

**ANÁLISE ESPACIAL E DE VIZINHANÇA NO  
MELHORAMENTO GENÉTICO DE PLANTAS**

**JANAINA RIBEIRO COSTA**

**2003**

**JANAINA RIBEIRO COSTA**

**ANÁLISE ESPACIAL E DE VIZINHANÇA NO  
MELHORAMENTO GENÉTICO DE PLANTAS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Agronomia, área de concentração em Genética e Melhoramento de Plantas, para obtenção do título de "Doutor".

Orientador

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

LAVRAS  
MINAS GERAIS - BRASIL  
2003

Costa, Janaina Ribeiro

Análise espacial e de vizinhança no melhoramento genético de plantas /  
Janaina Ribeiro Costa. -- Lavras : UFLA, 2003.

82 p. : il.

Orientador: Julio Silvio de Souza Bueno Filho.

Tese (Doutorado) – UFLA.

Bibliografia.

1. Melhoramento genético vegetal. 2. Análise espacial. 3. Método das  
médias móveis. 4. Método de Papadakis. I. Universidade Federal de Lavras. II.  
Título.

CDD -631.523

**JANAINA RIBEIRO COSTA**

**ANÁLISE ESPACIAL E DE VIZINHANÇA NO  
MELHORAMENTO GENÉTICO DE PLANTAS**

Tese apresentada à Universidade Federal de Lavras,  
como parte das exigências do Programa de  
Agronomia, área de concentração em Genética e  
Melhoramento de Plantas, para obtenção do título  
de "Doutor".

Aprovada em 7 de fevereiro de 2003.

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho	UFLA
Prof. Dr. Magno Antônio Patto Ramalho	UFLA
Prof. Dr. Eduardo Bearzoti	UFLA
Prof <sup>ª</sup> . Dr <sup>ª</sup> . Elaine Aparecida de Souza	UFLA
Prof. Dr. Antônio Carlos de Oliveira	Embrapa -Milho e Sorgo

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

## SUMÁRIO

	Página
<b>RESUMO</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>1 INTRODUÇÃO</b> .....	1
<b>2 REFERENCIAL TEÓRICO</b> .....	3
2.1 Princípios básicos da experimentação.....	3
2.2 Tipos de experimentos mais usados e m genética e melhoramento de plantas.....	5
2.3 Análise de modelos mistos.....	10
2.4 Modelagens espaciais.....	22
2.4.1 Análise de vizinhança: método de Papadakis.....	23
2.4.2 Análise de vizinhança: método das médias móveis.....	25
2.4.3 Análise espacial.....	26
<b>3 MATERIAL E MÉTODOS</b> .....	32
3.1 Experimentos utilizados.....	32
3.2 Local e condução.....	34
3.3 Formas de análise.....	34
3.3.1 Modelo linear sem controle local: delineamento inteiramente casualizado (DIC).	35
3.3.2 Modelo linear com controle local: delineamento e m blocos casualizados completos (DBC).....	36
3.3.3 Modelo linear com controle local: delineamento em látice.....	36
3.3.4 Modelagem de Gauss-Markov Normal (GMN), considerando erros independentes.....	37
3.3.5 Modelagem da análise espacial (AE), considerando erros dependentes.....	38
3.3.6 Modelagem das médias móveis (MM).....	38
3.3.7 Modelagem de Papadakis (PPD).....	39
3.4 Modelo misto.....	40

3.5 Avaliação das formas de análise.....	44
3.6 Programa computacional.....	47
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>49</b>
<b>5 CONCLUSÕES.....</b>	<b>70</b>
<b>6 REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>71</b>
<b>ANEXOS.....</b>	<b>76</b>

## RESUMO

COSTA, Janaina Ribeiro. **Análise espacial e de vizinhança no melhoramento genético de plantas**. 2003. 82 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, MG\*.

O objetivo deste trabalho foi procurar alternativas para melhorar a eficiência dos experimentos de avaliação de progênies de feijão e milho. O foco foi em técnicas analíticas que podem melhorar a precisão experimental. Dados de produção de grãos foram submetidos a diferentes tipos de análise, ignorando ou considerando o controle local do delineamento. Para cada análise, quatro tipos de modelos foram usados: Gauss-Markov normal (GMN), médias móveis nos dados brutos (MM), médias móveis nos dados residuais (Papadakis-PPD) e análise espacial modelando covariâncias residuais (AE). Também foram verificadas as conseqüências de considerar os efeitos de tratamentos como sendo fixos ou aleatórios na modelagem para ranqueamento dos genótipos e para seleção. A comparação dos diferentes tipos de análise foi baseada nas suas eficiências. Foram calculadas as variâncias residuais ( $\hat{\sigma}^2$ ), as variâncias de progênies ( $\hat{\sigma}_p^2$ ), as semi-amplitudes (SA) relativas dos intervalos de confiança de  $\hat{\sigma}_p^2$  e os critérios de informação de Akaike (AIC). Os resultados mostraram que ignorar o delineamento experimental em geral não é efetivo, com a ajuda da informação espacial. Os modelos MM e PPD em muitos casos melhoraram o modelo original justificado pelo delineamento, embora a AE não o tenha melhorado. Ademais, houve diferenças no ranqueamento de progênies quando representadas por efeitos aleatórios (comparadas à usual recuperação da informação interbloco, com efeitos fixos para tratamentos). A AE, se inefetiva, não mudou as estimativas dos componentes de variância. Esta propriedade garante que a combinação de efeitos aleatórios para tratamentos e a AE não viola as suposições (algumas destas justificadas pelo delineamento). Isto é especialmente útil com experimentos grandes (com um grande número de progênies).

