ao número de repetições. Delineamentos parcialmente balanceados não usam todas as possíveis repetições. Por exemplo, para látices 3×3 , 4 repetições são necessárias para um delineamento balanceado, mas se tiver apenas duas ou três repetições, estes são parcialmente balanceados. Os látices quadrados que usam apenas duas repetições são chamados látices simples, os que usam três repetições são chamados látices triplos.

2.3.2.1 Construção do delineamento látice quadrado

Suponha que o pesquisador tenha v tratamentos identificados como $1, 2, \dots, k^2$ com os números dos tratamentos dispostos em um quadrado $k \times k$. Por exemplo, em um quadrado 3×3 formando a Repetição I, como pode ser visto no quadro abaixo. Há nove tratamentos dispostos em uma ordem específica de tal modo que cada linha da matriz quadrada é considerada como um bloco contendo três tratamentos. Para construir a Repetição II, cada coluna da combinação para a Repetição I é usada para formar os três blocos na Repetição II. As repetições III e IV são baseadas em dois quadrados latinos ortogonais.

Bloco	Repetição I			Bloco	Repetição II		
1	1	2	3	4	1	4	7
2	4	5	6	5	2	5	8
3	7	8	9	6	3	6	9

Bloco	Repetição III			Bloco	Repetição IV		
7	1	6	8	10	1	5	9
8	2	4	9	11	2	6	7
9	3	5	7	12	3	4	8

Dois quadrados latinos são ortogonais, quando sobrepostos, todos os pares ordenados são distintos como por ser visto no esquema abaixo. A partir da formação padrão, os números de tratamentos que caíam com a mesma letra em um quadrado latino formam um bloco. Por exemplo, no quadrado latino 1 do esquema abaixo, os tratamentos de números 1, 6 e 8 caem na letra A, logo os tratamentos 1, 6 e 8 estarão no mesmo bloco na Repetição III (quadro acima). Analogamente à Repetição III, a Repetição IV será construída a partir do segundo quadrado latino.

Quadrado Latino 1			Quadrado Latino 2			Sobrepostos		
A	B	C	A	B	C	(A,A)	(B,B)	(C,C)
(1)	(2)	(3)	(1)	(2)	(3)			
B	C	A	C	A	B	(B,C)	(C,A)	(A,B)
(4)	(5)	(6)	(4)	(5)	(6)			
C	A	B	B	C	A	(C,B)	(A,C)	(B,A)
(7)	(8)	(9)	(7)	(8)	(9)			

É possível criar automaticamente desenhos de delineamentos reticulados com a função `design.lattice` no pacote `agricolae` (MENDIBURU, 2015) do R Core Team (2015), mas apenas para delineamentos simples e triplos.

2.3.3 Análises estatísticas convencionais para delineamentos em látice quadrado

Experimentos montados no delineamento látice quadrado podem ser analisados de maneiras distintas. Snedecor (1946) *apud* Silva (1997) aplicou aos látices simples, a estrutura

de análise de blocos casualizados. A análise como blocos casualizados se faz considerando que cada uma das repetições é um bloco completo casualizado. Essa análise não leva em conta a provável heterogeneidade dos blocos e, portanto, fornece, em geral, um maior quadrado médio do resíduo. Diante da maior facilidade atual em modelar os dados de uma maneira mais realista e eficiente, esta análise deve ser preterida, a não ser para o cálculo da eficiência do látice (PIMENTEL-GOMES, 2009).

Uma outra maneira é a análise intrablocos dos látices que se baseia no tradicional método dos quadrados mínimos, tal como o que é feito para os delineamentos inteiramente casualizados ou em blocos completos casualizados. Silva, Ferreira e Pacheco (2000) demonstraram ser a análise intrablocos a melhor para estimação de componentes de variância sempre que a eficiência relativa do delineamento em látice, em relação ao delineamento em blocos casualizados, for superior a 100% e, em caso contrário, deve-se optar pelo modelo de análise em blocos casualizados completos. Na análise intrablocos apenas comparações entre parcelas do mesmo bloco são usadas nas estimativas de efeito de tratamentos.

As formas de análise discutidas até agora estimam os efeitos de tratamentos tendo por base apenas sua própria informação, contida nos blocos onde aparece. Yates (1939), tendo como base o fato da não ortogonalidade dos delineamentos BIB, por não conterem os blocos todos os tratamentos, propôs uma nova forma de analisar esse tipo de experimento, de modo que os efeitos de tratamentos contidos nos blocos pudessem ser recuperados. Isso se chama análise com recuperação da informação interblocos.

A análise intrablocos pode ser usada para qualquer experimento em blocos incompletos e se baseia em métodos estatísticos exatos. A análise com recuperação da informação interblocos, por outro lado, permite aproveitar melhor os dados, mas usa métodos estatísticos apenas aproximados e só deve ser usada para experimentos com número de graus de liberdade relativamente grande para blocos e para o resíduo (PIMENTEL-GOMES; GARCIA, 2002).

2.3.4 Análises individuais e conjuntas

Dados de produtividade provenientes de culturas agrícolas melhoradas ou de programas de avaliação surgem, geralmente, de uma série de ensaios de campo conhecidos como ensaios multi-ambientes. Esses ensaios permitem a investigação do desempenho produtivo varietal através da instalação do experimento em várias localizações geográficas (SMITH; CULLIS; THOMPSON, 2005).

A seleção de genótipos superiores é feita com base em observações do fenótipo e, portanto, extremamente dependente da composição do valor fenotípico relativo ao caráter sob seleção (MAURO, 1991). O fenótipo é resultante dos efeitos genéticos e daqueles devidos aos ambientes, aos quais o genótipo foi exposto durante o seu desenvolvimento.

Diversos autores, como Moore e Dixon (2015), Silva, Ferreira e Pacheco (2000), Veronesi

(1995) entre outros, estudaram a análise conjunta de experimentos. Este tipo de análise é também de grande interesse para os melhoristas de plantas, pois, segundo Ramalho (1977), as estimativas de parâmetros genéticos baseadas em experimentos conduzidos em um único ambiente são superestimadas. Isto ocorre devido ao fato de que, além do componente genético, há o componente da interação genótipo-ambiente envolvido nestas estimativas (REGAZZI *et al.*, 1999). A detecção e estimativa da interação genótipo por ambiente, e o conhecimento da existência desta interação é de primordial importância no processo de seleção de novos híbridos, pois esta interação pode mascarar a performance média de um indivíduo ou linhagem na análise conjunta em vários ambientes (SOLFERINI, 2010).

A interação genótipo \times ambiente foi definida por Dessotti *et al.* (2013) como sendo o comportamento diferencial de genótipos em diferentes ambientes. Conclui-se com isso que os efeitos genéticos e ambientais não são independentes, uma vez que as respostas fenotípicas dos genótipos podem diferir com as variações ambientais (SOUZA-JÚNIOR; VENCOVSKY, 1989).

A resposta diferenciada dos genótipos aos vários ambientes, conhecida como interação genótipo \times ambiente é um fenômeno natural que faz parte da evolução das espécies. Segundo Vencovsky e Barriga (1992), a interação genótipo \times ambiente é de natureza genética, mas não no sentido usual, e sim da decorrência de instabilidades das manifestações genotípicas entre ambientes. O comportamento dos genótipos em relação ao ambiente tem merecido especial atenção, devido a sua interferência nos processos de seleção. Por esta razão, torna-se importante o conhecimento dessas interações, principalmente a sua aplicação nos estudos de estabilidade fenotípica das espécies.

A interação genótipo \times ambiente é de extrema importância nos programas de melhoramento, pois torna possível a seleção de genótipos com adaptação ampla ou específica, a escolha de locais de seleção e a determinação do número ideal de ambientes e de genótipos a serem avaliados (zonas de melhoramento) durante a seleção (KEMPTON; FOX; CERESO, 1997).

Se existem interações significativas envolvendo o fator tratamento, por exemplo, entre tratamento e local, a recomendação usual é a de fazer e relatar análise separadamente para cada local. Esta abordagem claramente faz conclusões individuais para cada local (MOORE; DIXON, 2015).

2.4 Modelos lineares mistos

O melhoramento e avaliação de culturas melhoradas é um dos mais antigos objetivos da pesquisa agrícola e teve o maior impacto na produção mundial de alimentos nos últimos 100 anos. O desafio agora é manter o melhoramento das culturas agrícolas a uma taxa que seja equivalente ao rápido aumento da população mundial (projeção de ser 8,3 bilhões de pessoas em 2025 (SMITH; CULLIS; THOMPSON, 2005)). Para chegar a isso, o melhoramento convencional e metodologias de biotecnologia são necessários. Enquanto ferramentas desta pesquisa oferecem grandes

promessas, também é importante reconhecer que os métodos convencionais de melhoramento de plantas continuam fazendo contribuições significativas no aumento da produção de alimentos. Por isso, é de fundamental importância que os métodos estatísticos usados para planejar e analisar dados de programas de melhoramento vegetal e avaliação sejam o mais precisos, eficientes e informativos possível.

Em genética, principalmente em estudos de melhoramento vegetal, muito se utiliza os modelos mistos. Nesses estudos grandes experimentos são instalados em vários locais com exatamente os mesmos delineamentos. Geralmente, o interesse do pesquisador está em identificar novas variedades de uma determinada espécie que apresente boas respostas em diversas condições ambientais. Muitas vezes, a formulação de modelos para experimentos conduzidos em mais de um local não é simples devido a possibilidade de opções em indicar se os fatores em estudo serão de efeito fixo ou aleatório.

O enquadramento de efeitos como fixos ou aleatórios está relacionado ao objetivo da análise. Se os níveis do efeito constituem-se amostras de uma população sobre a qual se quer tirar conclusões, este deve ser considerado como de efeitos aleatórios, ou se são parâmetros constantes sobre os quais se quer identificar diferenças e magnitudes, deve ser considerado de efeito fixo.

Segundo Camarinha-Filho (2002), as variáveis resposta provenientes de blocos incompletos, parcelas subdivididas, dados longitudinais, coeficientes aleatórios, análise de curvas de crescimento, curvas polinomiais e BLUPs (melhor preditor linear não viesado), entre outros, podem ser descritos por meio de um modelo linear misto. Dessa forma, a estrutura da matriz de variâncias e covariâncias deve estar inserida no modelo para melhor explicar o comportamento dos dados.

Tanto o desenvolvimento de novas cultivares, bem como a recomendação de novas variedades lançadas requerem uma seleção a ser feita entre um conjunto maior dos demais bons genótipos, logo, a estimação dos valores genotípicos é o foco em qualquer tentativa de melhoramento (PIEPHO *et al.*, 2007). Análise de dados de ensaios de melhoramento de plantas e variedades usualmente pode ser baseada em um modelo linear misto que pode ser expresso matricialmente da seguinte forma:

$$y = X\beta + Zu + \epsilon, \quad (1)$$

em que:

y é o vetor das observações;

X é a matriz de incidência dos efeitos fixos;

β é o vetor de parâmetros de efeitos fixos a serem estimados;

Z é a matriz de incidência dos efeitos aleatórios;

\mathbf{u} é o vetor de parâmetros de efeito aleatório a serem preditos tal que $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$;

ϵ é o vetor de erros aleatórios, associados a cada observação tal que $\epsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$ sendo que ϵ e \mathbf{u} são independentes.

As dimensões dessas matrizes e vetores dependem do delineamento experimental empregado.

Matricialmente, o modelo (1) é dado por

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_n & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \cdots \\ \mathbf{u}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Segundo Alcarde (2012), em termos de modelos hierárquicos, o modelo linear misto geral pode ser reescrito de tal modo que:

$$\mathbf{Y}|\mathbf{u} \sim N(\mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \mathbf{\Sigma}),$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}).$$

Assim, marginalmente, $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V})$, em que $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{\Sigma}$.

Henderson *et al.* (1959) apresentaram uma forma para a obtenção do melhor preditor linear não viesado (BLUP) dos efeitos aleatórios desse modelo misto, através da maximização da função densidade de probabilidade conjunta de \mathbf{y} e \mathbf{u} , em relação aos efeitos fixos e aleatórios.

Tomando o modelo (1), as pressuposições acerca da distribuição de \mathbf{y} , \mathbf{u} e ϵ podem ser descritas como:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \epsilon \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{X}\beta \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{\Sigma} & \mathbf{Z}\mathbf{G} & \mathbf{\Sigma} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} & \phi \\ \mathbf{\Sigma} & \phi & \mathbf{\Sigma} \end{bmatrix} \right),$$

em que:

$\mathbf{0}$ é um vetor nulo;

ϕ é uma matriz nula;

\mathbf{G} é a matriz de variâncias e covariâncias dos efeitos aleatórios presente no vetor \mathbf{u} ;

$\mathbf{\Sigma}$ é a matriz de variâncias e covariâncias dos resíduos.

A função densidade de probabilidade conjunta de \mathbf{y} e \mathbf{u} , $f(\mathbf{y}, \mathbf{u})$ de acordo com (SEARLE; CASELLA; MCCULLOCH, 2006), pode ser escrita como o produto entre a função densidade de probabilidade condicional de \mathbf{y} dado \mathbf{u} e a função densidade de probabilidade de \mathbf{u} , ou seja,

$$f(\mathbf{y}, \mathbf{u}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}[(\mathbf{y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{u})'\Sigma^{-1}(\mathbf{y}-\mathbf{X}\beta-\mathbf{Z}\mathbf{u})]} \frac{1}{(2\pi)^{n/2} |\mathbf{G}|^{1/2}} e^{-\frac{1}{2}(\mathbf{u}'\mathbf{G}^{-1}\mathbf{u})} .$$

Para se proceder à maximização de $f(\mathbf{y}, \mathbf{u})$, pode-se usar a transformação por logaritmo. Assim, fazendo-se $L = \log [f(\mathbf{y}, \mathbf{u})]$, tem-se:

$$L = \frac{1}{2}2n\log(2\pi) - \frac{1}{2}(\log|\Sigma| + \log|\mathbf{G}|) - \frac{1}{2}(\mathbf{y}'\Sigma^{-1}\mathbf{y} - 2\mathbf{y}'\Sigma^{-1}\mathbf{X}\beta - 2\mathbf{y}'\Sigma^{-1}\mathbf{Z}\mathbf{u}) + \\ + 2\beta'\mathbf{X}'\Sigma^{-1}\mathbf{Z}\mathbf{u} + \beta'\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta + \mathbf{u}'\mathbf{Z}'\Sigma^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}'\mathbf{G}^{-1}\mathbf{u} .$$

Derivando-se L em relação a β e \mathbf{u} , tem-se:

$$\begin{bmatrix} \frac{\partial L}{\partial \beta} \\ \frac{\partial L}{\partial \mathbf{u}} \end{bmatrix} = \begin{bmatrix} -\mathbf{X}'\Sigma^{-1}\mathbf{y} + \mathbf{X}'\Sigma^{-1}\mathbf{X}\beta + \mathbf{X}'\Sigma^{-1}\mathbf{Z}\mathbf{u} \\ -\mathbf{Z}'\Sigma^{-1}\mathbf{y} - \mathbf{Z}'\Sigma^{-1}\mathbf{X}\beta + \mathbf{Z}'\Sigma^{-1}\mathbf{Z}\mathbf{u} + \mathbf{G}^{-1}\mathbf{u} \end{bmatrix} .$$

Igualando-se tais derivadas a zero obtém-se:

$$\begin{bmatrix} \mathbf{X}'\Sigma^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\Sigma^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\Sigma^{-1}\mathbf{X}\hat{\beta} + \mathbf{Z}'\Sigma^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{G}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\Sigma^{-1}\mathbf{y} \\ \mathbf{Z}'\Sigma^{-1}\mathbf{y} \end{bmatrix} .$$

Dessa forma, a solução conjunta para os efeitos fixos e aleatórios, segundo (HENDERSON, 1984) é dada a partir do seguinte sistema de equações:

$$\begin{bmatrix} \mathbf{X}'\Sigma^{-1}\mathbf{X} & \mathbf{X}'\Sigma^{-1}\mathbf{Z} \\ \mathbf{Z}'\Sigma^{-1}\mathbf{X} & \mathbf{Z}'\Sigma^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\Sigma^{-1}\mathbf{y} \\ \mathbf{Z}'\Sigma^{-1}\mathbf{y} \end{bmatrix} .$$

Essas são as chamadas Equações de Modelos Mistos (EMM), que permitem obter o melhor estimador não viesado (BLUE) para os efeitos fixos ($\hat{\beta}$), bem como o melhor preditor linear não viesado (BLUP) para os efeitos aleatórios ($\hat{\mathbf{u}}$). A solução do sistema será:

$$\hat{\beta} = \left\{ \mathbf{X}' \left[\Sigma^{-1} - \Sigma^{-1}\mathbf{Z} (\mathbf{Z}'\Sigma^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\Sigma^{-1} \right] \mathbf{X} \right\}^{-1} \mathbf{X}' \left[\Sigma^{-1} - \Sigma^{-1}\mathbf{Z} (\mathbf{Z}'\Sigma^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\Sigma^{-1} \right] \mathbf{y}$$

e

$$\hat{\mathbf{u}} = (\mathbf{Z}'\Sigma^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\Sigma^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) .$$

Algumas propriedades da solução para os efeitos fixos são:

i) A solução $\hat{\beta}$ obtida pelas equações de modelo misto é uma solução de mínimos quadrados generalizados, ou seja,

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

em que: $V^{-1} = \Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1}Z'\Sigma^{-1})^{-1}Z'\Sigma^{-1}$.

ii) A variância de $\hat{\beta}$ é dada por:

$$\text{Var}(\hat{\beta}) = \left[X'\Sigma^{-1}X - X'\Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1}X \right]^{-1} .$$

Algumas propriedades para os efeitos aleatórios são:

i) O preditor \hat{u} é o melhor preditor linear não-viesado (BLUP) de u .

O termo predição refere-se a fatores aleatórios, e ao melhor preditor linear não viesado (BLUP) pode ser definido como o resultado da regressão dos efeitos de um fator aleatório u , em função das observações y , corrigidas para os efeitos dos fatores fixos $X\beta$.

ii) A variância de \hat{u} é dada por:

$$\begin{aligned} \text{Var}(\hat{u}) = & GZ' \left\{ \Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} - \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] \right. \\ & \left. X \left\{ X' \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] X \right\}^{-1} X' \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] \right\} ZG' . \end{aligned}$$

iii) A variância do erro de predição é dada por:

$$\begin{aligned} \text{Var}(u - \hat{u}) = & G - GZ' \left\{ \Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} - \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] \right. \\ & \left. X \left\{ X' \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] X \right\}^{-1} X' \left[\Sigma^{-1} - \Sigma^{-1}Z(Z'\Sigma^{-1}Z + G^{-1})^{-1}Z'\Sigma^{-1} \right] \right\} ZG' . \end{aligned}$$

Deve-se observar que o uso das equações de modelo misto, requer o conhecimento prévio das variâncias e covariâncias (G e Σ) envolvidas no modelo (SEARLE; CASELLA; MCCULLOCH, 2006). Esses parâmetros geralmente são desconhecidos, e o que se tem feito é substituí-los por suas estimativas.

2.4.1 Estimação dos componentes de variância

Componentes de variância são as variâncias associadas aos parâmetros de efeitos aleatórios de um modelo estatístico (BARBIN, 1993). A estimativa de componentes de variância tem larga aplicação no melhoramento de plantas, pois fornece subsídios na tomada de decisão durante o planejamento e execução de um programa de melhoramento. Auxilia, inclusive, na escolha da população base e do método de seleção a ser utilizado, bem como na avaliação para definir a

viabilidade da continuação de um programa em andamento (SILVA; FERREIRA; PACHECO, 2000).

Dentre os principais procedimentos para a estimação dos parâmetros genéticos em testes de progênies, destacam-se o tradicional método da análise de variância (ANOVA) e os métodos baseados na máxima verossimilhança (ML) e na máxima verossimilhança restrita (REML).

Com dados balanceados, estes podem ser analisados utilizando técnicas da ANOVA. Esta análise fornece os quadrados médios para variedades, ambientes e a tabela de dupla entrada para a interação. Os componentes da variância podem ser estimadas igualando-se os quadrados médios obtidos na tabela ANOVA com suas respectivas esperanças. Na situação mais comum de dados desbalanceados, o método da ANOVA é inadequado, portanto, uma alternativa é necessária (GOGEL; CULLIS; VERBYLA, 1995). O método mais popular para a estimação de componentes de variância em dados desbalanceados é a máxima verossimilhança restrita (PATTERSON; THOMPSON, 1971; SMITH; CULLIS; THOMPSON, 2005). Tendo em conta as estimativas dos componentes de variância, os efeitos fixos podem ser estimados usando o melhor estimador linear não viesado empírico (EBLUE) e os efeitos aleatórios preditos usando melhor preditor linear não viesado empírico (EBLUP).

No enfoque dos modelos mistos a análise da parte aleatória consiste na predição dos efeitos aleatórios, na presença de efeitos fixos, e na estimação dos componentes de variância. A análise da parte fixa consiste de estimação e testes de hipóteses sobre funções estimáveis dos efeitos fixos. Em geral, tanto a predição dos efeitos aleatórios quanto a estimação dos efeitos fixos dependem da estimação dos componentes de variância (FRITSCHÉ-NETO *et al.*, 2010).

Segundo Alcarde (2012), atualmente os métodos de máxima verossimilhança são os mais utilizados. Estes tem a vantagem de restringir o espaço paramétrico e com isso não obter componentes de variância negativos. Harville (1977) afirma que os estimadores de ML são funções de estatísticas suficientes, consistentes, assintoticamente normais e eficientes. No entanto, podem ser viesados para pequenas amostras e não consideram a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo.

Alternativamente, o método REML foi desenvolvido visando diminuir o viés em pequenas amostras e tem sido amplamente adotado, pois leva em consideração a perda dos graus de liberdade e tende a ser menos viesado que as estimativas de ML. Quando casos balanceados são analisados, as soluções fornecidas pelo REML coincidem com as apresentadas pelo método ANOVA que fornece estimadores não viesados e de variância mínima (SEARLE; CASELLA; MCCULLOCH, 2006).

Os métodos ML e REML exigem normalidade para que os estimadores tenham propriedades desejáveis. Entretanto, tais estimadores podem ser robustos aos desvios da normalidade, gerando estimativas razoáveis mesmo quando a forma da distribuição não é especificada (HARVILLE, 1977).

2.4.1.1 Método dos mínimos quadrados - ANOVA

A estimação de componentes de variância em modelos com dados balanceados é o caso mais simples e de onde deriva boa parte da metodologia para dados desbalanceados. O método dos momentos (ANOVA) é o mais comumente empregado constituindo em se igualarem formas quadráticas às suas respectivas esperanças, obtendo-se assim um conjunto de equações que permite a estimação (BARBIN, 1993). O princípio do método dos mínimos quadrados usado com dados balanceados pode ser generalizado para dados desbalanceados com o uso de qualquer forma quadrática em lugar das somas de quadrados.

Segundo Silva (1997), o método de estimação de componentes de variância mais utilizado pelos melhoristas de plantas é o método da análise de variância. Este método consiste em igualar os quadrados médios do quadro de análise de variância aos seus respectivos valores esperados, e a solução do sistema de equações resultante nos fornece os estimadores dos componentes de variância. Este método, quando aplicado a dados balanceados, o que normalmente ocorre nos experimentos em melhoramento de plantas, fornece estimadores não tendenciosos e de variância mínima, independentemente do modelo ser aleatório ou misto, o que é desejável. No entanto, pode gerar estimativas negativas para os componentes de variância, ainda que estes por definição sejam positivos.

Para Searle, Casella e McCulloch (2006), embora muito úteis e eficientes no melhoramento de espécies anuais e perenes, em experimentos com dados balanceados, os modelos matemáticos tradicionais ajustados à ANOVA não são adequados à seleção genética, para dados não balanceados sob qualquer delineamento com blocos, pois fornecem estimativas não exatas e tendenciosas das diferenças entre os tratamentos genéticos.

2.4.1.2 Método da Máxima Verossimilhança

Os estimadores de máxima verossimilhança são funções de estatísticas suficientes, além de serem consistentes, assintoticamente normais e eficientes, sendo que, geralmente, são obtidos sob suposição de normalidade. Por outro lado, esses estimadores podem ser viesados para pequenas amostras, além de não considerarem a perda de graus de liberdade resultante da estimação dos efeitos fixos do modelo (HARVILLE, 1977).

Uma das causas do viés encontrado nos estimadores de máxima verossimilhança é o fato deste método não levar em consideração a perda de graus de liberdade devido a estimação dos efeitos fixos pertencentes ao modelo matemático de interesse (CUSTÓDIO; BARBIN, 2005), pois mesmo os modelos matemáticos aleatórios possuem um efeito fixo, aquele relativo (usualmente) à média geral das observações.

2.4.1.3 Método da Máxima Verossimilhança Restrita

O método REML consiste em maximizar a função de verossimilhança conjunta de todos os contrastes de erros, $Y^* = L'Y$, em que L é a matriz com $[n - \text{posto}(X)]$ colunas, de posto completo, com colunas ortogonais às colunas da matriz X , isto é $LX = \mathbf{0}$. Ou seja, maximiza a parte da função de verossimilhança que é invariante aos efeitos fixos.

Usando $L = [L_1 : L_2]$ não singular, tal que: $L'_1X = I_p$ e $L'_2X = \mathbf{0}$.

Então:

$$\begin{cases} Y^*_1 = L'_1Y = I_p\beta + L'_1Z\gamma + L'_1\epsilon \\ Y^*_2 = L'_2Y = L'_2Z\gamma + L'_2\epsilon \end{cases} .$$

Portanto:

$$E(Y^*) = E \begin{bmatrix} Y^*_1 \\ Y^*_2 \end{bmatrix} = \begin{pmatrix} \beta \\ \mathbf{0} \end{pmatrix}$$

e

$$\text{Var}(Y^*) = \text{Var} \begin{bmatrix} Y^*_1 \\ Y^*_2 \end{bmatrix} = \begin{pmatrix} L'_1\Sigma L_1 & L'_1\Sigma L_2 \\ L'_2\Sigma L_1 & L'_2\Sigma L_2 \end{pmatrix} .$$

As distribuições de $L'Y$ pode ser subdividida em uma distribuição condicional $Y^*_1|Y^*_2$ para a estimação de β e uma distribuição marginal, baseada em Y^*_2 para a estimação dos componentes de variância.

Assim, com o logaritmo da função de verossimilhança restrita, calcula-se a função escore e a iguala a zero para obter as estimativas dos parâmetros. Sendo o vetor de efeitos fixos, BLUE (melhor estimador não viesado), dado por:

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y .$$

E o vetor de parâmetros de efeitos aleatórios, BLUP (melhor preditor linear não viesado):

$$\hat{u} = UZ'\Sigma^{-1} (Y - X\hat{\beta}) ,$$

tal que $\hat{\beta} \sim N(\beta, (X'\Sigma^{-1}X)^{-1})$ e o BLUP é equivalente a $E[u|Y]$ sob suposição de normalidade multivariada. Para obtenção dos resultados um método iterativo é necessário e às vezes o processo computacional pode ser intenso caso exista um grande número de parâmetros e/ou observações.

As equações REML com dados balanceados são idênticas aos estimadores ANOVA que são não viesados e de variância mínima. O estimador REML leva em conta os graus de liberdade envolvidos nas estimativas dos efeitos fixos, ao passo que os estimadores ML não. No caso de dados não balanceados os estimadores ML e os estimadores REML são viesados (SEARLE; CASELLA; MCCULLOCH, 2006).

No método da máxima verossimilhança restrita, cada observação é dividida em duas partes independentes uma referente aos efeitos fixos e outra aos efeitos aleatórios, de maneira que a função densidade de probabilidade das observações é dada pela soma das funções densidade de probabilidade de cada parte. A maximização da função densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o viés resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo (CAMARINHA-FILHO, 2002).

Segundo, Fritsche-Neto (2008) o método mais recomendado na estimação de componentes de variância em situações com dados desbalanceados ou não é o REML. Diversos trabalhos com dados de melhoramento genético de milho, como o de (WARDYN; EDWARDS; LAMKEY, 2007; IEMMA, 2003) tiveram os componentes de variância estimados por este método, que apresenta muitas das propriedades que os estimadores de componentes de variância devem possuir, dentre as quais pode-se citar (FRITSCHÉ-NETO, 2008):

1. Não viesado, tal que a esperança matemática do preditor seja o próprio parâmetro;
2. Consistência, tal que, com o aumento do tamanho da amostra, a esperança do preditor se aproxime do parâmetro e a variância para zero;
3. Eficiência, tal que o preditor apresente variância mínima;
4. Suficiência, tal que o preditor condene o máximo possível a informação contida na amostra e não seja função dependente do parâmetro;
5. Completitude, que está ligada à unicidade do estimador;
6. Invariância a translação, tal que a estimação dos componentes de variância não seja afetada por mudanças nos efeitos fixos.

Henderson *et al.* (1959) demonstraram que as equações de modelos mistos produzem como soluções BLUE - melhor estimação linear não-viesada dos efeitos fixos, além dos preditores BLUP - melhor predição linear não-viesada dos efeitos aleatórios.

Os estimadores ML e REML dos componentes de variância não são formas explícitas, isto é, o estimador de cada componente está em função dos estimadores dos demais componentes, e só podem ser encontrados por métodos numéricos iterativos (CAMARINHA-FILHO, 2002).

2.4.1.4 Predição do efeitos aleatórios

A obtenção das soluções das equações de modelos mistos depende do conhecimento dos componentes de variância. Por esse motivo, alguns autores preferem denominar os BLUPs obtidos dessa forma como EBLUPs, em que E significa “*empirical*”, indicando que os valores

paramétricos das variáveis são substituídos por suas estimativas (IEMMA, 2003). Para tanto foram usadas estimativas dessas variâncias pelo método da máxima verossimilhança restrita (REML), usando um processo iterativo baseado no algoritmo EM (*Expectation-Maximization*) (DEMPSTER; N.M.; D.B., 1977), ou de maximização da esperança (HENDERSON, 1984).

O algoritmo EM é estável numericamente, apresentando convergência mesmo que os valores iniciais não tenham sido adequados. Entretanto, uma inconveniência do algoritmo EM, é a lentidão para estimativas próximas ao limite do espaço paramétrico (RESENDE *et al.*, 1996). Os algoritmos livres de derivadas (*derivative-free*) apresentam vantagens computacionais, pois evitam a inversão de matrizes de ordem elevada. Os estimadores obtidos pelo EM são idênticos aos de Harville (1977). Se valores positivos forem inicialmente utilizados, a convergência para valores não negativos é garantida (HARVILLE, 1977).

2.4.2 Seleção de modelos

2.4.2.1 Teste da Razão de Verossimilhança

Quando deseja-se comparar os valores das funções de verossimilhança para dois modelos aninhados, o teste de razão de verossimilhanças é indicado. No contexto dos modelos lineares mistos, o teste da razão de verossimilhanças pode ser empregado para testar hipóteses sobre parâmetros de covariância ou parâmetros de efeito fixo, no contexto de modelos lineares mistos. A estatística do teste é calculada por:

$$2\log\left(\frac{L_2}{L_1}\right) = 2 [\log(L_2) - \log(L_1)] ,$$

em que L_1 é a função de verossimilhança do modelo aninhado e L_2 é a função de verossimilhança do modelo de referência.

Sob certas condições de regularidade, a estatística teste de razão de verossimilhanças segue, assintoticamente, a distribuição χ^2 , em que o número de graus de liberdade é obtido subtraindo-se o número de parâmetros do modelo aninhado do número de parâmetros do modelo de referência. Em situações que os parâmetros encontram-se na fronteira do espaço paramétrico, a estatística teste da razão de verossimilhanças segue uma mistura de χ^2 (SELF; LIANG, 1987).

2.5 Herdabilidade

Herdabilidade é frequentemente utilizada por melhoristas de plantas e geneticistas como uma medida de precisão de experimentos individuais ou conjuntos. Seu principal uso é para calcular a resposta à seleção. A maioria das fórmulas propostas para o cálculo de herdabilidade assumem implicitamente dados balanceados e efeitos genotípicas independentes. Ambos os pressupostos são frequentemente violados em ensaios de melhoramento de plantas (PIEPHO; MÖHRING, 2007).

A herdabilidade é um dos parâmetros genéticos que mais contribui para o trabalho do melhorista, pois ela fornece a proporção de variância genética presente na variância fenotípica total. Desta forma, ela mede a confiabilidade do valor fenotípico como indicador do valor reprodutivo. Devido a isso, a herdabilidade participa quase sempre de todas as fórmulas relacionadas com a predição de ganhos dos métodos de melhoramento, e também de inúmeras decisões práticas que os melhoristas tomam, tais como o método de seleção a ser utilizado (SILVA, 1997).

As estimativas de herdabilidade são de grande importância à predição de ganhos genéticos. Como outros parâmetros, tal como o coeficiente de variação genética, que é um indicador da magnitude de variação genética entre as progênies, a herdabilidade subsidia a escolha da melhor estratégia de melhoramento a ser adotada, mostra claramente o potencial de progresso genético esperado no programa de melhoramento e é essencial à predição de valores genéticos (FALCONER; MACKAY, 1996). A estimação e uso da herdabilidade, portanto, torna-se imprescindível aos programas de melhoramento.

A herdabilidade pode ser estimada pela razão entre a variância genética e a variância total, e observada no sentido amplo, expressa pela proporção de variância genética total em relação à variância fenotípica, ou no sentido restrito, determinada pelo quociente entre a variância genética aditiva e a variância fenotípica (DUDLEY; MOLL, 1969), ou seja, o coeficiente de herdabilidade expressa a quantidade de variação genética em relação à variação fenotípica total (VENCOVSKY; BARRIGA, 1992).

Valores altos para as estimativas de herdabilidade indicam boas possibilidades de ganho nos estudos de melhoramento, pois o progresso esperado pela seleção depende diretamente da herdabilidade e da intensidade da seleção, e inversamente ao desvio padrão fenotípico (DUDLEY; MOLL, 1969).

Grandes conjuntos de genótipos são geralmente testados em delineamentos que envolvem blocos incompletos e definições comuns de herdabilidade não se aplicam em tais casos (PIEPHO; MÖHRING, 2007). Além disso, as definições padrão assumem que os ensaios são analisados por modelos com efeitos aleatórios independentes para blocos, parcelas, plantas, etc., enquanto que a análise de ensaios de campo é muitas vezes feito por modelos espaciais, que implicam estruturas de variâncias e covariâncias complexas relativas às unidades de observação.