---

\*Comitê Orientador: Júlio Sílvio de Sousa Bueno Filho - UFLA (Orientador) e Magno Antônio Patto Ramalho - UFLA.

## ABSTRACT

COSTA, Janaina Ribeiro. **Spatial and neighbourhood analysis in plant breeding**. 2003. 82 p. Thesis (Doctor Degree in Genetics and Plant Breeding) – Universidade Federal de Lavras, Lavras, MG\*.

The aim of this work was to look for alternatives to improve the efficiency of progeny evaluation experiments with common bean and maize. The focus was on analytical techniques that could enhance the experimental precision. Grain yield data were submitted to different types of analyses, ignoring or taking into account the local control of the design. For each analysis, four types of models were used: normal Gauss-Markov (GMN), moving averages on raw data (MM), moving averages on residual data (Papadakis - PPD) and spatial analysis modeling residual covariances (AE). We also investigated the consequences of considering the treatment effects as being fixed or random on modeling for genotype ranking and selection. The comparison of the different types of analysis was based on their efficiencies. Residual variances ( $\hat{\sigma}^2$ ), progeny variances ( $\hat{\sigma}_p^2$ ), relative semi-amplitudes (SA) of the confidence intervals of  $\hat{\sigma}_p^2$  and Akaike information criteria (AIC) were calculated. Results showed that ignoring the experimental design in the model can rarely be effective, with the aid of spatial information. MM and PPD models in many cases improved the original model justified by design, although AE did not. In addition, there were differences in the progeny ranking when representing them by random effects (compared to the usual recovery of interblock information with fixed effects of treatments). AE, if ineffective, does not change variance components estimates. This property guarantees that the combination of random effects of treatments and AE do not violate the assumptions (some of these justified by the design). This is specially useful with large experiments (with a huge number of progenies).

---

\*Guidance Committee: Júlio Sílvio de Sousa Bueno Filho - UFLA (Major Professor) and Magno Antônio Patto Ramalho - UFLA.



## 1 INTRODUÇÃO

Programas de melhoramento genético de plantas envolvem etapas como a escolha de genitores, realização de cruzamentos, escolha e realização do processo seletivo e avaliação prévia dos genótipos em campos experimentais, antes de serem recomendados para produção comercial. A etapa de avaliação é a mais cara, mais demorada e que exige maior rigor na decisão. Diante deste fato, fica evidente a necessidade de cuidados especiais durante esta fase experimental. Assim, os melhoristas têm como desafio aumentar a precisão dos experimentos com os recursos disponíveis.

Muito embora o sucesso obtido no melhoramento genético seja fato conhecido, é também sabido que as diferenças a serem detectadas no futuro serão cada vez menores. Assim, na tentativa de detectar as pequenas variações existentes, os melhoristas são forçados a avaliar números cada vez maiores de progênies (superior a 100) sob várias condições ambientais.

Nestes casos, em que estão envolvidos muitos genótipos, o pesquisador defronta-se com o problema de distribuição dos tratamentos na área experimental, pois torna-se difícil a obtenção de áreas homogêneas, como exemplo quanto à fertilidade do solo, para manter blocos que proporcionem comparações entre os tratamentos e estimações de parâmetros genotípicos e fenotípicos mais precisas. Com o objetivo de controlar o problema de heterogeneidade do solo são adotados freqüentemente, no melhoramento, os **delineamentos em blocos incompletos (DBI)**.

Atualmente, outras estratégias baseadas em modelagens espaciais têm sido sugeridas no melhoramento de plantas para exercer o controle local e aumentar a precisão experimental, tais como a análise de vizinhança e a análise espacial dos dados. Estas formas de análise são mais apropriadas para

experimentos em que é detectada dependência espacial entre os erros experimentais. Nestes casos, a eficiência dos estimadores de contrastes de tratamentos não dependerá apenas da variação residual, mas também das posições das parcelas a que foram alocados os tratamentos.

Estudos envolvendo métodos de análise espacial foram feitos por Bartlett (1978), Vivaldi (1990), Brownie et al. (1993), Cressie & Hartfield (1996), Souza (1997), Duarte (2000) entre vários outros autores.

Entre os métodos de análise de vizinhança de interesse para o melhoramento, encontram-se os denominados médias móveis e método de Papadakis, que modelam a dependência espacial na forma de análise de covariância. Já na análise espacial esta dependência é modelada na forma de modelagem da matriz de variância e covariâncias residuais, considerando a distância entre parcelas no campo experimental.

A análise espacial, desenvolvida na década de 50, foi pouco explorada até algum tempo atrás devido a dificuldades de cálculo. Contudo, em suas diversas formas, esta análise seria uma alternativa para aumentar a precisão experimental sem aumentar o número de repetições, com reduzidas exigências de delineamento.

É importante notar que sempre que há um delineamento casualizado, pode-se ignorar a análise espacial mas, ainda assim, é de se esperar que tal análise possa aumentar a precisão das estimativas de interesse.

Neste sentido, o objetivo do presente trabalho foi procurar alternativas que possam melhorar a eficiência dos experimentos de avaliação de linhagens e , ou, progênies de feijão e milho por meio de técnicas de análises que contribuam para melhorar a precisão experimental.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Princípios básicos da experimentação**

Os delineamentos experimentais clássicos são baseados em três conceitos estabelecidos por Fisher (1935): o da repetição, o da casualização e o do controle local.

O princípio da repetição consiste na aplicação de cada tratamento em mais de uma parcela. Os tratamentos devem ser repetidos, possibilitando, assim, estimar o erro experimental sem o qual não seria possível realizar testes de hipóteses. Mead & Curnow (1983) afirmam que quanto mais parcelas forem usadas para cada tratamento, menor será o erro padrão das estimativas dos efeitos de tratamentos e do contrastes entre eles e, portanto, maior a acurácia do experimento. Conforme Davis (1986), a existência de variabilidade nos solos experimentais poderia ser compensada pelo uso de um maior número de parcelas repetidas. Ramalho et al. (2000) concordam que o uso de um número apropriado de repetições contribui para aumentar a precisão das estimativas dos parâmetros e, ao mesmo tempo, para aumentar o poder dos testes estatísticos.

Conforme Steel & Torrie (1997), a razão para o princípio da casualização em delineamentos experimentais é fornecer estimativas não tendenciosas de efeitos de tratamentos ou contrastes entre eles e fornecer estimativas válidas da variação residual ou da precisão. Hinkelmann & Kempthorne (1994) indicam a utilidade da casualização na validação da aproximação pelo modelo Gauss-Markov normal. De acordo com Mead & Curnow (1983), a casualização deve ser usada na alocação dos tratamentos às parcelas e tem por finalidade propiciar a todos os tratamentos a mesma

probabilidade de serem designados a qualquer um a das unidades experimentais. Com isso evita-se que qualquer tratamento seja sistematicamente favorecido e garante que os erros sejam independentes. Desta forma, a casualização pode reduzir o efeito da correlação espacial entre parcelas adjacentes, validando a análise de variância (Grondona & Cressie, 1991). Por isso, as posições dos tratamentos (configuração espacial) no mapa de campo são normalmente ignoradas na abordagem tradicional.

Para Hinkelmann & Kempthorne (1994), o princípio do controle local é o reconhecimento de padrões supostamente associados às parcelas. A idéia básica é a partição do conjunto total de parcelas em subconjuntos (blocos) que são os mais homogêneos possíveis. Segundo Brownie et al. (1993), a finalidade da blocagem é reduzir a variação residual e assim aumentar a precisão experimental, pois ela garante que os tratamentos sejam avaliados em condições ambientais (e operacionais) similares dentro de um mesmo bloco. De acordo com Es & Es (1993), a heterogeneidade do solo inflaciona a estimativa do erro experimental, reduzindo a capacidade do melhorista em detectar diferenças entre tratamentos. Assim, se os efeitos são pequenos, a alta variabilidade no campo pode não permitir a detecção destas diferenças.

Como estratégias de controle local tem-se a utilização de variáveis blocantes discretas, associadas a delineamentos em blocos completos ou incompletos, quadrados latinos completos ou incompletos, e outros, ou contínuas, que são técnicas conhecidas como análises de covariância, da qual a análise de vizinhança é um caso particular.

O delineamento mais empregado é o de blocos casualizados, introduzido por Fisher (1935), em que todos os tratamentos ocorrem dentro de um mesmo bloco, que espera-se ser o mais homogêneo possível, e a heterogeneidade entre os blocos não é incluída no erro. Este tipo de delineamento é utilizado quando se quer controlar a heterogeneidade ambiental em apenas uma direção. Muitas

vezes se observa heterogeneidade nas parcelas em duas ou mais direções e os delineamentos recomendados, neste caso, são generalizações dos delineamentos em blocos. Em especial, os quadrados latinos (completos) têm utilização pouco difundida em genética e melhoramento de plantas devido ao grande número de repetições que estes delineamentos exigem.

## **2.2 Tipos de experimentos mais usados em genética e melhoramento de plantas**

A maioria dos caracteres de interesse econômico que são manuseados pelos melhoristas são quantitativos, ou seja, controlados por muitos genes e muito influenciados pelo ambiente. De acordo com Ramalho et al. (2000), este fato implica em uma baixa probabilidade de associar, em um indivíduo, todos os alelos favoráveis. Na expectativa de aumentar a chance de sucesso, os melhoristas normalmente avaliam algumas centenas de linhagens e, ou, famílias, o que permite encontrar valores genéticos extremos para os efeitos de tratamentos.

Nesse contexto, a eficiência dos delineamentos com blocos completos é reduzida à medida que o número de parcelas por bloco passa de dez, tornando-se difícil obter um bloco homogêneo na dimensão desejada. Conforme Es & Es (1993), a utilização de blocos pequenos enquadra-se na pressuposição de uniformidade dentro dos blocos e, portanto, pode fornecer uma maior captação na quantificação de variabilidade localizada. Ramalho et al. (2000) enfatizam que a heterogeneidade dentro do bloco tende a aumentar o erro experimental, dificultando a discriminação das linhagens e, ou, famílias em teste ou a obtenção de estimativas mais precisas, em razão de uma menor precisão do experimento.