3 MATERIAL E MÉTODOS

3.1 Material

Os dados utilizados nesse trabalho foram fornecidos pelo Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz”. Estes são oriundos do experimento instalado nos anos agrícolas de 2002/2003 até 2006/2007 em 4 locais do município de Piracicaba-SP, como pode ser visto na Tabela 1. Os locais são Estação Experimental Fazenda Anhembi (AN), Estação Experimental Fazenda Areão (AR), Estação Experimental Fazenda Caterpillar (CA) e Estação Experimental Departamento de Genética da ESALQ/USP (ES). Cada ambiente corresponde à combinação local \times ano.

Tabela 1 – Estações experimentais e anos em que os experimentos foram instalados

E.E.	02/03	03/04	05/06	06/07
Anhembi	×	×	×	×
Areão	×			
Caterpillar	×	×		
ESALQ	×	×	×	×

É interessante ressaltar que mesmo tratando-se de experimentos conduzidos no mesmo local, em diferentes anos, a análise será conduzida como se fossem 11 ambientes distintos, devido às condições climáticas, como por exemplo, diferentes temperaturas e precipitações registradas nos diferentes anos. Portanto o efeito de ano não será considerado individualmente na análise.

As linhagens L14-04B e L08-05F foram cruzadas entre si para a obtenção da geração F_1 . Por sua vez, as plantas F_1 foram autofecundadas para a obtenção da população F_2 . Foram amostradas, aleatoriamente, 256 plantas da população F_2 . Posteriormente, para o aumento do número de sementes, as plantas F_2 foram autofecundadas obtendo-se progênies $F_{2:3}$ (Figura 1). Cada progênie $F_{2:3}$ representa, em média, a planta F_2 da qual foi originada.

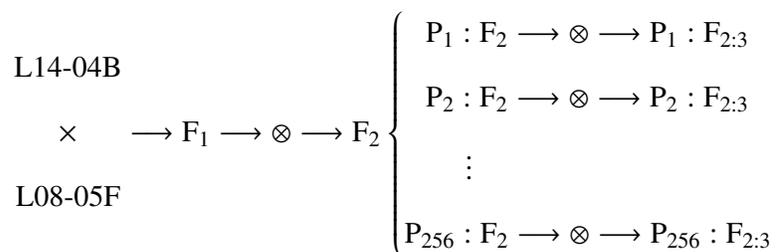


Figura 1 – Esquema para a obtenção das 256 plantas e populações utilizadas, em que as letras L, F e P representam linhagem, geração filial e planta respectivamente; \times representa cruzamento entre as linhagens e \otimes representa a autofecundação das progênies

Adaptado de Coelho (2010).

As 256 progênies foram plantadas em experimentos no delineamento látice simples 16×16 , sendo as parcelas constituídas de linhas de 4,0 m, com espaçamento de 0,80 m entre linhas e 0,20 m entre plantas. Foram distribuídas 50 sementes por parcela e, aproximadamente 30 dias após a semeadura foi realizado o desbaste, restando cerca de 20 plantas em cada parcela. Essa densidade de semeadura corresponde a um estande médio de 62.500 plantas ha^{-1} . As adubações de plantio e cobertura, além do controle de plantas daninhas, foram realizadas de acordo com o recomendado para a cultura do milho.

Considerando todas as plantas da parcela foram coletados os seguintes dados: florescimento masculino e florescimento feminino (dias para que 50% das plantas da parcela tivessem emitido a inflorescência masculina e feminina, respectivamente), estande (número de plantas por parcela na época da colheita), acamamento (número de plantas da parcela com inclinação superior a 30° em relação à vertical), quebramento (número de plantas quebradas abaixo da primeira espiga na parcela), número de espigas por parcela, peso de grãos ($Kg.parcela^{-1}$) e umidade dos grãos (%). Considerando cinco plantas competitivas da parcela foram coletados os dados de altura da planta e altura da espiga (distância, em cm, do solo até a folha bandeira e até a inserção da primeira espiga, respectivamente). Para a realização das análises estatísticas foi considerado apenas o caráter peso de grãos que foi ajustado para 15,5% de umidade e convertido para $t.ha^{-1}$ sendo então referida como produção de grãos.

3.2 Métodos

Um experimento em látice pode ser analisado individualmente, caso as repetições sejam instaladas em apenas um ambiente, ou conjunta, caso o mesmo experimento tenha sido instalado em mais de um ambiente. Para que seja possível a análise conjunta, é recomendado que a análise individual seja feita em cada um dos ambientes. Após a verificação da homogeneidade de variâncias entre os experimentos, pode-se pensar em estudar os experimentos conjuntamente, a fim de obter informações mais abrangentes sobre os efeitos estudados.

Usualmente, a análise de maior interesse para a genética, é a análise conjunta, já que estudar a interação entre genótipos e ambiente é de interesse para os melhoristas (RAMALHO; FERREIRA; OLIVEIRA, 2005). No presente trabalho foi feita a análise individual por ambiente e em seguida a análise conjunta dos 11 ambientes para comparar a ordenação das progênies nas duas situações.

Com o uso dos modelos mistos, os EBLUPs foram estimados pelo método da máxima verossimilhança restrita, que tende a apresentar menor viés quando comparado com o método de estimação da máxima verossimilhança (GILMOUR; THOMPSON; CULLIS, 1995). A comparação de modelos aninhados foi feita utilizando-se o teste de razão de verossimilhanças em que testou-se a significância das interações presentes na análise conjunta.

Com o auxílio de gráficos de diagnósticos foram encontrados bons ajustes para os dados

que é um dos principais objetivos de programas de melhoramento de plantas. Extraiu-se então o componente de variância para as progênies na análise individual e o componente de variância para a interação ambiente \times progênie na análise conjunta. Em ambos os casos ordenou-se os valores preditos de 13, 26, 39 e 52 melhores progênies, que correspondem a 5, 10, 15 e 20% das progênies estudadas no experimento. Os valores preditos foram calculados pela soma dos EBLUPs com a média geral.

As estimativas dos intervalos de confiança para os componentes de variância, com 95% de significância, foram estimados segundo Burdick e Graybill (1992).

O objetivo dos experimentos com um número grande de progênies providas do melhoramento genético, não é a comparação dos tratamentos, mas a predição e a ordenação das novas variedades (SANTOS, 2015).

Os ajustes foram realizados utilizando o *software* R (R Core Team, 2015), com o pacote estatístico lme4 (BATES *et al.*, 2015). Este fornece estimativas pelo método da máxima verossimilhança restrita (REML) do parâmetro de variância como padrão, sendo possível alterar o método de ajuste. É possível extrair também os melhores estimadores lineares não viesados empíricos (EBLUEs) dos efeitos fixos e melhores preditores lineares não viesados empíricos (EBLUPs) dos efeitos aleatórios com facilidade para inclusão/exclusão de efeitos.

Para facilitar as interpretações e para os valores serem passíveis de comparação, somou-se o valor da média geral do experimento aos EBLUPs obtidos.

Toda a análise foi feita respeitando a estrutura dos experimentos em látice quadrado simples em que cada um dos ambientes constam 2 repetições com 16 blocos incompletos e cada um dos blocos com 16 parcelas, comportando então as 256 progênies em estudo.

Baseado na tese de Coelho (2010) que considerou cada uma das repetições do látice quadrado como um bloco completo, a análise também foi feita com essa abordagem para efeito de comparação dos ranqueamentos, nas análises individuais e conjunta. Esta comparação tem como objetivo verificar a efetividade de se conduzir experimentos considerando a estrutura do látice. Cecon (1992) também estudou diferentes alternativas de análise dos látices, incluindo a que assume as repetições como sendo blocos casualizados completos, o que, segundo Bueno-Filho e Vencovsky (2000) também é usual, especialmente nos casos de erros no estabelecimento do arranjo experimental.

3.3 Especificação dos modelos

O modelo matemático é considerado misto, pois possui efeitos aleatórios, além do efeito fixo relativo à média geral.

3.3.1 Análise convencional como látice quadrado

3.3.1.1 Análises individuais

Primeiro, as análises individuais foram efetuadas de acordo com o seguinte modelo:

$$y = X_g\mu + Z_r\gamma + Z_{rb}\theta + Z_p\tau + \epsilon \quad (2)$$

em que:

y é o vetor de dados fenotípicos, de dimensão $n \times 1$, em que, n é o número de observações;

X_g é a matriz de incidência do efeito da média geral, de dimensão $n \times 1$;

μ é uma constante inerente a todas as observações;

Z_r é a matriz de incidência dos efeitos de repetições, de dimensão $n \times 2$;

γ é o vetor dos efeitos de repetições (aleatórios), de dimensão 2×1 ;

Z_{rb} é a matriz de incidência dos efeitos de blocos dentro de repetições, de dimensão $n \times 16$;

θ é o vetor dos efeitos de blocos dentro de repetições (aleatórios), de dimensão 16×1 ;

Z_p é a matriz de incidência dos efeitos genotípicos, de dimensão $n \times 256$;

τ é o vetor dos efeitos genotípicos (aleatórios), de dimensão 256×1 ;

ϵ é o vetor de erros, de dimensão $n \times 1$, sendo que $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ e $\mathbf{R} = \mathbf{I}\sigma_\epsilon^2$, em que, \mathbf{I} é uma matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é o erro experimental.

3.3.1.2 Análises conjuntas

No modelo de análise conjunta para os locais, o modelo é dado por:

$$y = X_g\mu + Z_a\alpha + Z_{ar}\gamma + Z_{arb}\theta + Z_p\tau + Z_{lp}\psi + \epsilon \quad (3)$$

em que:

y é o vetor de dados fenotípicos, de dimensão $n \times 1$, em que, n é o número de observações;

X_g é a matriz de incidência do efeito da média geral de dimensão $n \times 1$;

μ é uma constante inerente a todas as observações;

Z_a é a matriz de incidência dos efeitos de ambientes, de dimensão $n \times 4$;

α é o vetor dos efeitos de ambientes (aleatórios), de dimensão 4×1 ;

Z_{ar} é a matriz de incidência da interação ambiente e repetição, de dimensão $n \times (4 \times 2)$;

γ é o vetor dos efeitos de repetição (aleatórios), de dimensão 2×1 ;

Z_{arb} é a matriz de incidência da interação ambiente e bloco dentro de repetição, de dimensão $n \times (4 \times 2 \times 16)$;

θ é o vetor dos efeitos de blocos (aleatórios), de dimensão 16×1 ;

Z_p é a matriz de incidência dos efeitos genotípicos, de dimensão $n \times 256$;

τ é o vetor dos efeitos genotípicos (aleatórios), de dimensão 256×1 ;

Z_{ap} é a matriz de incidência da interação ambiente e progênes, de dimensão 256×4 ;

ψ é o vetor dos efeitos da interação aleatória entre ambiente e progênes, de dimensão 4×1 ;

ϵ é o vetor de erros, de dimensão $n \times 1$, sendo que $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ e $\mathbf{R} = \mathbf{I}\sigma_\epsilon^2$, em que, \mathbf{I} é uma matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é o erro experimental.

Tendo em vista um modelo inicial a ser ajustado, faz-se necessária a escolha de um método para a estimação dos componentes de variância, uma vez que os modelos mistos incorporam termos aleatórios adicionais ao modelo. Neste trabalho o método utilizado foi o da máxima verossimilhança restrita, que fornece estimativas com menor viés.

3.3.2 Análise considerando repetição como um bloco completo

Nas análises individuais, o modelo é semelhante ao descrito anteriormente com a exceção da exclusão do efeito de blocos dentro de repetições. Com isso considerou-se as repetições atuando como se fossem blocos completos. O mesmo ocorreu na análise conjunta.

Essa abordagem reduz a quantidade de matrizes a serem manipuladas, entretanto deixam de explicar possíveis variações que seriam atribuídas aos efeitos dos blocos incompletos, inflacionando assim o componente de variância atribuído ao resíduo.

3.3.2.1 Análises individuais

Primeiro, as análises individuais foram efetuadas em conformidade com o seguinte modelo:

$$\mathbf{y} = \mathbf{X}_g\mu + \mathbf{Z}_r\gamma + \mathbf{Z}_p\tau + \epsilon \quad (4)$$

em que:

\mathbf{y} é o vetor de dados fenotípicos, de dimensão $n \times 1$, em que, n é o número de observações;

\mathbf{X}_g é a matriz de incidência do efeito da média geral, de dimensão $n \times 1$;

μ é uma constante inerente a todas as observações;

\mathbf{Z}_r é a matriz de incidência dos efeitos de repetições, de dimensão $n \times 2$;

γ é o vetor dos efeitos de repetições (aleatórios), de dimensão 2×1 ;

Z_p é a matriz de incidência dos efeitos genotípicos, de dimensão $n \times 256$;

τ é o vetor dos efeitos genotípicos (aleatórios), de dimensão 256×1 ;

ϵ é o vetor de erros, de dimensão $n \times 1$, sendo que $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ e $\mathbf{R} = \mathbf{I}\sigma_\epsilon^2$, em que, \mathbf{I} é uma matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é o erro experimental.

3.3.2.2 Análise conjunta

No modelo de análise conjunta para os locais, o modelo é dado por:

$$y = X_g\mu + Z_a\alpha + Z_{ar}\gamma + Z_p\tau + Z_{ap}\psi + \epsilon \quad (5)$$

em que:

y é o vetor de dados fenotípicos, de dimensão $n \times 1$, em que, n é o número de observações;

X_g é a matriz de incidência do efeito da média geral, formadas por valores 1, de dimensão $n \times 1$;

μ é uma constante inerente a todas as observações;

Z_a é a matriz de incidência dos efeitos de ambientes, de dimensão $n \times 4$;

α é o vetor dos efeitos de ambientes (aleatórios), de dimensão 4×1 ;

Z_{ar} é a matriz de incidência da interação ambiente e repetição, de dimensão $n \times (4 \times 2)$;

γ é o vetor dos efeitos de repetição (aleatórios), de dimensão 2×1 ;

Z_p é a matriz de incidência dos efeitos genotípicos, de dimensão $n \times 256$;

τ é o vetor dos efeitos genotípicos (aleatórios), de dimensão 256×1 ;

Z_{ap} é a matriz de incidência da interação ambiente e progênes, de dimensão 256×4 ;

ψ é o vetor dos efeitos da interação aleatória entre ambiente e progênes, de dimensão 4×1 ;

ϵ é o vetor de erros, de dimensão $n \times 1$, sendo que $\epsilon \sim N(\mathbf{0}, \mathbf{R})$ e $\mathbf{R} = \mathbf{I}\sigma_\epsilon^2$, em que, \mathbf{I} é uma matriz identidade de dimensão $n \times n$ e σ_ϵ^2 é o erro experimental.

Foram comparadas também as ordenações das progênes pelos dois métodos empregados, realizando-se a seleção truncada de 5, 10, 15 e 20% das progênes, o que corresponde à seleção de 13, 26, 39 e 52 progênes respectivamente. A percentagem de coincidência na seleção foi obtida por meio do número de progênes selecionadas em comum entre os dois métodos.

4 RESULTADOS E DISCUSSÃO

4.1 Análise descritiva

Em estatística genética, estudam-se certas características dos indivíduos sem descontinuidades naturais, chamadas de “variação” ou “caracteres métricos”. O estudo de um caráter métrico centraliza-se em torno da análise de sua variação, porque é em termos de variância que são formuladas as questões primárias da genética. O valor observado, quando um caráter é medido no indivíduo, é o valor fenotípico desse indivíduo; todas as observações de médias, variâncias e covariâncias devem, claramente, ser baseadas na medida do valor fenotípico (PEREIRA-NETO, 1994).

Inicialmente foi realizada uma análise descritiva dos dados. Na Tabela 2 são apresentadas as médias observadas das produções de grãos em cada ambiente, e seus respectivos desvios padrões (D.P.), valores máximos, mínimos, amplitudes e coeficientes de variação C.V.(%). Tem-se que a média de produção dos grãos variou de 2,21 a 7,29 t.ha⁻¹ na Estação Experimental Fazenda Areão (AR) no ano agrícola 2003/2004 e Estação Experimental Departamento de Genética da ESALQ/USP (ES) em 2005/2006, respectivamente. A maior produção foi de 14,64 t.ha⁻¹ na Estação Experimental Departamento de Genética da ESALQ/USP (ES) em 2005/2006. Essas variações nos valores médios das produções de grãos obtidos entre os ambientes ocorreram, principalmente, em decorrência das características edafoclimáticas existentes nos diferentes ambientes em que os ensaios foram conduzidos.

Tabela 2 – Análise exploratória da produtividade em t.ha⁻¹ por ambiente

	Média	D.P.	Máximo	Mínimo	Amplitude	C.V.(%)
AN0203	3,28	1,23	9,13	0,57	8,56	21,61
AR0203	2,21	1,41	8,74	0,12	8,63	32,99
CA0203	2,45	1,31	9,26	0,15	9,11	27,43
ES0203	2,64	1,26	9,06	0,25	8,81	29,02
AN0304	2,33	1,26	8,67	0,07	8,61	26,39
CA0304	3,67	1,28	9,63	1,27	8,36	17,14
ES0304	4,94	1,69	13,67	1,43	12,24	16,25
AN0506	6,32	1,87	13,66	2,24	11,42	12,91
ES0506	7,29	2,19	14,64	2,40	12,25	13,94
AN0607	3,69	1,61	9,66	0,53	9,13	22,42
ES0607	5,47	2,06	12,73	0,92	11,81	20,54
Geral	4,03	2,30	14,64	0,07	14,58	19,92

Nas análises individuais o coeficiente de variação variou de 12,91% na Estação Experimental Fazenda Anhembi (AN) na safra 2005/2005 a 32,99% na Estação Experimental Fazenda Areão (AR) no ano agrícola 2003/2004. De acordo com Pimentel-Gomes (2009) se o coeficiente

de variação for maior do que 20%, é considerado alto, todavia todos os valores obtidos estão dentro dos limites aceitáveis na experimentação agrônômica (SCAPIM; CARVALHO; CRUZ, 1995; FRITSCHÉ-NETO, 2008).