Com o objetivo de controlar a heterogeneidade dentro de blocos, e assim contornar os problemas mencionados, são adotados freqüentemente, no melhoramento genético vegetal, os delineamentos em blocos incompletos introduzidos por Yates (1936). Estes delineamentos caracterizam-se pelo fato de que cada bloco não contém todos os tratamentos. Isto enquadra-se em uma situação de não ortogonalidade. Assim, as comparações entre tratamentos são influenciadas pelos blocos o que exige, na análise estatística, o ajustamento de totais e médias de tratamentos para os efeitos de blocos.

Os delineamentos em blocos incompletos podem ser classificados em balanceados e parcialmente balanceados. Segundo Yates (1936), quando quaisquer pares de tratamentos ocorrem juntos em um mesmo número de blocos, todos os contrastes entre tratamentos têm a mesma precisão e o delineamento é chamado de blocos incompletos balanceados (**BIB – Balanced Incomplete Block**).

Bose & Nair (1939) afirmam que os delineamentos do tipo BIB, apesar de mais eficientes, em algumas situações tornam-se impraticáveis, dado o número excessivo de repetições requerido e as dificuldades de construção. Assim, na tentativa de contornar este problema, os autores introduziram os delineamentos em blocos incompletos parcialmente balanceados (**PBIB – Partially Balanced Incomplete Block**), o que consiste na utilização de apenas parte das repetições contidas no BIB.

Nos delineamentos em PBIB existem as chamadas classes de associados representadas pelo parâmetro  $\lambda$ . Este parâmetro indica a associação entre os pares de tratamentos com respeito a disposição nos blocos. Assim, há delineamentos com várias classes de associados, como exemplo os que possuem três classes de associados, em que dois tratamentos nunca estão juntos ( $\lambda_1=0$ ) ou estão juntos em um único bloco do experimento ( $\lambda_2=1$ ) ou estão juntos em dois blocos ( $\lambda_3=2$ ). Normalmente os delineamentos mais utilizados são os que

possuem duas classes de associados pois são mais eficientes (PBIB); os tratamentos que estão juntos em um mesmo bloco são comparados com maior precisão do que aqueles que estão em blocos diferentes.

Na terminologia de blocos incompletos é comum encontrar os termos “resolvable designs” referindo-se aos casos em que os blocos podem ser agrupados em repetições completas e “non resolvable design”, caso contrário.

De acordo com Ramalho et al. (2000), os delineamentos do tipo “resolvable” devem ser os preferidos, pois permitem inclusive que as análises sejam efetuadas de acordo com o modelo de blocos completos. É comum entre os melhoristas conduzir a análise de variância do látice como se fosse um delineamento em blocos completos, reunindo os blocos de modo a formarem repetições. A razão entre as estimativas das variâncias das estimativas dos contrastes entre médias de dois tratamentos, dos dois delineamentos, fornece a chamada eficiência relativa (ER) do látice em relação aos blocos casualizados completos.

Na categoria de delineamentos “resolvables” mais usados no melhoramento de plantas enquadram-se os látices quadrados e os retangulares e na categoria de “non resolvable” encontram-se, por exemplo, os delineamentos em blocos aumentados.

Pode-se definir 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 e, para atender à condição de balanceamento, tem-se como consequência que  $r=k+1$ , sendo  $r$  o número de repetições ortogonais. Estes delineamentos só existem para  $k$  sendo número primo ou potência de primo. Já o látice parcialmente balanceado é obtido a partir do delineamento balanceado, tomando-se algumas das  $k+1$  repetições, podendo-se utilizar de duas até  $k$  repetições quando  $k$  é primo ou potência de primo (Ramalho et al., 2000).

Os autores sugerem que seria interessante, principalmente nas fases finais de um programa de melhoramento de plantas, o uso do látice quadrado com tratamentos comuns, possibilitando, deste modo, comparar com alto grau de precisão as linhagens e, ou, progênies melhoradas com outras de comportamento conhecido (testemunhas). Uma desvantagem, neste caso, seria o aumento no tamanho do experimento.

Os látices retangulares, propostos por Harshbarger (1949), permitem a comparação de  $v = p(p-1)$  tratamentos em blocos de  $k = p-1$  parcelas, em que  $p$  é um número inteiro. Dessa forma, podem ser utilizados números de tratamentos que se situam entre os permitidos pelos látices quadrados. Tais látices são, portanto, PBIB. Quando dois arranjos básicos ( $x$  e  $y$ ) são utilizados, os látices retangulares são ditos simples, e são denominados triplos quando três arranjos ( $x$ ,  $y$ ,  $z$ ) são utilizados.

Patterson & Williams (1976) propuseram os delineamentos alfa látices, em que os  $v$  tratamentos são dispostos em blocos de tamanho  $k$ , tal que  $v$  seja múltiplo de  $k$  ( $v = ks$ , para  $s > 1$ ), sendo  $s$  o número de blocos por repetição. Uma vantagem dos alfa látices, quando comparados aos látices quadrados e retangulares, está na possibilidade de se poder utilizar uma ampla classe de valores de  $v$ , ou seja, a relação  $v = ks$  no alfa látice é menos restritiva do que a mesma relação para os látices quadrados e retangulares. Na verdade, todos os látices podem ser considerados como delineamentos alfa.

Um dos maiores problemas no uso de todos estes delineamentos citados é que eles requerem um grande número de parcelas. Ocorre que nem sempre é possível atender a esta restrição devido à limitação de recursos financeiros, área experimental disponível e quantidade de sementes para o material genético a ser avaliado, em especial para ensaios iniciais de um programa de melhoramento. A alternativa para estas situações é o uso do delineamento aumentado introduzido por Federer (1956), que é definido como um delineamento padrão (blocos ao



acaso, quadrado latino, látice, etc.) e que recebe tratamentos adicionais em seus blocos, linhas ou colunas. Nos delineamentos aumentados, dois tipos de tratamentos são considerados, os tratamentos comuns, constituídos por testemunhas, incluídos uma única vez em cada bloco do experimento, e os tratamentos a serem avaliados, denominados de regulares, incluídos uma única vez em um dos blocos. Porém, nas fases finais de um programa de melhoramento, os blocos aumentados não devem ser utilizados em razão da baixa precisão que estes delineamentos proporcionam, principalmente nas comparações entre tratamentos alocados em blocos diferentes.

Com respeito à análise de experimentos em blocos incompletos, vale ressaltar que são geralmente encontradas, na literatura, duas formas, a análise intrablocos e a análise com recuperação da informação interblocos. Na análise intrablocos, consideram-se apenas as comparações de parcelas dentro do mesmo bloco, na estimação das diferenças entre tratamentos. Tais estimativas estão sujeitas, portanto, a um erro intrablocos ( $\sigma^2$ ) que depende unicamente das diferenças de fertilidade do solo dentro dos blocos. Assim, quando o material experimental for muito heterogêneo, uma maneira de melhorar a eficiência das comparações pode ser obtida diminuindo-se o tamanho dos blocos.

Já a segunda forma de análise, com recuperação da informação interblocos, consiste em uma informação combinada com análise intrablocos (informação fornecida pelos contrastes entre tratamentos dentro do mesmo bloco) e a informação relativa aos contrastes entre tratamentos, presentes em blocos distintos. Neste caso, as estimativas estão sujeitas a um erro ( $\sigma_1^2$ ) que depende não só de diferenças intrablocos, mas também de diferenças interblocos, ou seja,  $\sigma_1^2 = \sigma^2 + k\sigma_b^2$ , em que  $\sigma^2$  é a variância residual intrablocos,  $\sigma_b^2$  é a variância de blocos e  $k$  é o tamanho dos blocos.

Embora o uso da informação interblocos possa melhorar a qualidade das estimativas de médias de tratamentos, bem como dos contrastes entre elas, esta análise fornece resultados dos testes de significância (F e de comparações múltiplas) apenas aproximados, enquanto em uma análise intrablocos estes são exatos. Assim, autores como Pimentel Gomes (1990), Cochran & Cox (1957) e outros recomendam o uso da análise com recuperação da informação interblocos somente para experimentos com número de graus de liberdade (GL) relativamente grande para blocos e para o resíduo, algo igual ou superior a 10 GL para ambos.

A análise estatística intrablocos é obtida admitindo-se o modelo como fixo, ou seja, com todos os efeitos representados por constantes, exceto o erro experimental, que é uma variável aleatória. Já a análise com recuperação da informação interblocos é baseada num modelo linear misto, assumindo-se os efeitos de blocos dentro de repetições como aleatórios e os demais efeitos como fixos, exceto o erro experimental (Robinson, 1991).

White & Hodge (1989) alertam que, para fins de ordenação do material genético para seleção, o ideal é considerar o efeito de tratamentos como aleatório. Esta abordagem está despertando interesse nos melhoristas de plantas, pois origina uma terceira forma de análise que permite recuperar a informação intergenotípica.

### **2.3 Análise de modelos mistos**

No melhoramento vegetal, bem como em diversas outras áreas de conhecimento, o pesquisador tem interesse em ajustar um modelo estatístico aos dados que seja o mais próximo da realidade. Os modelos podem apresentar

efeitos de natureza fixa ou aleatória. Um modelo se diz fixo quando os parâmetros nele existentes são todos de efeito fixo, com exceção do efeito do erro experimental, que é de natureza aleatória. Se o modelo contiver apenas efeitos aleatórios, com exceção de uma constante que é de natureza fixa, ele é dito aleatório. Se, no entanto, aparecerem no modelo efeitos fixos e aleatórios, além da constante e do erro experimental, estes serão denominados modelos mistos (Eisenhart, 1947).

Searle (1971) discute as diferenças entre efeitos fixos e aleatórios. Para ele, se os níveis do fator podem ser considerados uma amostra aleatória de níveis, os efeitos devem ser tratados como aleatórios; caso contrário, as inferências ficarão restritas aos níveis incidentes nos dados e então os efeitos deverão ser tratados como fixos. Logo, se as inferências ficarem confinadas aos efeitos do modelo, estes são considerados fixos; e se as inferências forem feitas sobre uma população de efeitos dos quais uma amostra aleatória faz parte dos dados, então os efeitos são ditos aleatórios.