A análise conjunta dos materiais genéticos foi realizada com os 11 ambientes, uma vez que a razão entre o maior e o menor componente de variância residual foi inferior a sete (PIMENTEL-GOMES, 2009). Considerando-se todos os ambientes, a média geral da produção de grãos foi de 4,03 t.ha⁻¹ e o coeficiente de variação foi 19,92%, índice considerado aceitável para experimentos de campo.

4.2 Coeficiente de determinação fenotípico entre valores observados e preditos

Os coeficientes de determinação (R^2) obtidos entre os valores observados e os preditos pelos modelos das análises individuais e conjuntas (Tabela 3) foram altos tanto para a análise do experimento considerando-se a estrutura do látice quadrado como na estrutura de blocos completos, indicando confiabilidade dos modelos. Nas duas abordagens, os coeficientes de correlação foram maiores nas análises conjuntas, 96% para o látice e 95% para blocos completos.

Os coeficientes de determinação entre os valores observados e os preditos pelo modelos são de natureza fenotípica, que é provocada por fatores genéticos e ambientais. No melhoramento genético a preocupação maior é com a associação de natureza genética.

Tabela 3 – Coeficiente de determinação fenotípico entre os valores observados e os preditos pelos modelos nas análises individuais e conjuntas

LÁTICE		DBC	
Ambiente	R^2 (%)	Ambiente	R^2 (%)
AN0203	0,92	AN0203	0,91
AR0203	0,94	AR0203	0,91
CA0203	0,94	CA0203	0,93
ES0203	0,91	ES0203	0,89
AN0304	0,94	AN0304	0,90
CA0304	0,94	CA0304	0,93
ES0304	0,95	ES0304	0,92
AN0506	0,95	AN0506	0,95
ES0506	0,95	ES0506	0,94
AN0607	0,94	AN0607	0,92
ES0607	0,93	ES0607	0,90
Conjunta	0,96	Conjunta	0,95

4.3 Estimação dos componentes de variância

Após processar o código desenvolvido no R para o caráter produção de grãos em t.ha⁻¹, pode-se encontrar os valores dos componentes de variância, obtidos pelo método REML, relativos aos efeitos aleatórios do modelo, juntamente com outras informações do modelo e do processo

de estimação. Com as estimativas dos componentes de variância, calculou-se o intervalo de confiança (I.C.) das análises individuais por ambiente e da análise conjunta dos ambientes e também o coeficiente de herdabilidade (h^2).

4.3.1 Considerando a estrutura do látice quadrado

Nas análises individuais para o caráter produção de grãos, foram obtidos os seguintes componentes de variância por ambiente (Tabela 4), juntamente de seus respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade, de acordo com o modelo descrito pela equação (2).

Tabela 4 – Estimativa dos componentes de variância das análises individuais por ambiente, com os respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade considerando-se a estrutura do látice quadrado

Ambiente	$\hat{\sigma}_p^2$	I.C.	$\hat{\sigma}_e^2$	I.C.	h^2
AN0203	1,01	[0,73 ; 1,47]	0,50	[0,42 ; 0,60]	0,67
AR0203	1,25	[0,94 ; 1,73]	0,53	[0,45 ; 0,64]	0,70
CA0203	1,22	[0,95 ; 1,65]	0,45	[0,38 ; 0,54]	0,73
ES0203	0,91	[0,59 ; 1,60]	0,59	[0,50 ; 0,70]	0,61
AN0304	0,93	[0,71 ; 1,27]	0,38	[0,32 ; 0,45]	0,71
CA0304	1,23	[0,97 ; 1,62]	0,40	[0,34 ; 0,47]	0,76
ES0304	2,00	[1,57 ; 2,63]	0,64	[0,55 ; 0,77]	0,76
AN0506	2,83	[2,28 ; 3,60]	0,67	[0,56 ; 0,80]	0,81
ES0506	3,74	[2,98 ; 4,83]	1,03	[0,88 ; 1,24]	0,78
AN0607	1,88	[1,45 ; 2,52]	0,68	[0,58 ; 0,82]	0,73
ES0607	2,71	[2,00 ; 3,86]	1,26	[1,07 ; 1,51]	0,68

A análise conjunta foi realizada segundo o modelo descrito pela equação (3). Os componentes de variância obtidos constam na tabela 5, juntamente de seus respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade.

Tabela 5 – Estimativa dos componentes de variância da análise conjunta, com os respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade considerando-se a estrutura do látice quadrado

Parâmetro	Estimativa	I.C.
$\hat{\sigma}_p^2$	1,27	[0,96 ; 1,77]
$\hat{\sigma}_{ap}^2$	0,54	[0,38 ; 0,84]
$\hat{\sigma}_e^2$	0,64	[0,38 ; 0,84]
h^2	0,52	-

Com o objetivo de determinar o efeito da interação genótipo \times ambiente foi feita a comparação de modelos aninhados com e sem esta interação, pelo teste de razão de verossimilhanças. Este indicou significância da interação, portanto conclui-se que as progênies apresentam comportamento diferenciado em ambientes distintos.

O coeficiente de herdabilidade é um parâmetro de elevada importância no melhoramento genético e indica a porcentagem de variação fenotípica que é atribuída ao efeito genotípico. O experimento conduzido na Estação Experimental Departamento de Genética da ESALQ/USP (ES) durante a safra 2002/2003 apresentou a menor estimativa do coeficiente de herdabilidade (61%), enquanto que a maior estimativa foi registrada na Estação Experimental Fazenda Anhembi (AN) na safra 2005/2006 (81%). Na análise conjunta, entretanto, encontrou-se baixa estimativa do coeficiente de herdabilidade, inferior a todos os ambientes em estudo.

4.3.2 Considerando cada repetição como um bloco completo

Abandonando-se a estrutura do látice quadrado e considerando cada repetição como se fosse um bloco completo, houve pouca variação nos resultados obtidos em relação à metodologia do látice, como pode ser visto na Tabela 6, em que constam os componentes de variância obtidos juntamente de seus respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade.

Tabela 6 – Estimativa dos componentes de variância das análises individuais por ambiente, com os respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade considerando-se a estrutura de blocos completos

Ambiente	$\hat{\sigma}_p^2$	I.C.	$\hat{\sigma}_e^2$	I.C.	h^2
AN0203	0,99	[0,68 ; 1,58]	0,52	[0,44 ; 0,63]	0,66
AR0203	1,28	[0,88 ; 2,04]	0,67	[0,57 ; 0,81]	0,66
CA0203	1,21	[0,91 ; 1,70]	0,49	[0,41 ; 0,59]	0,71
ES0203	0,93	[0,51 ; 2,17]	0,65	[0,55 ; 0,78]	0,59
AN0304	0,99	[0,63 ; 1,80]	0,60	[0,51 ; 0,72]	0,62
CA0304	1,22	[0,94 ; 1,64]	0,42	[0,35 ; 0,50]	0,75
ES0304	1,99	[1,46 ; 2,86]	0,86	[0,73 ; 1,03]	0,70
AN0506	2,83	[2,27 ; 3,62]	0,67	[0,57 ; 0,80]	0,81
ES0506	3,73	[2,94 ; 4,88]	1,06	[0,90 ; 1,28]	0,78
AN0607	1,84	[1,37 ; 2,60]	0,75	[0,64 ; 0,90]	0,71
ES0607	2,61	[1,65 ; 4,74]	1,59	[1,34 ; 1,90]	0,62

De maneira semelhante às análises individuais, pode-se observar também que não há grande diferença nos resultados obtidos pela análise conjunta convencional do experimento, considerando a estrutura do látice, como pode ser visto na Tabela 7.

Tabela 7 – Estimativa dos componentes de variância da análise conjunta, com os respectivos intervalos de confiança a 95% e do coeficiente de herdabilidade considerando-se a estrutura de blocos completos

Parâmetro	Estimativa	I.C.
$\hat{\sigma}_p^2$	1,26	[0,95 ; 1,74]
$\hat{\sigma}_{ap}^2$	0,52	[0,43 ; 0,65]
$\hat{\sigma}_e^2$	0,75	[0,71 ; 0,80]
h^2	0,50	-

4.4 Diagnósticos

Para verificar a qualidade de ajuste dos modelos selecionados, gráficos de diagnósticos foram obtidos para os modelos das análises individuais. Os resíduos condicionais em relação aos valores ajustados para os modelos apresentaram um comportamento aleatório em torno de zero, possuindo poucas observações atípicas, como pode ser visto na Figura 2 para os modelos em que se considerou a estrutura do látice na análise dos dados e na Figura 3 para aqueles em que se considerou cada repetição como um bloco completo. Nas duas abordagens, o ambiente que mais apresentou pontos fora do intervalo $(-2, 2)$ foi a Estação Experimental Departamento de Genética da ESALQ/USP (ES) no ano agrícola 2006/2007. Uma outra forma de verificação da qualidade do ajuste são os gráficos de dispersão em que compara-se as médias de cada um dos tratamentos com os EBLUPs obtidos acrescidos das médias dos tratamentos, como pode ser visto nas Figuras 4 e 5.

Os gráficos quantil-quantil com envelope de simulação da distribuição normal para verificação da adequação do ajuste de um determinado modelo a um conjunto de observações consiste em se plotarem os valores absolutos ordenados de uma determinada medida de diagnóstico versus as estatísticas de ordem esperadas da distribuição meio-normal (MORAL; HINDE; DEMETRIO, 2016). Estes gráficos indicam que se a maior parte dos pontos estiverem contidos no envelope simulado, há indícios de bom ajuste do modelo. Os gráficos obtidos, para cada um dos ambientes, apresentaram tendências, no entanto, poucos resíduos ficaram fora dos envelopes, atendendo então à pressuposição de normalidade (Figuras 6 e 7). Os ambientes que mais apresentaram pontos fora do envelope foram Estação Experimental Fazenda Areão (AR), Estação Experimental Fazenda Caterpillar (CA) e Estação Experimental Departamento de Genética da ESALQ/USP (ES), todas no ano agrícola 2002/2003 para as duas situações de análise.

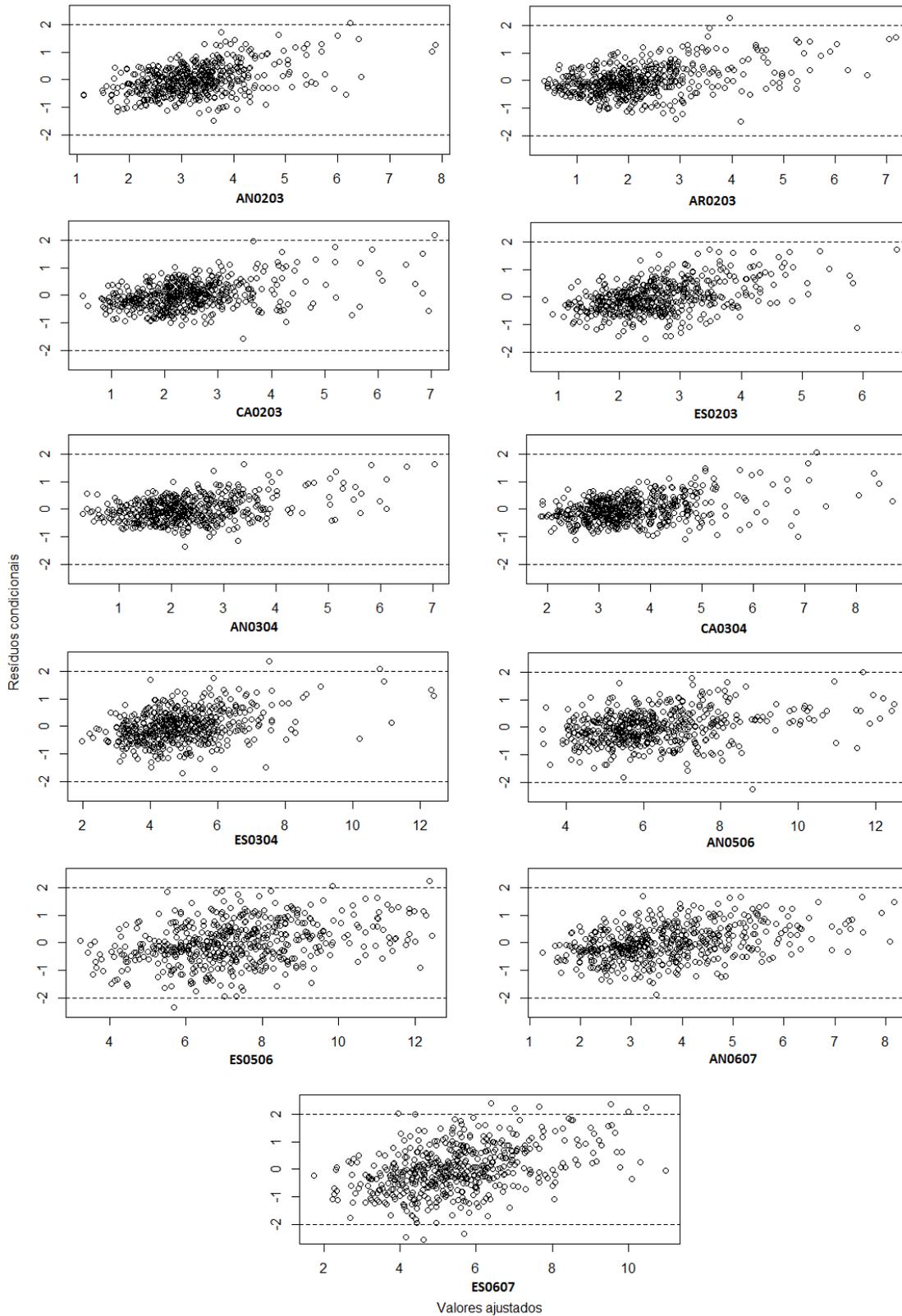


Figura 2 – Resíduos condicionais das análises individuais em cada um dos ambientes - estrutura látice

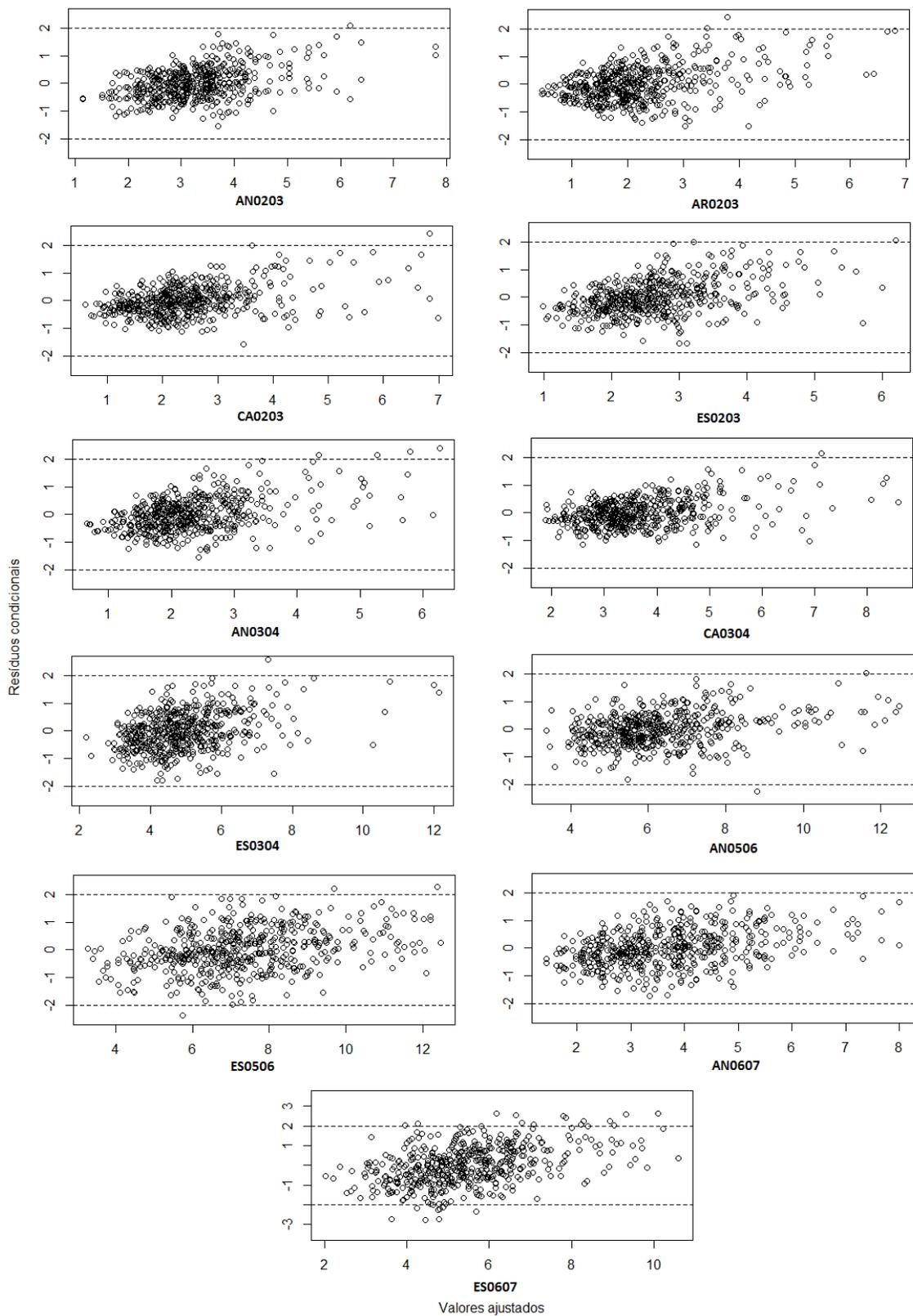


Figura 3 – Resíduos condicionais das análises individuais em cada um dos ambientes - estrutura de blocos completos