Robinson (1991) opina que na tomada de decisão se os efeitos do modelo são fixos ou aleatórios, deve-se considerar se os níveis dos fatores vieram de uma distribuição de probabilidade. Se afirmativo, tais efeitos podem ser considerados aleatórios; caso contrário, o ideal é considerá-los como fixos.

De acordo com White & Hodge (1989), os valores genéticos devem ser considerados como efeitos aleatórios pois os indivíduos possuem alelos que segregam e novamente se agregam aleatoriamente formando novos genótipos, cuja observação da expressão se realiza na descendência. O autor acrescenta que em um determinado experimento, não se avalia todos os descendentes de cada indivíduo, mas apenas uma amostra deste quando cruzado com a população.

Bueno Filho (1997) alerta que a decisão de analisar um determinado fator como fixo ou aleatório pode resultar em diferentes análises, produzindo médias ajustadas diferentes, com possível alteração no ranqueamento dos

genótipos. Assim, Duarte (2000) recomenda um bom entrosamento entre melhoristas e biometristas para tomar uma decisão mais acertada sobre a natureza dos efeitos, resultando em uma análise mais precisa.

A Metodologia de Modelos Mistos (MMM), proposta por Henderson (1950), consiste na avaliação dos valores genéticos tratados como sendo de efeito aleatório, corrigindo-os para os demais efeitos fixos contidos no modelo, sendo apropriada para graus elevados de desbalanceamento (André, 1999). Esta metodologia combina, por meio dos quadrados mínimos generalizados, a melhor predição linear não tendenciosa dos efeitos aleatórios, “Best Linear Unbiased Prediction - BLUP”, e a melhor estimativa linear não tendenciosa dos efeitos fixos, “Best Linear Unbiased Estimation - BLUE”.

É comum entre os estatísticos denominar estimadores de efeitos aleatórios de “preditores” e estimadores de efeitos fixos de “estimadores”. Robinson (1991) discute a diferença entre estes dois termos e declara que prefere usar o termo estimador para ambos os efeitos, fixo e aleatório, reservando a palavra preditor para combinações lineares posteriores das estimativas dos parâmetros isolados do modelo.

O BLUP foi originalmente desenvolvido nos Estados Unidos e Canadá para ranqueamento e seleção no contexto do melhoramento genético animal, sendo atualmente também amplamente utilizado no melhoramento genético vegetal, como por Soh (1994), em dendê; por Bernardo (1994), em milho; por Panter & Allen (1995a e 1995b), em soja; por Bueno Filho (1997), em eucalipto; e por André (1999), em feijão.

Robinson (1991) esclarece que as estimativas BLUP dos valores realizados de variáveis aleatórias  $\mathbf{p}$  são lineares no sentido de que elas são combinações lineares dos dados,  $\mathbf{y}$ ; não tendenciosas no sentido de que a média do valor da estimativa é igual à média do valor da quantidade a ser estimada, ou seja,  $E(\hat{\mathbf{p}}) = E(\mathbf{p})$ ; melhores no sentido de que estas estimativas têm quadrado

médio do erro mínimo dentro da classe de estimadores não tendenciosos; e preditores para distingui-las dos estimadores de efeitos fixos.

No caso de dados balanceados e não havendo parentesco entre os genótipos, as soluções do modelo misto (BLUP) e as soluções de quadrados mínimos (BLUE) levam ao mesmo ranqueamento dos genótipos (Henderson et al., 1959).

No contexto do modelo linear misto a seguir, o estimador de efeitos aleatórios é o BLUP e o de efeitos fixos é o BLUE.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{p} + \mathbf{e} \quad (1)$$

em que

**y**: é o vetor das observações, de dimensões (n) x (1);

**X**: é a matriz de incidência dos efeitos fixos (por exemplo, locais) do delineamento, de dimensões (n) x (u);

**β**: é o vetor de efeitos fixos desconhecidos a serem estimados (por exemplo, efeito a ser estimado para o local 1, local 2, etc.), de dimensões (u) x (1);

**Z**: é a matriz de incidência dos efeitos aleatórios (por exemplo, genótipos) do delineamento, de dimensões (n) x (v);

**p**: é o vetor de efeitos aleatórios desconhecidos a serem preditos (por exemplo, efeito a ser predito do genótipo 1, do genótipo 2, etc.), de dimensões (v) x (1);

**e**: é o vetor de erros aleatórios, de dimensões (n) x (1).

Para o modelo acima, n é o número de observações, u é o número de efeitos fixos e v é o número de efeitos aleatórios.

Em geral, tem-se que:

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} ,$$

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} ,$$

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

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}) ;$$

em que

$\mathbf{0}$  é um vetor nulo;

$\mathbf{G}$  é a matriz de variâncias e covariâncias genéticas dada por:

$$\mathbf{G} = \mathbf{A}\sigma_p^2 ,$$

sendo  $\mathbf{A}$  a matriz de coeficientes de parentesco ou similaridade genética entre genótipos e  $\sigma_p^2$  a variância genética;

$\mathbf{R}$  é a matriz de variâncias e covariâncias residuais.

As matrizes  $\mathbf{G}$  e  $\mathbf{R}$  são admitidas matrizes simétricas e positivas definidas.

Muitas vezes, por não conhecer a matriz  $\mathbf{A}$  ou por simplificação de cálculo, o pesquisador assume:

$$\mathbf{G} = \mathbf{I}\sigma_p^2 ,$$

$$\mathbf{R} = \mathbf{I}\sigma^2 ;$$

em que  $\sigma^2$  é a variância residual.

Duarte (2000) afirma que o procedimento de utilizar a matriz identidade ( $\mathbf{I}$ ) em  $\mathbf{G}$  e  $\mathbf{R}$  nem sempre corresponde às melhores estruturas para as diversas situações reais. O autor acrescenta que, em um modelo de ensaio de competição de cultivares, cujas parcelas vizinhas mostrem similaridade intrínseca à variabilidade local (como por exemplo, gradientes de fertilidade do solo dentro de blocos), ou em experimentos com medidas repetidas em uma mesma parcela, assumir  $\mathbf{R}$  como sendo  $\mathbf{I}\sigma^2$  não seria apropriado. Em modelos espaciais, as parcelas não são independentes; neste caso, assume-se que  $\mathbf{R} = [f(d)]\sigma^2$ , em que  $f(d)$  é uma função da distância ( $d$ ) entre duas parcelas (veja mais detalhes no item 2.4.3).

Nos modelos lineares de Gauss-Markov fixos em que  $\mathbf{y}=\mathbf{X}\beta+\mathbf{e}$ , com  $\mathbf{e}\sim\mathbf{N}(\mathbf{0},\mathbf{R}=\mathbf{I}\sigma^2)$ , baseados no método dos quadrados mínimos ordinários, a matriz  $\mathbf{V}$  de variâncias e covariâncias dos dados é igual à matriz de variâncias e covariâncias dos erros,  $\mathbf{V}=\mathbf{R}=\mathbf{I}\sigma^2$ , o que significa que os resíduos são não correlacionados, ou seja, todas as observações tomadas são independentes, não havendo, portanto, covariância entre elas. As estimativas  $\beta^0$  obtidas neste caso são ditas estimativas de quadrados mínimos ordinários (OLS). Já em um modelo misto, de acordo com Duarte (2000), esta covariância entre as observações existe pois os níveis de um fator aleatório estão relacionados entre si por uma população de referência. Neste caso, a matriz  $\mathbf{V}$  será diferente de  $\mathbf{I}\sigma^2$ , mesmo admitindo-se independência do erro, pois em seu cálculo estará presente a estrutura de correlação presente nos dados, ou seja,  $\mathbf{V}=\mathbf{ZGZ}' + \mathbf{R}$ . As estimativas obtidas, no caso do modelo misto, são ditas estimativas de quadrados mínimos generalizados (GLS).

As Equações de Modelos Mistos (EMM) baseadas no modelo em (1) são dadas por:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \beta^{\circ} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad (2)$$

Robinson (1991) chama a atenção para os casos em que a matriz  $\mathbf{G}^{-1}$  tende para a matriz nula. Nestes casos, as equações em (2) tendem para as equações de quadrados mínimos generalizados para estimar  $\beta$  e  $\mathbf{p}$  quando os componentes de  $\mathbf{p}$  são considerados de efeitos fixos.

Efetuada a multiplicação em (2), têm-se:

$$\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\beta^{\circ} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{p}} = \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} , \quad (3)$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\beta^{\circ} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\mathbf{p}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} . \quad (4)$$

Henderson et al. (1959), após realizar várias operações algébricas, apresenta as seguintes soluções do sistema de equações do modelo misto:

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

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

em que  $\mathbf{C} = \mathbf{G}\mathbf{Z}'$  é a matriz de covariâncias entre  $\mathbf{y}$  e  $\mathbf{p}$  (covariâncias entre observações fenotípicas e valores genotípicos verdadeiros).

Ocorre que muitos pesquisadores alegam encontrar dificuldades computacionais no processo de inversão da matriz  $\mathbf{V}$ , normalmente de grande dimensão ( $n \times n$ ). Portanto, existem outras alternativas para obtenção do BLUP, apesar de esta alegação não ser mais tão importante nos dias de hoje.



Robinson (1991) apresenta quatro derivações para obtenção do BLUP, dentre as quais destaca-se, a seguir, a derivação de máxima verossimilhança (primeira justificativa de Henderson).

Henderson (1950) descreve as estimativas BLUP como sendo “estimativas de máxima verossimilhança conjuntas”. Henderson (1973) assumiu, em sua derivação, que  $\mathbf{p}$  e  $\mathbf{e}$  são normalmente distribuídos.

A função densidade conjunta de  $\mathbf{y}$  e  $\mathbf{p}$  é:

$$F(\mathbf{y}, \mathbf{p}) = (2\pi\sigma^2)^{\frac{1}{2}n - \frac{1}{2}v} \left( \det \begin{bmatrix} \mathbf{G} & \boldsymbol{\phi} \\ \boldsymbol{\phi} & \mathbf{R} \end{bmatrix} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \begin{pmatrix} \mathbf{p} \\ \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p} \end{pmatrix}' \begin{bmatrix} \mathbf{G} & \boldsymbol{\phi} \\ \boldsymbol{\phi} & \mathbf{R} \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p} \end{pmatrix} \right\} \quad (7);$$

em que  $\boldsymbol{\phi}$  é uma matriz nula e os demais termos já foram descritos anteriormente.