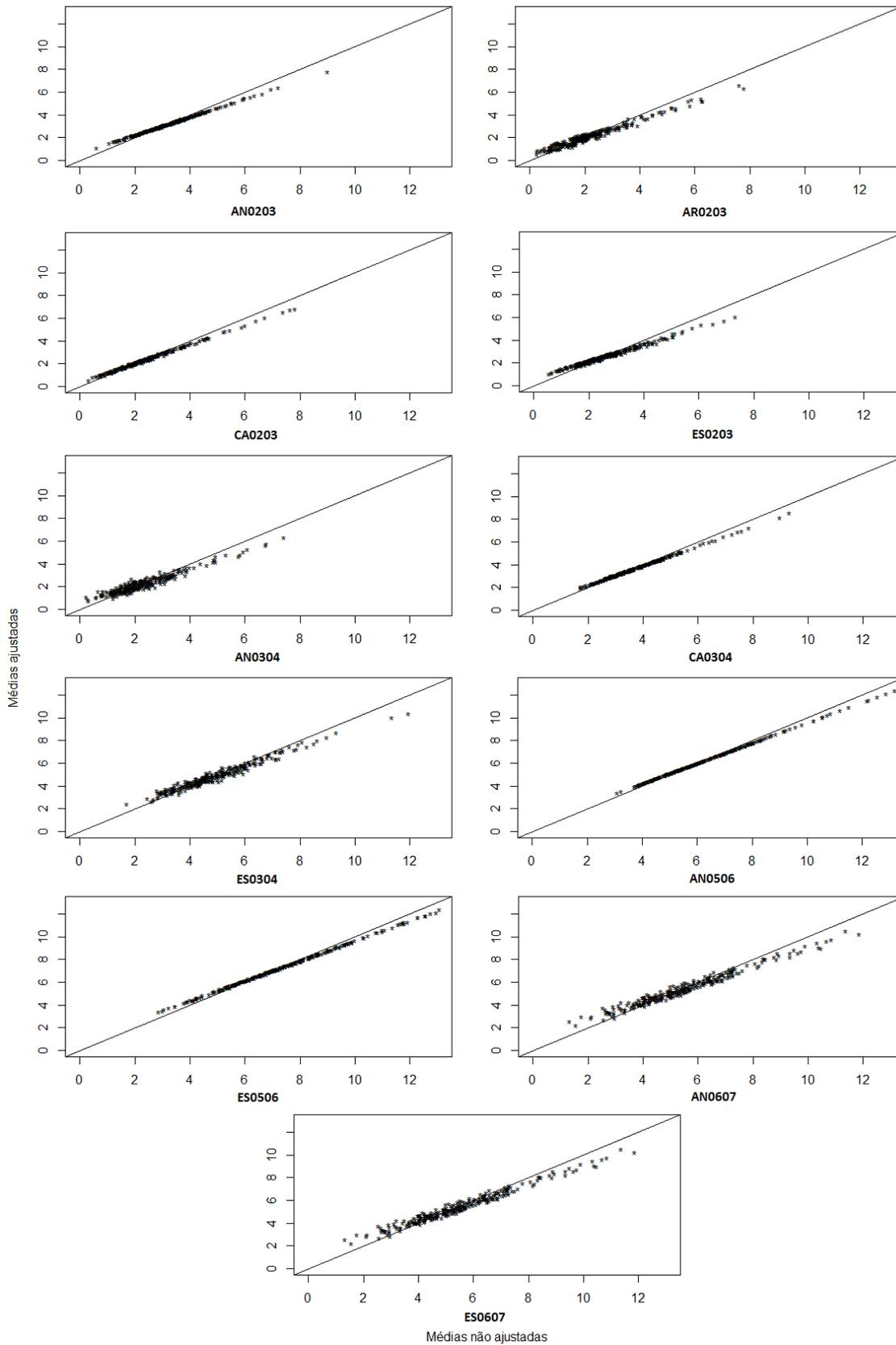


Figura 4 – Comparação das médias de tratamentos em relação aos EBLPs obtidos após ajuste do modelo - estrutura látice

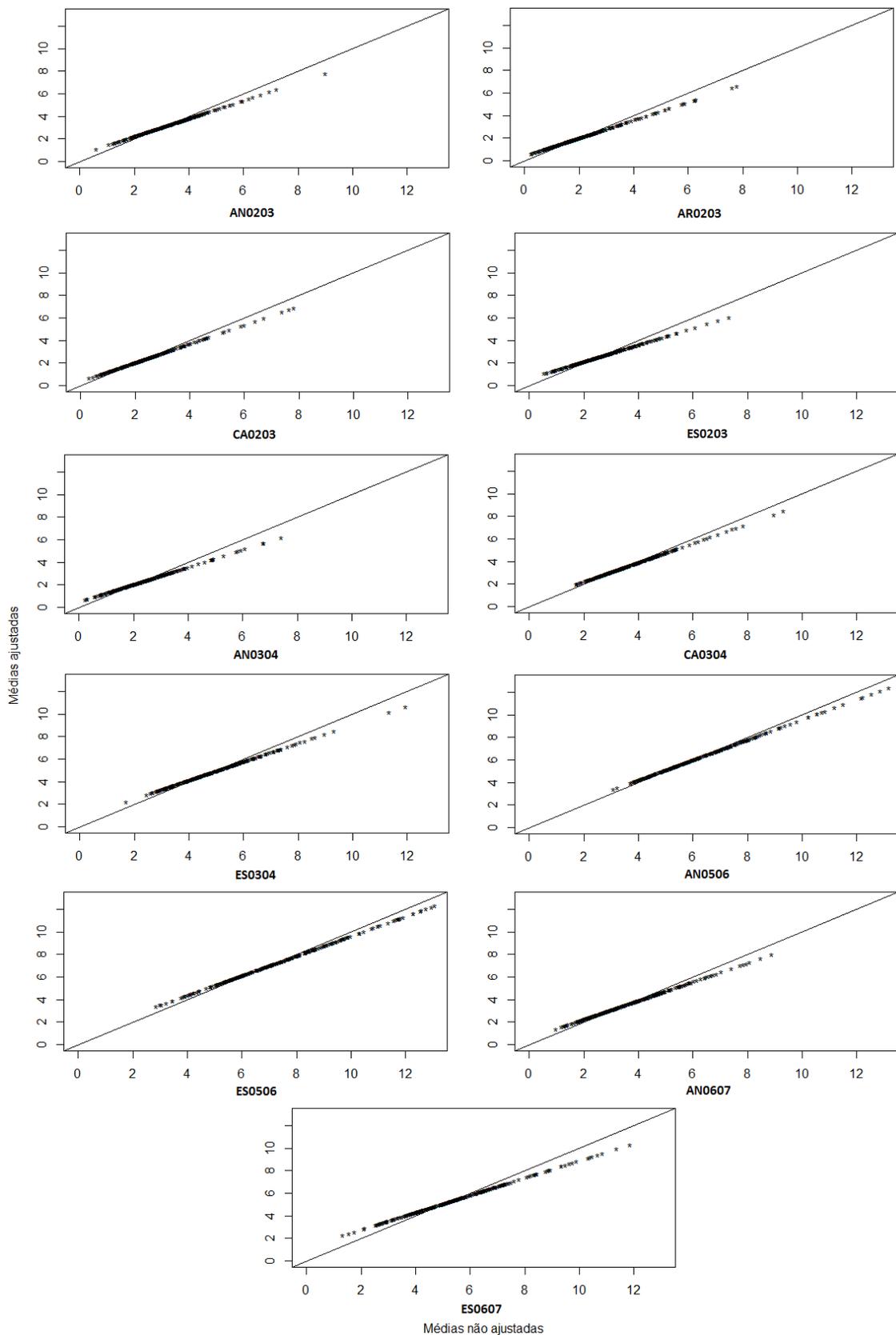


Figura 5 – Comparação das médias de tratamentos em relação aos EBLPs obtidos após ajuste do modelo - estrutura de blocos completos

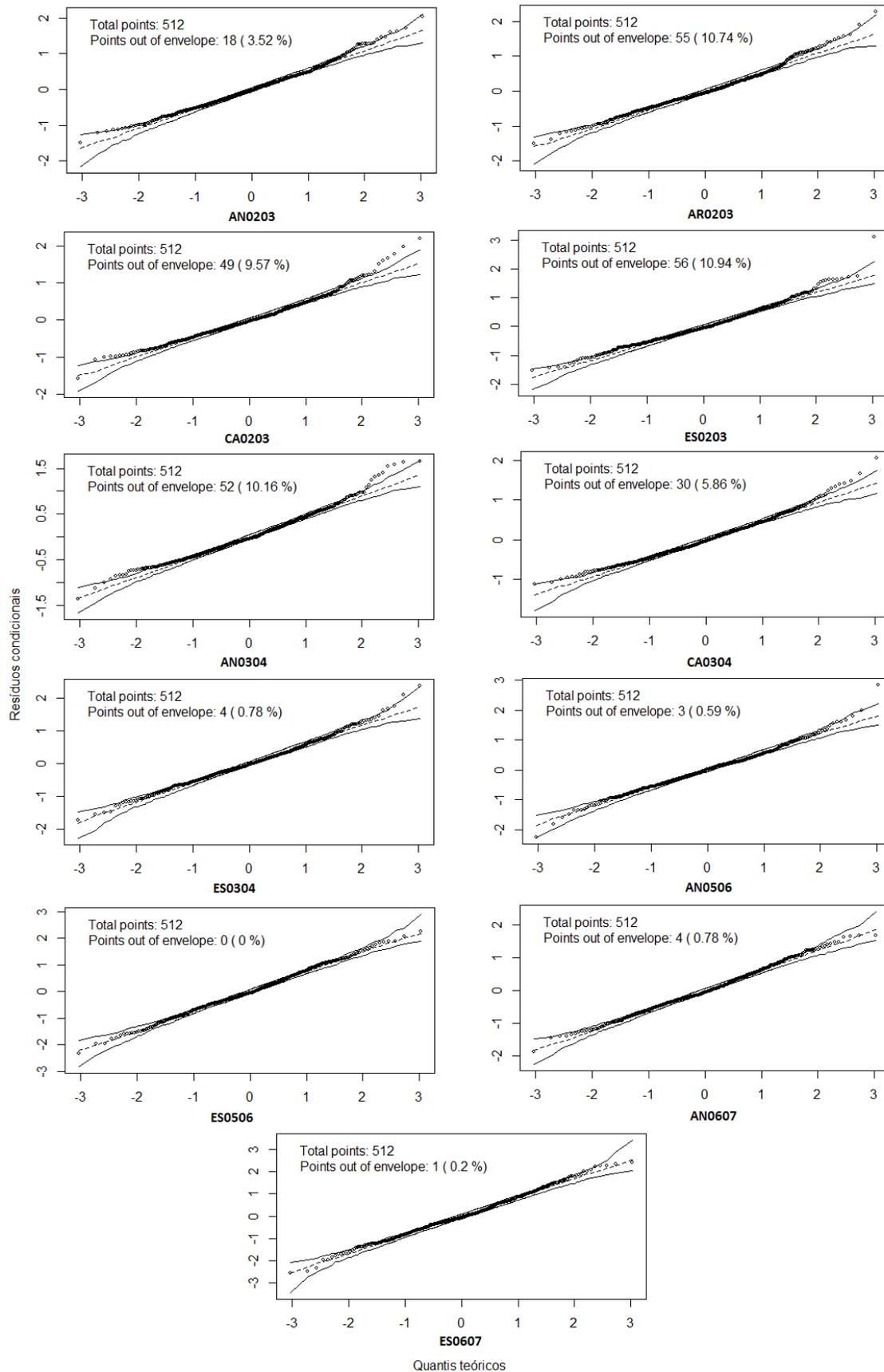


Figura 6 – Resíduos condicionais das análises individuais dos em cada um dos ambientes - estrutura látice

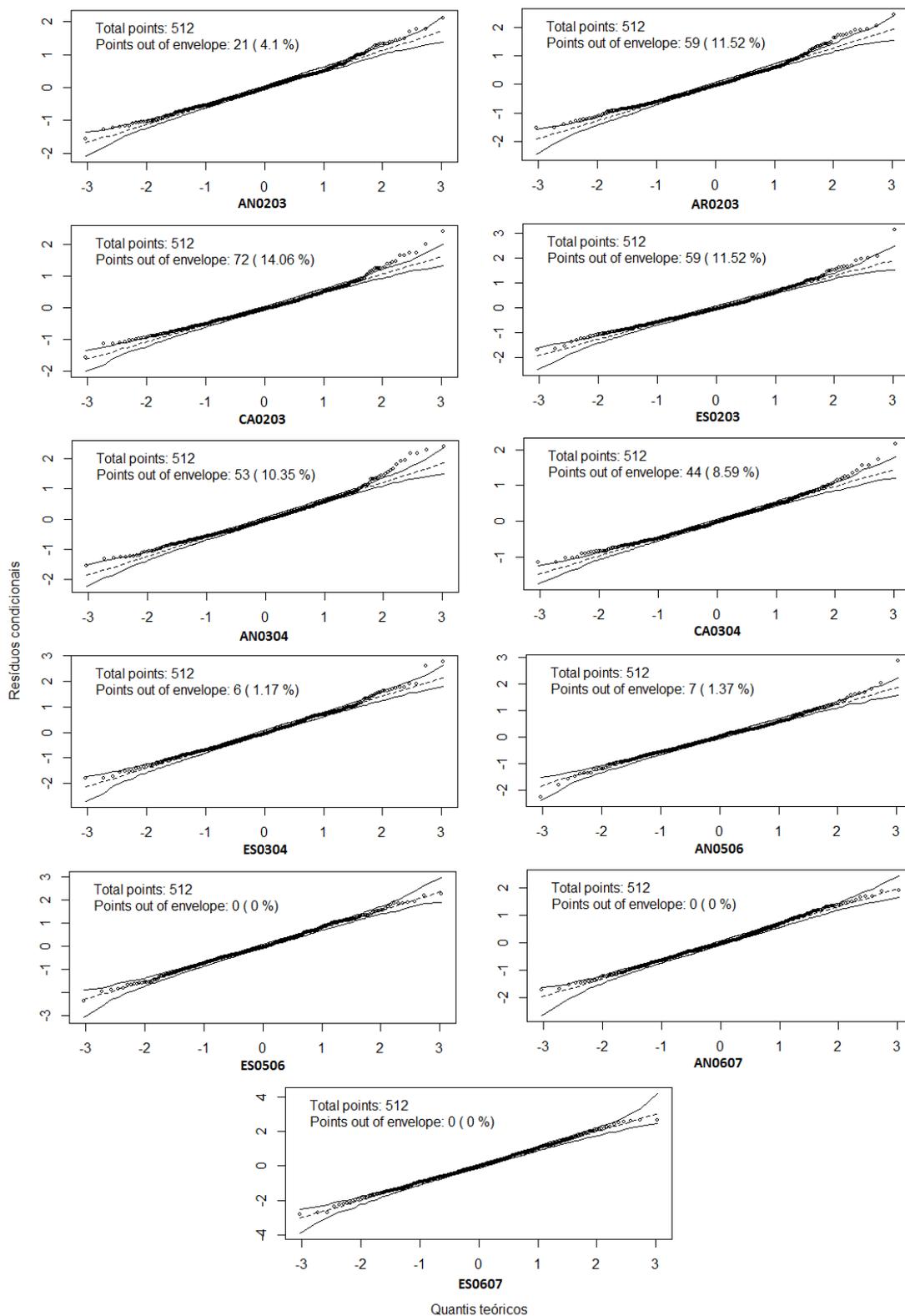


Figura 7 – Resíduos condicionais das análises individuais em cada um dos ambientes - estrutura de blocos completos

4.5 Ordenação e seleção das progênes obtidos a partir da predição dos efeitos aleatórios

Para comparar a porcentagem de progênes selecionadas levando-se os 11 ambientes combinados dois a dois, os EBLUPs obtidos foram colocados em ordem crescente em cada um dos ambientes e verificou-se quantas progênes foram coincidentes considerando-se 5, 10, 15 e 20% de seleção nas análises individuais e conjuntas, o que corresponde a selecionar por ambiente as 13, 26, 39 e 52 progênes mais produtivas, respectivamente, como pode ser visto nas Tabelas 8 e 9. Nota-se que a porcentagem de progênes coincidentes entre ambiente foi bem maior nas análises individuais do que na análise conjunta, confirmando assim a significância da interação genótipo \times ambiente. Isso ajudará o melhorista na seleção das progênes mais produtivas a serem comparadas nos próximos ensaios.

Nas análises individuais do experimento respeitando a estrutura do látice quadrado e tomando-se as progênes 5% mais produtivas, 15,4% delas foram coincidentes entre os 11 ambientes, já quando o percentual de seleção sobe para 20% das melhores progênes, 19,2% das progênes coincidem. Na análise conjunta nenhuma progênie foi mais produtiva em todos os ambientes, mesmo com a seleção das 52 melhores progênes (20% delas).

Nas análises individuais do experimento levando-se agora em consideração cada repetição do látice como um bloco completo, ao comparar todos os ambiente, tomando-se as progênes 5% mais produtivas, 23,1% delas foram coincidentes, já quando o percentual de seleção sobe para 20% das melhores progênes, 19,2% das progênes coincidem (mesmo percentual da análise anterior). Da mesma maneira que na abordagem anterior, na análise conjunta nenhuma progênie foi mais produtiva em todos os ambientes, mesmo com a seleção das 52 melhores progênes (20% delas) de todos ambientes.

É interessante comparar também a ordenação das progênes em cada ambiente nas duas abordagens consideradas. Tomando-se 5, 10, 15 e 20% de seleção nas análises individuais não foram encontradas grandes diferenças na ordenação e seleção das progênes com a utilização da análise pela estrutura do látice quadrado em confronto com a utilização de cada repetição do látice como um bloco completo (Tabela 10).

Tabela 8 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas nas análises individuais e conjunta como látice

Ambientes	Análises individuais				Análise conjunta			
	5% (13)	10% (26)	15% (39)	20% (52)	5% (13)	10% (26)	15% (39)	20% (52)
AN0203 e AR0203	53,8%	46,2%	53,8%	55,8%	7,7%	15,4%	23,1%	28,8%
AN0203 e CA0203	69,2%	57,7%	53,8%	50,0%	7,7%	11,5%	15,4%	17,3%
AN0203 e ES0203	61,5%	46,2%	56,4%	53,8%	15,4%	7,7%	10,3%	17,3%
AN0203 e AN0304	61,5%	61,5%	56,4%	51,9%	23,1%	15,4%	17,9%	28,8%
AN0203 e CA0304	53,8%	50,0%	56,4%	55,8%	23,1%	19,2%	25,6%	25,0%
AN0203 e ES0304	61,5%	50,0%	61,5%	59,6%	15,4%	23,1%	28,2%	26,9%
AN0203 e AN0506	61,5%	50,0%	51,3%	53,8%	7,7%	19,2%	23,1%	23,1%
AN0203 e ES0506	46,2%	46,2%	38,5%	40,4%	7,7%	26,9%	25,6%	28,8%
AN0203 e AN0607	53,8%	50,0%	48,7%	51,9%	0,0%	3,8%	2,6%	3,8%
AN0203 e ES0607	38,5%	42,3%	38,5%	42,3%	0,0%	0,0%	0,0%	3,8%
AR0203 e CA0203	61,5%	61,5%	59,0%	69,2%	0,0%	3,8%	5,1%	7,7%
AR0203 e ES0203	53,8%	53,8%	53,8%	67,3%	0,0%	3,8%	7,7%	5,8%
AR0203 e AN0304	69,2%	57,7%	61,5%	63,5%	15,4%	30,8%	28,2%	30,8%
AR0203 e CA0304	76,9%	50,0%	51,3%	57,7%	38,5%	34,6%	30,8%	34,6%
AR0203 e ES0304	61,5%	53,8%	66,7%	71,2%	15,4%	7,7%	7,7%	23,1%
AR0203 e AN0506	53,8%	57,7%	46,2%	48,1%	23,1%	30,8%	28,2%	30,8%
AR0203 e ES0506	53,8%	53,8%	43,6%	46,2%	23,1%	26,9%	33,3%	32,7%
AR0203 e AN0607	46,2%	46,2%	41,0%	51,9%	0,0%	0,0%	0,0%	1,9%
AR0203 e ES0607	38,5%	53,8%	46,2%	48,1%	0,0%	0,0%	0,0%	1,9%
CA0203 e ES0203	61,5%	69,2%	61,5%	65,4%	38,5%	34,6%	43,6%	46,2%
CA0203 e AN0304	61,5%	57,7%	61,5%	65,4%	0,0%	7,7%	10,3%	11,5%
CA0203 e CA0304	69,2%	61,5%	56,4%	57,7%	0,0%	0,0%	0,0%	0,0%
CA0203 e ES0304	61,5%	53,8%	61,5%	57,7%	0,0%	3,8%	10,3%	15,4%
CA0203 e AN0506	61,5%	42,3%	43,6%	46,2%	0,0%	0,0%	0,0%	0,0%
CA0203 e ES0506	53,8%	42,3%	35,9%	46,2%	7,7%	7,7%	5,1%	5,8%
CA0203 e AN0607	53,8%	46,2%	41,0%	44,2%	7,7%	46,2%	48,7%	51,9%
CA0203 e ES0607	38,5%	46,2%	41,0%	38,5%	7,7%	34,6%	38,5%	38,5%
ES0203 e AN0304	53,8%	53,8%	59,0%	59,6%	0,0%	7,7%	7,7%	9,6%
ES0203 e CA0304	53,8%	65,4%	46,2%	57,7%	0,0%	3,8%	2,6%	3,8%
ES0203 e ES0304	69,2%	50,0%	59,0%	55,8%	0,0%	11,5%	10,3%	13,5%
ES0203 e AN0506	61,5%	46,2%	43,6%	40,4%	7,7%	7,7%	12,8%	9,6%
ES0203 e ES0506	46,2%	50,0%	46,2%	46,2%	0,0%	0,0%	7,7%	11,5%
ES0203 e AN0607	69,2%	53,8%	48,7%	50,0%	7,7%	23,1%	35,9%	40,4%
ES0203 e ES0607	38,5%	50,0%	41,0%	46,2%	15,4%	23,1%	25,6%	36,5%
AN0304 e CA0304	76,9%	53,8%	46,2%	53,8%	23,1%	38,5%	35,9%	38,5%
AN0304 e ES0304	76,9%	61,5%	64,1%	61,5%	15,4%	15,4%	17,9%	19,2%
AN0304 e AN0506	69,2%	53,8%	53,8%	46,2%	15,4%	19,2%	28,2%	36,5%
AN0304 e ES0506	61,5%	50,0%	43,6%	48,1%	23,1%	26,9%	35,9%	38,5%
AN0304 e AN0607	46,2%	42,3%	41,0%	46,2%	0,0%	3,8%	7,7%	13,5%
AN0304 e ES0607	46,2%	38,5%	46,2%	42,3%	0,0%	3,8%	7,7%	9,6%
CA0304 e ES0304	69,2%	61,5%	59,0%	59,6%	30,8%	19,2%	25,6%	36,5%
CA0304 e AN0506	61,5%	50,0%	48,7%	55,8%	15,4%	23,1%	28,2%	32,7%
CA0304 e ES0506	61,5%	46,2%	43,6%	50,0%	23,1%	23,1%	28,2%	28,8%
CA0304 e AN0607	46,2%	53,8%	48,7%	51,9%	0,0%	0,0%	7,7%	9,6%
CA0304 e ES0607	38,5%	46,2%	43,6%	46,2%	0,0%	0,0%	0,0%	1,9%
ES0304 e AN0506	76,9%	50,0%	48,7%	51,9%	7,7%	7,7%	10,3%	17,3%
ES0304 e ES0506	53,8%	50,0%	51,3%	51,9%	23,1%	23,1%	30,8%	34,6%
ES0304 e AN0607	61,5%	46,2%	46,2%	55,8%	0,0%	0,0%	2,6%	9,6%
ES0304 e ES0607	46,2%	42,3%	51,3%	50,0%	0,0%	3,8%	10,3%	15,4%
AN0506 e ES0506	69,2%	73,1%	61,5%	63,5%	15,4%	15,4%	25,6%	28,8%
AN0506 e AN0607	76,9%	73,1%	71,8%	69,2%	0,0%	11,5%	7,7%	5,8%
AN0506 e ES0607	53,8%	65,4%	61,5%	61,5%	0,0%	3,8%	5,1%	9,6%
ES0506 e AN0607	53,8%	61,5%	71,8%	61,5%	7,7%	3,8%	7,7%	7,7%
ES0506 e ES0607	69,2%	73,1%	61,5%	67,3%	0,0%	3,8%	10,3%	13,5%
AN0607 e ES0607	53,8%	69,2%	56,4%	55,8%	30,8%	34,6%	48,7%	50,0%
TODOS	15,4%	15,4%	17,9%	19,2%	0,0%	0,0%	0,0%	0,0%

Tabela 9 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas nas análises individuais e conjunta como blocos casualizados completos

Ambientes	Análises individuais				Análise conjunta			
	5% (13)	10% (26)	15% (39)	20% (52)	5% (13)	10% (26)	15% (39)	20% (52)
AN0203 e AR0203	53,8%	50,0%	51,3%	51,9%	0,0%	15,4%	17,9%	32,7%
AN0203 e CA0203	61,5%	61,5%	53,8%	53,8%	0,0%	7,7%	15,4%	17,3%
AN0203 e ES0203	61,5%	46,2%	53,8%	51,9%	0,0%	11,5%	10,3%	17,3%
AN0203 e AN0304	61,5%	53,8%	53,8%	53,8%	23,1%	19,2%	17,9%	26,9%
AN0203 e CA0304	53,8%	46,2%	56,4%	55,8%	7,7%	23,1%	20,5%	26,9%
AN0203 e ES0304	61,5%	46,2%	59,0%	59,6%	15,4%	26,9%	28,2%	26,9%
AN0203 e AN0506	61,5%	50,0%	53,8%	53,8%	23,1%	19,2%	25,6%	28,8%
AN0203 e ES0506	46,2%	42,3%	41,0%	42,3%	7,7%	15,4%	28,2%	28,8%
AN0203 e AN0607	53,8%	50,0%	51,3%	50,0%	0,0%	0,0%	2,6%	3,8%
AN0203 e ES0607	46,2%	42,3%	35,9%	38,5%	0,0%	0,0%	5,1%	5,8%
AR0203 e CA0203	61,5%	65,4%	59,0%	67,3%	0,0%	3,8%	7,7%	5,8%
AR0203 e ES0203	53,8%	46,2%	53,8%	63,5%	0,0%	3,8%	7,7%	13,5%
AR0203 e AN0304	69,2%	61,5%	56,4%	59,6%	15,4%	23,1%	23,1%	30,8%
AR0203 e CA0304	76,9%	53,8%	51,3%	53,8%	30,8%	38,5%	33,3%	32,7%
AR0203 e ES0304	69,2%	61,5%	66,7%	67,3%	7,7%	11,5%	10,3%	21,2%
AR0203 e AN0506	53,8%	53,8%	46,2%	46,2%	7,7%	15,4%	20,5%	30,8%
AR0203 e ES0506	53,8%	53,8%	43,6%	48,1%	23,1%	19,2%	23,1%	30,8%
AR0203 e AN0607	46,2%	53,8%	38,5%	46,2%	0,0%	3,8%	2,6%	1,9%
AR0203 e ES0607	46,2%	53,8%	46,2%	50,0%	0,0%	0,0%	0,0%	1,9%
CA0203 e ES0203	61,5%	65,4%	66,7%	67,3%	30,8%	34,6%	48,7%	46,2%
CA0203 e AN0304	61,5%	53,8%	56,4%	65,4%	0,0%	0,0%	10,3%	7,7%
CA0203 e CA0304	69,2%	61,5%	53,8%	61,5%	0,0%	0,0%	0,0%	1,9%
CA0203 e ES0304	61,5%	50,0%	61,5%	61,5%	0,0%	3,8%	5,1%	17,3%
CA0203 e AN0506	53,8%	42,3%	43,6%	50,0%	0,0%	0,0%	0,0%	3,8%
CA0203 e ES0506	46,2%	42,3%	35,9%	50,0%	7,7%	7,7%	7,7%	9,6%
CA0203 e AN0607	46,2%	50,0%	41,0%	48,1%	23,1%	42,3%	53,8%	50,0%
CA0203 e ES0607	46,2%	50,0%	41,0%	44,2%	15,4%	30,8%	35,9%	38,5%
ES0203 e AN0304	53,8%	38,5%	48,7%	50,0%	0,0%	3,8%	10,3%	9,6%
ES0203 e CA0304	61,5%	57,7%	43,6%	55,8%	0,0%	3,8%	2,6%	5,8%
ES0203 e ES0304	69,2%	42,3%	51,3%	61,5%	0,0%	3,8%	7,7%	11,5%
ES0203 e AN0506	61,5%	38,5%	41,0%	40,4%	7,7%	7,7%	10,3%	9,6%
ES0203 e ES0506	46,2%	42,3%	46,2%	46,2%	0,0%	3,8%	5,1%	7,7%
ES0203 e AN0607	61,5%	50,0%	46,2%	48,1%	0,0%	19,2%	43,6%	40,4%
ES0203 e ES0607	46,2%	46,2%	43,6%	48,1%	15,4%	19,2%	33,3%	36,5%
AN0304 e CA0304	76,9%	50,0%	51,3%	46,2%	23,1%	26,9%	30,8%	32,7%
AN0304 e ES0304	76,9%	57,7%	61,5%	57,7%	23,1%	15,4%	17,9%	17,3%
AN0304 e AN0506	69,2%	50,0%	46,2%	42,3%	23,1%	23,1%	25,6%	38,5%
AN0304 e ES0506	61,5%	46,2%	41,0%	44,2%	30,8%	30,8%	35,9%	34,6%
AN0304 e AN0607	46,2%	42,3%	38,5%	44,2%	0,0%	0,0%	10,3%	11,5%
AN0304 e ES0607	53,8%	34,6%	35,9%	34,6%	0,0%	3,8%	12,8%	9,6%
CA0304 e ES0304	76,9%	61,5%	56,4%	61,5%	23,1%	30,8%	25,6%	30,8%
CA0304 e AN0506	61,5%	50,0%	46,2%	55,8%	15,4%	23,1%	28,2%	34,6%
CA0304 e ES0506	61,5%	46,2%	38,5%	50,0%	15,4%	15,4%	23,1%	23,1%
CA0304 e AN0607	46,2%	53,8%	46,2%	50,0%	0,0%	0,0%	7,7%	7,7%
CA0304 e ES0607	46,2%	46,2%	41,0%	46,2%	0,0%	0,0%	0,0%	3,8%
ES0304 e AN0506	76,9%	53,8%	48,7%	57,7%	15,4%	7,7%	10,3%	23,1%
ES0304 e ES0506	53,8%	50,0%	51,3%	51,9%	23,1%	19,2%	33,3%	32,7%
ES0304 e AN0607	61,5%	46,2%	46,2%	53,8%	0,0%	0,0%	0,0%	7,7%
ES0304 e ES0607	53,8%	50,0%	43,6%	50,0%	0,0%	3,8%	5,1%	11,5%
AN0506 e ES0506	69,2%	73,1%	64,1%	65,4%	23,1%	19,2%	20,5%	28,8%
AN0506 e AN0607	76,9%	76,9%	69,2%	71,2%	0,0%	3,8%	7,7%	5,8%
AN0506 e ES0607	61,5%	65,4%	56,4%	65,4%	0,0%	3,8%	7,7%	9,6%
ES0506 e AN0607	53,8%	61,5%	71,8%	61,5%	7,7%	3,8%	10,3%	11,5%
ES0506 e ES0607	76,9%	69,2%	61,5%	63,5%	0,0%	3,8%	7,7%	11,5%
AN0607 e ES0607	61,5%	69,2%	61,5%	61,5%	30,8%	42,3%	46,2%	48,1%
TODOS	23,1%	19,2%	17,9%	19,2%	0,0%	0,0%	0,0%	0,0%

Tabela 10 – Porcentagem das progênies coincidentes entre ambientes ao selecionar 5, 10, 15 e 20% das progênies mais produtivas pelas duas abordagens de análise individuais

látice × blocos	5% (13)	10% (26)	15% (39)	20% (52)
AN0203	100%	96,2%	97,4%	100%
AR0203	92,3%	92,3%	97,4%	92,3%
CA0203	92,3%	96,2%	97,4%	96,2%
ES0203	84,6%	88,5%	89,7%	98,1%
AN0304	100%	92,3%	87,2%	84,6%
CA0304	100%	100%	94,9%	98,1%
ES0304	92,3%	88,5%	89,7%	90,4%
AN0506	100%	100%	100%	100%
ES0506	100%	96,2%	97,4%	98,1%
AN0607	100%	92,3%	100,0%	92,3%
ES0607	92,3%	96,2%	92,3%	88,5%

4.5.1 Correlação entre os EBLUPs obtidos por ambiente nas diferentes estruturas de análise

A correlação é considerada como uma medida de influência mútua ou conjunta entre variáveis. Foi utilizado o coeficiente de correlação não-paramétrico de Spearman (GIBBONS; CHAKRABORTI, 2003), pois como algumas das variáveis não têm distribuição Normal, é necessário utilizar esse coeficiente para podermos testar sua significância e para verificar a concordância quanto à classificação das progênies. Este coeficiente varia de -1 a 1, e quanto mais próximo desses valores mais forte é a correlação entre duas variáveis. Se for positivo, significa que as variáveis têm o mesmo comportamento: se uma aumenta, a outra também aumenta; caso contrário, se uma variável aumenta a outra diminui.

A correlação entre os EBLUPs obtidos pelos diferentes ambientes é de natureza genotípica. De modo geral, as correlações genotípicas foram superiores às de ambiente, evidenciando a maior influência das causas genéticas na determinação da correlação fenotípica, que é provocada pelos fatores genéticos e ambientais.

As Figuras 8, 9, 10 e 11 apresentam os coeficientes de correlação de Spearman entre as variáveis. Quanto mais próxima de uma elipse for a forma das figuras, mais forte é a correlação entre as duas variáveis correspondentes no cruzamento das linhas e colunas. E quanto mais forte o azul ou o vermelho da cor, também mais forte é a correlação.

Respeitando a estrutura do látice quadrado e analisando as correlações entre os EBLUPs dos 11 ambientes de experimentação comparados dois a dois nota-se que pela análise individual, todos os ambientes apresentaram correlações positivas entre si, destacando-se com maiores índices aqueles ambientes de mesmo local em anos diferentes, como pode ser visto na Figura 8.

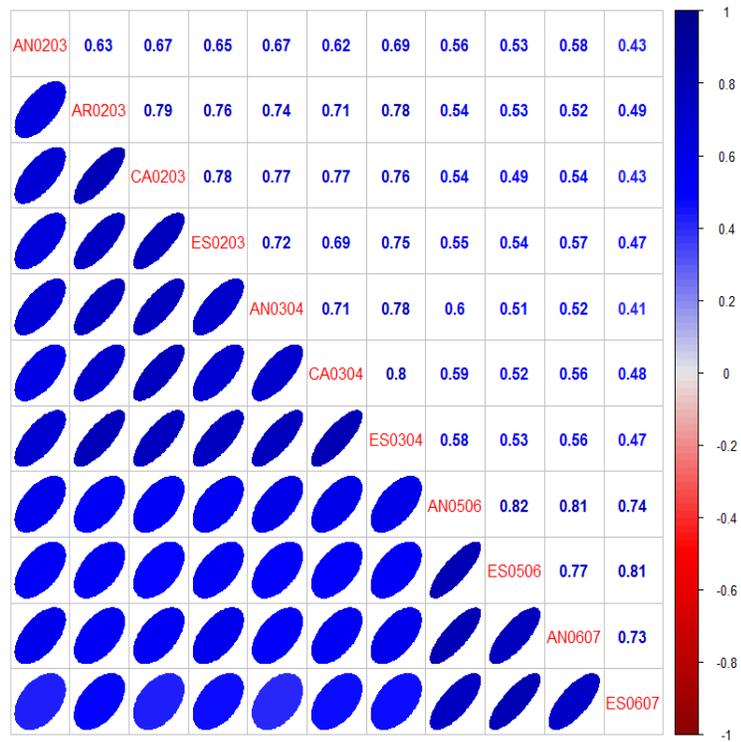


Figura 8 – Correlação de Spearman entre os EBLUPs obtidos pelo modelo em cada um dos ambientes nas análises individuais considerando-se a estrutura do látice quadrado

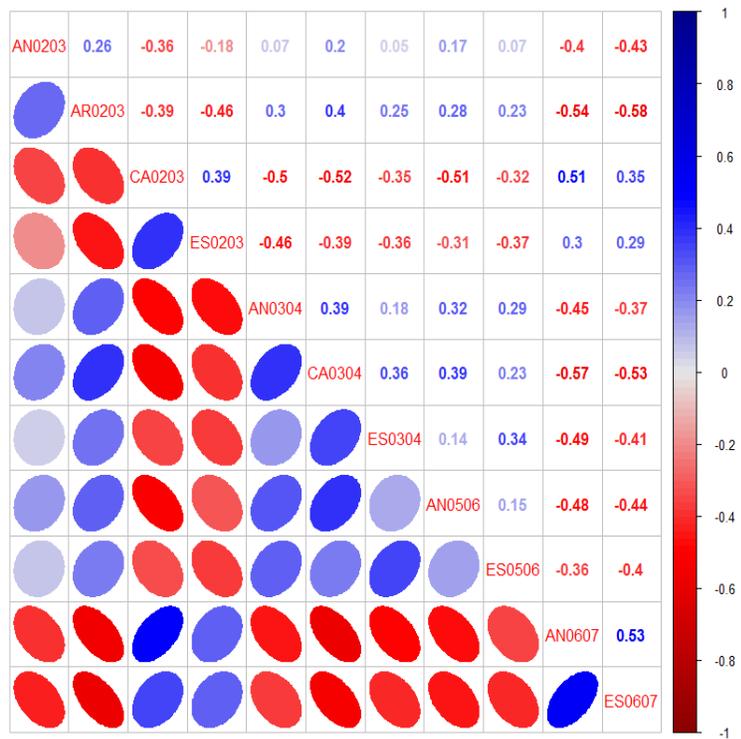


Figura 9 – Correlação de Spearman entre os EBLUPs obtidos pelo modelo em cada um dos ambientes nas análises conjuntas considerando-se a estrutura do látice quadrado

Já na Figura 9 são apresentados os coeficiente de correlação obtidos pela análise conjunta dos experimentos analisados em látice quadrado. Nota-se um comportamento totalmente diferente daqueles obtidos nas análises individuais. Uma provável explicação para este fato deve-se à interação genótipo \times ambiente ter sido altamente significativa, o que indica grandes diferenças no desempenho individual dos materiais avaliados nos diferentes ambientes.

Tomando-se como exemplo, os ambientes AN0203 e AR0203 verifica-se que a estimativa do coeficiente de correlação entre esses ambientes na análise individual foi 0,63 e na conjunta 0,26, mostrando que o ambiente pode interferir, modificando o grau de associação entre os ambientes.

As correlações de todos ambientes, na análise em que se considera cada repetição do látice como um bloco completo, foram pouco menores ou iguais em alguns poucos casos em relação à primeira análise, tanto nas correlações entre EBLUPs provenientes das análises individuais ou da conjunta, confirmando mais uma vez a forte interação genótipo \times ambiente.

Considerando-se os mesmos ambientes comparados na estrutura de análise anterior, AN0203 e AR0203, verifica-se que a estimativa do coeficiente de correlação entre esses ambientes na análise individual foi 0,61 e na conjunta 0,23, coeficientes menores dos registrados na abordagem anterior. Cecon (1992) chegou à conclusão de que, quando um experimento é montado na estrutura de látice com o objetivo de estimar parâmetros genéticos a análise mais indicada é a análise do látice quadrado como um látice, independentemente de sua eficiência, e não como blocos casualizados.

Quando os sinais dos coeficientes de correlação genotípica e de ambiente são opostos, implica a existência de dificuldades para se efetuar uma seleção eficiente (SCAPIM, 1994). Grande parte das combinações envolvendo o caráter peso de grãos apresentaram sinais diferentes entre as correlações genotípicas e de ambiente.

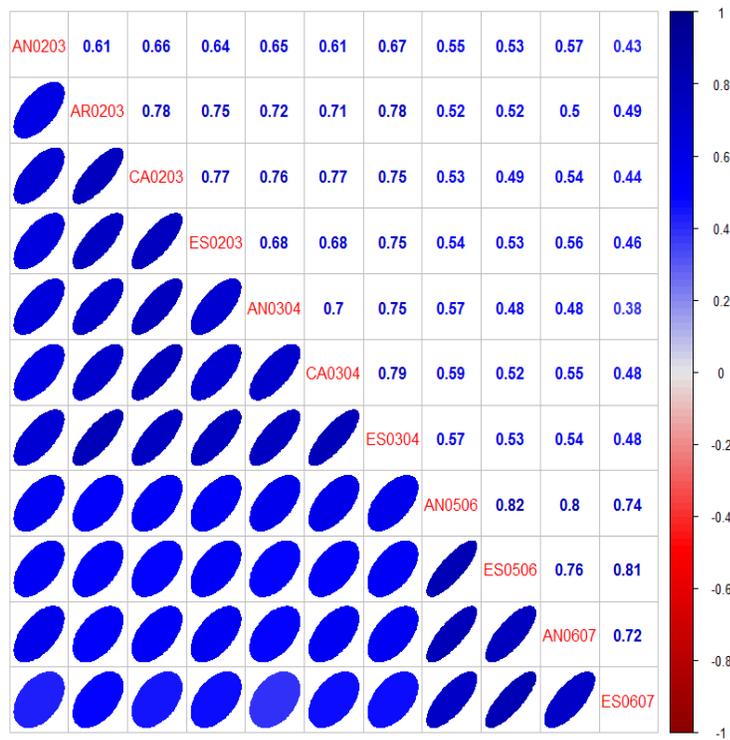


Figura 10 – Correlação de Spearman entre os EBLUPs obtidos pelo modelo em cada um dos ambientes nas análises individuais considerando-se a estrutura de blocos completos

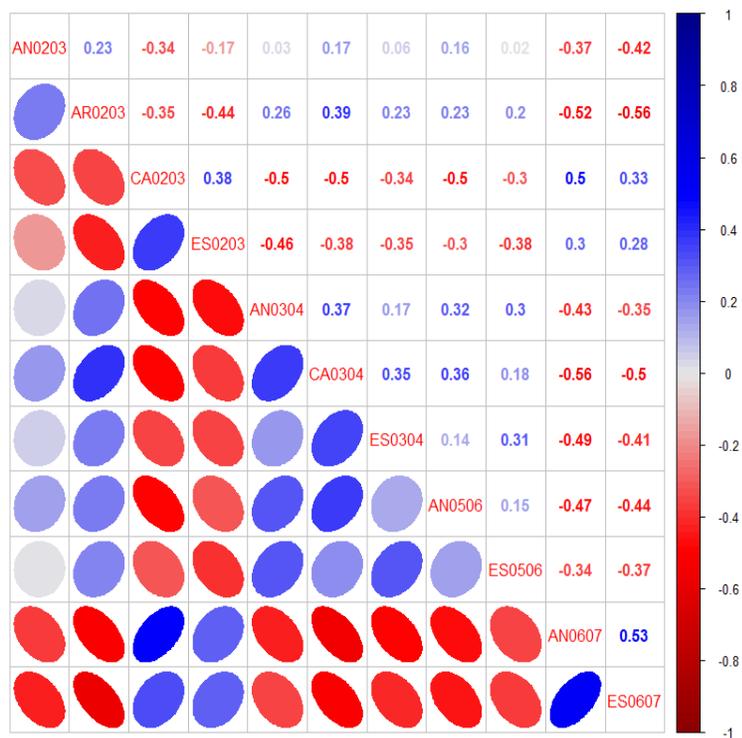


Figura 11 – Correlação de Spearman entre os EBLUPs obtidos pelo modelo em cada um dos ambientes nas análises conjuntas considerando-se a estrutura de blocos completos

5 CONSIDERAÇÕES FINAIS

A análise de experimentos instalados em delineamento látice quadrado por modelos mistos é uma maneira eficiente de determinar a contribuição da interação genótipo \times ambiente em programas de melhoramento genético com elevado número de materiais genéticos distintos e auxilia na decisão dos melhoristas visando a seleção de progênies que ainda serão avaliadas em etapas futuras.

Por meio desse trabalho foi possível confirmar que a utilização do método da máxima verossimilhança restrita (REML) é também uma alternativa para a estimação dos componentes de variância em experimentos de melhoramento genético. A partir dos componentes de variância é possível fazer então a predição de valores genotípicos no melhoramento vegetal de espécies anuais. Esse método era consolidado apenas no melhoramento animal ou no caso do melhoramento vegetal, apenas em espécies perenes, como o eucalipto.

Mesmo sendo pequena a diferença nos resultados entre as duas abordagens de análise do conjunto de dados, a análise em látice quadrado mostrou-se superior à abordagem dos blocos completos, de acordo com os gráficos de diagnósticos obtidos.

A interação é o principal entrave ao selecionar as progênies mais produtivas, pois, se a mesma existir, não será possível realizar uma recomendação de forma generalizada, exigindo adoção de medidas que controlem ou minimizem os efeitos dessa interação. Existem pelo menos três alternativas para se atenuar o efeito da interação genótipo \times ambiente, ou seja, identificar as progênies específicas para cada ambiente, realizar o zoneamento ecológico e identificar as progênies com maior estabilidade fenotípica.

Quando se dispõe de progênies avaliadas em vários ambientes, o resultado desta nada mais é do que uma tabela de dupla entrada na qual se tem as estimativas das médias dos genótipos por ambiente. De posse desses resultados o objetivo dos programas de melhoramento é identificar aquelas com maior estabilidade fenotípica ou até mesmo os genótipos com adaptação a ambientes específicos.

Análises complementares que seriam interessantes para esse conjunto de dados inclui a modelagem de outros caracteres e a estimativa das correlações entre eles por ser um aspecto genético de grande interesse para o melhoramento de plantas, e por isso deve receber atenção especial dos melhoristas. Uma outra sugestão é explorar diferentes matrizes de variâncias e covariâncias para inserir informações de parentesco entre as progênies com o auxílio de outros softwares ou mesmo com os pacotes ASReml-R ou nlme que contém diversas opções para as estruturas dessas matrizes.

REFERÊNCIAS

- ALCARDE, R. **Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R**. 2012. 156 p. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 2012.
- BARBIN, D. **Componentes de Variância: teoria e aplicações**. Piracicaba: FEALQ, 1993. 117 p.
- BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. **Fitting Linear Mixed-Effects Models Using lme4**. Vienna, 2015. Disponível em: <<https://cran.r-project.org/web/packages/lme4/lme4.pdf>>. Acesso em: 28 abr. 2016.
- BOSE, R.C.; NAIR, K.R. Partially balanced incomplete block designs. **Sankhya: The Indian Journal of Statistics**, Kolkata, v. 4, n. 3, p. 337–372, 1939.
- BUENO-FILHO, J.S.S.; VENCOSKY, R. Alternativas de análise de ensaios em látice no melhoramento vegetal. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 2, p. 259–269, 2000.
- BURDICK, R.K.; GRAYBILL, F.A. **Confidence intervals on variance components**. New York: Marcel Dekker, 1992. v. 127. 211 p.
- CAMARINHA-FILHO, J.A. **Modelos lineares mistos: estruturas de matrizes de variâncias e covariâncias e seleção de modelos**. 2002. 85 p. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 2002.
- CECON, P.R. **Alternativas de análise de experimentos em látice e aplicações no melhoramento vegetal**. 1992. 109 p. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 1992.
- COCHRAN, W.G.; COX, G.M. **Experimental designs**. New York: Wiley, 1957. 611 p.
- COELHO, A.M. **Comparação de métodos de estimação de componentes de variância e parâmetros genéticos considerando o delineamento III aplicado a caracteres quantitativos em milho**. 2010. 101 p. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 2010.
- COMPANHIA NACIONAL DE ABASTECIMENTO. **Séries Históricas**. Brasília: 2016. Disponível em: <http://www.conab.gov.br/conteudos.php?a=1252&ordem=produto&Pagina_objcmsconteudos=3#A_objcmsconteudos>. Acesso em: 04 maio 2016.
- COMSTOCK, R.E. Quantitative genetics and maize breeding. **Maize Breeding and Genetics**, New York, v. 13, p. 191–206, 1978.
- CUSTÓDIO, T.N.; BARBIN, D. Comparação de modelos mistos visando à estimação do coeficiente de herdabilidade para dados de proporção. **Revista de Matemática e Estatística**, Jaboticabal, v. 23, p. 23–31, 2005.
- DEMPSTER, A.P.; N.M., Laird.; D.B., Rubin. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistic Society**, London, v. 39, p. 1–38, 1977.

DESSOTTI, C.; PIEDADE, S.M.S.; BARBIN, D.; SANCHES, P.F.; BAUTISTA, E.A.L. Comparação de médias de tratamentos em grupos de experimentos. **Revista Brasileira de Biometria**, Jaboticabal, v. 31, p. 327–336, 2013.

DUARTE, J.; VENCOVSKY, R. Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. **Scientia Agricola**, Piracicaba, v. 58, n. 1, p. 109–117, 2001.

DUDLEY, J.W.; MOLL, R.H. Interpretation and use of estimation of heritability and genetic variances in plant breeding. **Crop Science**, Madison, v. 2, n. 3, p. 257–262, 1969.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. London: Longman Press, 1996. 464 p.

FANCELLI, A.L.; DOURADO-NETO, D. **Produção de milho**. Guaíba: Guaíba Agropecuária, 2004. 360 p.

FRITSCHÉ-NETO, R. **Predição de valores genotípicos de híbridos de milho com desbalançamentos de genótipos e ambientes**. 2008. 71 p. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 2008.

FRITSCHÉ-NETO, R.; GONÇALVES, M.C.; VENCOVSKY, R.; SOUZA-JÚNIOR, C.L. Prediction of genotypic values of maize hybrids in unbalanced experiments. **Crop Breeding and Applied Biotechnology**, Viçosa, MG, v. 10, p. 32–39, 2010.

GIBBONS, J.D.; CHAKRABORTI, S. **Nonparametric Statistical Inference**. 4ª. ed. New York: Marcel Dekker, 2003. 654 p.

GILMOUR, A.R.; THOMPSON, R.; CULLIS, B.R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Hoboken, v. 51, n. 4, p. 1440–1450, 1995.

GOGEL, B.J.; CULLIS, B.R.; VERBYLA, A.P. REML estimation of multiplicative effects in multi-environment variety trials. **Biometrics**, Hoboken, v. 51, n. 4, p. 744–749, 1995.

HARVILLE, D.A. Maximum likelihood approaches to variance component estimation and to related problems. **Journal of the American Statistical Association**, Hoboken, v. 72, n. 358, p. 320–338, 1977.

HENDERSON, C.R. **Applications of linear model in animal breeding**. Ontário, 1984. Disponível em: <<ftp://tech.obihiro.ac.jp/suzuki/Henderson.pdf>>. Acesso em: 24 out. 2015.

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**, Hoboken, v. 15, n. 2, p. 192–218, 1959.

HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. Hoboken: John Wiley, 2005. v.2. 736 p.

IEMMA, M. **Uso do melhor preditor linear não viesado (BLUP) em análises dialéticas e predição de híbridos**. 2003. 81 p. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 2003.

- JUGENHEIMER, R.W. **Corn improvement, seed production and uses**. Hoboken: John Wiley, 1976. 670 p.
- KEMPTON, R.A.; FOX, P.N.; CERESO, M. **Statistical methods for plant variety evaluation**. London: Chapman & Hall, 1997. 191 p.
- LEITE, M.S.O. **Análise multiambientes visando a recomendação regionalizada de clones de cana-de-açúcar**. 2011. 72 p. Tese (Tese de doutorado) — Universidade Federal de Viçosa, Viçosa, MG, 2011.
- MAURO, A.O. **Adaptabilidade, estabilidade e ganho genético com o processo seletivo em soja (*Glycine max L.*) em Ponta Porã, Mato Grosso do Sul**. 1991. 192 p. Tese (Tese de doutorado) — Universidade Federal de Viçosa, Viçosa, MG, 1991.
- MENDIBURU, F. **agricolae: Statistical Procedures for Agricultural Research**. 2015. R package version 1.2-3. Disponível em: <<http://CRAN.R-project.org/package=agricolae>>. Acesso em: 24 out. 2015.
- MOORE, K.J.; DIXON, P.M. Analysis of combined experiments revisited. **Agronomy Journal**, Madison, v. 107, n. 2, p. 763–771, 2015.
- MORAL, R. A.; HINDE, J.; DEMETRIO, C.G.B. **hnp: Half-Normal Plots with Simulation Envelopes**. 2016. R package version 1.2. Disponível em: <<http://CRAN.R-project.org/package=hnp>>. Acesso em: 29 jan. 2016.
- MORO, G.V.; SANTOS, M.F.; BENTO, D.A.V.; AGUIAR, A.M; SOUZA-JÚNIOR, C.L. Genetic analysis of kernel oil content in tropical maize with design III and qtl mapping. **Euphytica**, Wageningen, v. 185, p. 419–428, 2012.
- PATERNIANI, E. **Avaliação do método de seleção entre e dentro de famílias de meio-irmãos no melhoramento de milho *Zea Mays L.*** 1968. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 1968.
- PATERNIANI, E.; VIEGAS, G.P. **Melhoramento e produção do milho**. Campinas: Fundação Cargill, 1987. 795 p.
- PATTERSON, H.D.; THOMPSON, R. Recovery of interblock information when blocks sizes are unequal. **Biometrika**, Oxford, UK, v. 58, p. 545–554, 1971.
- PEREIRA-NETO, R. **Estudo comparativo de métodos de estimação da variância de coeficientes de herdabilidade**. 1994. 80 p. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 1994.
- PETERSEN, R.G. **Agricultural fiels experiments design and analysis**. New York: New York Marcel Dekker, 1994. 409 p.
- PIEPHO, H.P.; MÖHRING, J. Computing heritability and selection response from unbalanced plant breeding trials. **Genetics**, Bethesda, v. 177, n. 3, p. 1881–1888, 2007.
- PIEPHO, H.P.; MÖHRING, J.; MELCHINGER, A.E.; BÜCHSE, A. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, n. 1-2, p. 209–228, 2007.
- PIMENTEL-GOMES, F. **Curso de estatística experimental**. Piracicaba: FEALQ, 2009. 451 p.

PIMENTEL-GOMES, F.; GARCIA, C.H. **Estatística aplicada a experimentos agrônômicos e florestais**. Piracicaba: FEALQ, 2002. 309 p.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2015. 3501 p.

RAMALHO, M.A.P. **Eficiência relativa de alguns processos de seleção intrapopulacional no milho baseados em famílias não endógamas**. 1977. 121 p. Tese (Tese de doutorado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 1977.

RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. Book. **A experimentação em genética e melhoramento de plantas**. Lavras: Lavras UFLA, 2005. 322 p.

REGAZZI, A.J.; SILVA, H.D.; VIANA, J.M.S.; CRUZ, C.D. Análise de experimentos em látice quadrado com ênfase em componentes de variância. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 11, p. 1987–1997, 1999.

RESENDE, M.D.V; PRATES, D.F.; YAMADA, C.K.; JESUS, A. Estimação de componentes de variância e predição de valores genéticos pelo método da máxima verossimilhança restrita (reml) e melhor predição linear não viciada (blup) em *Pinus*. **Boletim de Pesquisa Florestal**, Curitiba, n. 32/33, p. 23–42, 1996.

RIOS, M.C.D. **Alternativas de prodecimentos estatísticos para avaliação de genótipos em programas de melhoramento genético vegetal**. 1997. 59 p. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo, Piracicaba, 1997.

SANTOS, A. **Experimentos casualizados em blocos incompletos com tratamentos comuns e testemunha intercalar**. Exame de qualificação. Piracicaba. 2015.

SCAPIM, C.A. **Cruzamentos dialéticos entre sete variedades de milho doce (*Zea mays*L.) e correlações entre caracteres agrônômicos**. 1994. 96 p. Dissertação (Dissertação de Mestrado) — Universidade Federal de Viçosa, Viçosa, MG, 1994.

SCAPIM, C.A.; CARVALHO, C.G.; CRUZ, C.D. Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira**, Brasília, v. 30, n. 5, p. 683–686, 1995.

SEARLE, S.R.; CASELLA, G.; MCCULLOCH, E. Book. **Variance Components**. Hoboken: John Wiley & Sons, 2006. 501 p.

SELF, S.G.; LIANG, K-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. **Journal of the American Statistical Association**, Hoboken, v. 82, n. 398, p. 605–610, 1987.

SILVA, H.D. **Análise de experimentos em látice quadrado ("Square lattice) com ênfase em componentes de variância e aplicações no melhoramento genético vegetal**. 1997. 139 p. Tese (Tese de doutorado) — Universidade Federal de Viçosa, Viçosa, MG, 1997.

SILVA, H.D.; FERREIRA, D.F.; PACHECO, C.A.P. Avaliação de quatro alternativas de análise de experimentos em látice quadrado, quanto à estimação de componentes de variância. **Bragantia**, Campinas, v. 59, n. 1, p. 117–123, 2000.

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

SOLFERINI, O.B. **Milho alto óleo - parâmetros genéticos, correlações entre modelos fixo e misto e seleção**. 2010. 145 p. Tese (Tese de doutorado) — Universidade Estadual Paulista, Jaboticabal, 2010.

SOLOGUREN, L. Demanda mundial cresce e Brasil tem espaço para expandir produção. **Visão Agrícola**, Piracicaba, v. 13, p. 8–11, 2015.

SOUZA-JÚNIOR, C.L.; VENCOVSKY, R. Covariância entre parentes na presença da interação genótipo-ambiente. In: SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 3., Lavras. **Resumos...** Lavras, ESAL, 1989. Resumo 34.

TOMÉ, M.A.; BUENO-FILHO, J.S.S.; FERREIRA, D.F. Blocos fixos ou aleatórios? O caso dos ensaios em látice no melhoramento vegetal. **Ciência e Agrotecnologia**, Lavras, v. 26, n. 3, p. 576–584, 2002.

UNITED STATES DEPARTMENT OF AGRICULTURE. **World corn production, consumption and stocks**. Washington, D.C., 2016. Disponível em: <<http://apps.fas.usda.gov/psdonline/>>. Acesso em: 03 maio 2016.

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

VERONESI, J.A. **Comparações de métodos e avaliação de adaptabilidade e estabilidade de milho (*Zea mays* L.) em dez ambientes do estado de Minas Gerais**. 1995. 90 p. Dissertação (Dissertação de Mestrado) — Universidade Federal de Viçosa, Viçosa, MG, 1995.

WARDYN, B.M.; EDWARDS, J.W.; LAMKEY, K.R. The genetic structure of a maize population: the role of dominance. **Crop Science**, Madison, v. 47, p. 467–476, 2007.

YATES, F. A new method of arranging variety trials involving a large number of varieties. **Journal of Agricultural Science**, Cambridge, v. 26, n. 3, p. 424–445, 1936.

YATES, F. The recovery of inter-block information in variety trials arranged in three-dimensional lattices. **Annals of Eugenics**, London, v. 9, n. 2, p. 136–156, 1939.

ZIMMERMANN, F.J.P. **Estatística aplicada à pesquisa agrícola**. Brasília: Brasília Embrapa, 2014. 582 p.

ANEXOS

ANEXO A - COMANDOS PARA O AJUSTE DOS MODELOS - SOFTWARE R

```

library(lme4) # pacote utilizado para ajuste dos modelos mistos
library(hnp) # pacote utilizado para os half-normal plots

## 1) Leitura dos dados:
# O arquivo com o conjunto de dados é da extensão .csv com as seguinte
  colunas:
#Amb, Rep, Bloco, Progenie e PGCHA
dados = read.csv2('progeniesmilho.csv')
dados$Rep = as.factor(dados$Rep)
dados$Bloco = as.factor(dados$Bloco)
dados$Progenie = as.factor(dados$Progenie)
str(dados)

## 2) Descritiva
#avaliação por Ambiente
desv.padr = tapply(dados$PGCHA,dados$Local,sd)
var = tapply(dados$PGCHA,dados$Local,var)
media = tapply(dados$PGCHA,dados$Local,mean)
maximo = tapply(dados$PGCHA,dados$Local,max)
minimo = tapply(dados$PGCHA,dados$Local,min)
f1=function(x) max(x)-min(x)
amplitude = tapply(dados$PGCHA,dados$Local,f1)

resumo=cbind(media, desv.padr, maximo, minimo, amplitude)

colnames(resumo)=c("média", "d.p.", "máximo", "mínimo", "Amplitude")
rownames(resumo)=c('AN0203', 'AR0203', 'CA0203', 'ES0203', 'AN0304', 'CA0304',
  'ES0304', 'AN0506', 'ES0506', 'AN0607', 'ES0607')
round(resumo,2)

## 3) Ajuste dos modelos individuais - análise como látice
#####
#   MODELOS PGCHA - análises individuais - Local AN0203(1)   #
#####
mmod1_PGCHAamb1 = lmer(PGCHA ~ (1|Rep) + (1|Rep:Bloco) + (1|Progenie),

```

```

data=dados[dados$Amb=='AN0203',]) #trocar código do ambiente para os
demais
summary(mmod1_PGCHAAmb1) #para extrair os componentes de variância
hnp(residuals(mmod1_PGCHAAmb1), xlab='Quantis teóricos', ylab='Resíduos
condicionais', scale=T, half=F, print=T)

# Cálculo do coeficiente de herdabilidade
herdAmb1 = VarCorr(mmod1_PGCHAAmb1)[[1]][1]/(VarCorr(mmod1_PGCHAAmb1)
[[1]][1] + summary(mmod1_PGCHAAmb1)$sigma^2)

# Estimação dos EBLUPs dos modelos individuais
(amb1_Progenie<- ranef(mmod1_PGCHAAmb1)$Progenie[,1])
names(amb1_Progenie)<- rownames(ranef(mmod1_PGCHAAmb1)$Progenie)
amb1_Progenie[order(amb1_Progenie, decreasing=T)][1:52]

## 4) Ajuste do modelo conjunto - análise como látice
#####
#   MODELOS PGCHA - ANÁLISE CONJUNTA - todos ambientes   #
#####
mmodc1_PGCHA = lmer(PGCHA ~ (1|Amb) + (1|Amb:Rep) + (1|Amb:Rep:Bloco) +
(1|Progenie) + (1|Amb:Progenie), data=dados)
summary(mmodc1_PGCHA) #para extrair os componentes de variância
hnp(residuals(mmodc1_PGCHA), xlab='Quantis teóricos', ylab='Resíduos
condicionais', scale=T, half=F, print=T)

## teste da razão de verossimilhança para a interação Amb:Progenie
mmodc2 <- lmer(PGCHA ~ (1|Amb) + (1|Amb:Rep) + (1|Amb:Rep:Bloco) + (1|
Progenie), data=dados) # modelo sem a interação Amb:Progenie
summary(mmodc2)
anova(mmodc1_PGCHA, mmodc2)

# Cálculo do coeficiente de herdabilidade
herdG = VarCorr(mmodc1_PGCHA)[[2]][1]/(VarCorr(mmodc1_PGCHA)[[2]][1] +
VarCorr(mmodc1_PGCHA)[[1]][1] + summary(mmodc1_PGCHA)$sigma^2)

# Estimação dos EBLUPs do modelo conjunto
names(ranef(mmodc1_PGCHA))
(ambLC_Progenie<- ranef(mmodc1_PGCHA)$"Amb:Progenie")

```

```

LC_Progenie<- ambLC_Progenie[1:256,1]
names(LC_Progenie)<- c(1:256)
LC_Progenie[order(LC_Progenie, decreasing=T)][1:52] #ordenação

## 5) Ajuste dos modelos individuais - análise como blocos completos
#####
#   MODELOS PGCHA - análises individuais - Local AN0203(1)   #
#####
mmod2_PGCHAAmb1 = lmer(PGCHA ~ (1|Rep) + (1|Progenie), data=dados[dados$
  Amb=='AN0203',]) #trocar código do ambiente para os demais
summary(mmod2_PGCHAAmb1) #para extrair os componentes de variância
hnp(residuals(mmod2_PGCHAAmb1), xlab='Quantis teóricos', ylab='Resíduos
  condicionais', scale=T, half=F, print=T)

# Cálculo do coeficiente de herdabilidade
herdAmb1 = VarCorr(mmod2_PGCHAAmb1)[[1]][1]/(VarCorr(mmod2_PGCHAAmb1)
  [[1]][1] + summary(mmod2_PGCHAAmb1)$sigma^2)

# Estimação dos EBLUPs dos modelos individuais
(amb1_Progenie<- ranef(mmod2_PGCHAAmb1)$Progenie[,1])
names(amb1_Progenie)<- rownames(ranef(mmod1_PGCHAAmb1)$Progenie)
amb1_Progenie[order(amb1_Progenie, decreasing=T)][1:52]

## 6) Ajuste do modelo conjunto - análise como blocos completos
#####
#   MODELOS PGCHA - ANÁLISE CONJUNTA - todos ambientes      #
#####
mmod2_PGCHAAmb11 = lmer(PGCHA ~ (1|Rep) + (1|Progenie), data=dados[dados
  $Amb=='ES0607',])
summary(mmod2_PGCHAAmb11) #para extrair os componentes de variância
hnp(residuals(mmodc2_PGCHA), xlab='Quantis teóricos', ylab='Resíduos
  condicionais', scale=T, half=F, print=T)

## teste da razão de verossimilhança para a interação Amb:Progenie
mmodc2 <- lmer(PGCHA ~ (1|Amb) + (1|Amb:Rep) + (1|Amb:Rep:Bloco) + (1|
  Progenie), data=dados) # modelo sem a interação Amb:Progenie
summary(mmodc2)

```

```
anova(mmodc2_PGCHA, mmodc2)

# Cálculo do coeficiente de herdabilidade
herdG = VarCorr(mmodc2_PGCHA)[[2]][1]/(VarCorr(mmodc2_PGCHA)[[2]][1] +
  VarCorr(mmodc2_PGCHA)[[1]][1] + summary(mmodc2_PGCHA)$sigma^2)

# Estimação dos EBLUPs do modelo conjunto
names(ranef(mmodc2_PGCHA))
(ambBC_Progenie<- ranef(mmodc2_PGCHA)$"Amb:Progenie")
BC_Progenie<- ambBC_Progenie[1:256,1]
names(BC_Progenie)<- c(1:256)
BC_Progenie[order(BC_Progenie, decreasing=T)][1:52] #ordenação
```

ANEXO B - COMANDOS PARA OS GRÁFICOS DE CORRELAÇÃO - SOFTWARE R

```
library(corrplot)
library(Hmisc)

## 1) Leitura dos dados:
dados = read.csv2('eblupsobtidos.csv')
str(dados)

variaveis = subset(dados, select=c(Amb1,Amb2,Amb3,Amb4,Amb5,Amb6,Amb7,
  Amb8,Amb9,Amb10,Amb11)) # coloque aqui as variáveis que você quer
  fazer a correlação

#names(variaveis) = c('AN0203','AR0203','CA0203','ES0203','AN0304','
  CA0304','ES0304','AN0506','ES0506','AN0607','ES0607') # use isso pra
  aparecer o nome das variáveis que você quer no gráfico

correl = round(rcorr(as.matrix(variaveis),type='pearson')$r,4) # tem
  Spearman e Pearson pra escolher no 'type'

#Você só pode testar se o coeficiente de Pearson é significativo se as
  duas variáveis seguirem distribuição normal. Se não for tem que usar
  o coeficiente de Spearman.

correl = round(rcorr(as.matrix(variaveis),type='spearman')$r,4) # tem
  Spearman e Pearson pra escolher no 'type'

col2 = colorRampPalette(c('darkred','red',grey(0.9),'blue','darkblue'))
  # não mexer

corrplot.mixed(correl,lower='ellipse',upper='number',col=col2(100))

round(rcorr(as.matrix(variaveis),type='spearman')$P,4) # testando a
  significância das correlações
```