O autor maximiza a equação (7) em relação a  $\boldsymbol{\beta}$  e  $\mathbf{p}$ , o que corresponde a minimizar:

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p} \end{pmatrix}' \begin{bmatrix} \mathbf{G} & \boldsymbol{\phi} \\ \boldsymbol{\phi} & \mathbf{R} \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p} \end{pmatrix} = \mathbf{p}' \mathbf{G}^{-1} \mathbf{p} + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{p}).$$

Derivando esta função em relação a  $\boldsymbol{\beta}$  e  $\mathbf{p}$  e igualando as derivadas a zero, chega-se às equações do modelo misto de Henderson em (3) e (4).

De acordo com André (1999), de (3) pode-se obter a solução para  $\boldsymbol{\beta}$  dada por:

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

Substituindo (8) em (4), têm-se:

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\boldsymbol{\beta}}) + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} ,$$

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\boldsymbol{\beta}} + (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}$$

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

$$\{\mathbf{Z}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}]\mathbf{Z} + \mathbf{G}^{-1}\}\hat{\boldsymbol{\beta}} = \mathbf{Z}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}]\mathbf{y} . \quad (9)$$

Fazendo  $\mathbf{S} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}^{-1}$  e substituindo S em (9), têm-se:

$$(\mathbf{Z}'\mathbf{S}\mathbf{Z} + \mathbf{G}^{-1})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{S}\mathbf{y} .$$

Isolando  $\hat{\boldsymbol{\beta}}$  chega-se à solução de quadrados mínimos generalizados para os efeitos aleatórios dada por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{S}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y} . \quad (10)$$

Substituindo  $\hat{\boldsymbol{\beta}}$  em (8), chega-se à solução de quadrados mínimos generalizados para os efeitos fixos dada por:

$$\hat{\boldsymbol{\beta}}^{\circ} = \{\mathbf{X}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{X}\}^{-1}\mathbf{X}'[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{y} . \quad (11)$$

Esta forma alternativa de obtenção das soluções das EMM tem a vantagem de reduzir o trabalho computacional, uma vez que as matrizes a serem invertidas possuem dimensões menores que a dimensão da matriz  $\mathbf{V}$ .

Vale ressaltar que a inversa da matriz  $\mathbf{V}$  pode ser obtida por:

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}.$$

Duarte (2000) comenta que quando as matrizes  $\mathbf{G}$  e  $\mathbf{R}$  são conhecidas,  $\beta^0$  (ou, mais provavelmente, alguma função estimável  $\mathbf{L}'\beta^0$ ) é chamado melhor estimador linear não viesado (**BLUE** – **B**est **L**inear **U**nbiased **E**stimator) de  $\beta$  (ou de  $\mathbf{L}'\beta^0$ ), e  $\hat{\mathbf{p}}$  é dito melhor preditor linear não viesado (**BLUP** – **B**est **L**inear **U**nbiased **P**redictor) de  $\mathbf{p}$ . O autor sugere que quando  $\mathbf{G}$  e  $\mathbf{R}$  são desconhecidas, dispondo-se apenas de suas estimativas  $\hat{\mathbf{G}}$  e  $\hat{\mathbf{R}}$ , deve-se substituir os termos BLUE e BLUP por **EBLUE** (**E**mpirical **B**est **L**inear **U**nbiased **E**stimator) e **EBLUP** (**E**mpirical **B**est **L**inear **U**nbiased **P**redictor), respectivamente.

Usando um enfoque da Escola Clássica, Henderson (1963) também mostrou, usando os multiplicadores de Lagrange, que as estimativas BLUP de combinações lineares de efeitos fixos e aleatórios são as estimativas que satisfazem os requisitos de serem lineares, não tendenciosas e possuem quadrado médio do erro mínimo.

Harville (1976) mostrou que a solução do modelo de Gauss-Markov pode ser estendida para os casos em que as matrizes  $\mathbf{G}$  e  $\mathbf{R}$  não apresentam posto completo.

Uma aproximação mais intuitiva para mostrar que as estimativas BLUP têm quadrado médio do erro mínimo dentro da classe de estimativas lineares não tendenciosas foi apresentada por Harville (1990).

White & Hodge (1989) apresentaram, entre outras, as seguintes vantagens do BLUP quando usado no melhoramento genético vegetal:

1. permite a estimação e predição não viesadas em apenas um procedimento linear;
2. pode levar em conta efeitos de seleção e endogamia ao longo das gerações, desde que o grau de relacionamento genético (parentesco) entre os indivíduos avaliados seja conhecido;
3. a correlação entre os valores genéticos verdadeiros e os preditos é máxima dentre todas as classes de preditores lineares não tendenciosos;
4. pode-se prever o valor genético de indivíduos observados ou não;
5. as variâncias e os erros das predições BLUP são menores em relação a outros métodos.

White et al. (1986) demonstraram que, nos ensaios em que se têm poucas observações, as predições de quadrados mínimos ordinários são menos confiáveis e os erros associados com estas predições são maiores. Contrariamente, quando os BLUP's foram obtidos utilizando as poucas observações existentes, as variâncias entre as predições foram menores, e os erros associados a estas predições foram minimizados. Assim, uma vez que os melhoristas de plantas freqüentemente desejam melhorar os valores preditos para vários indivíduos que não foram avaliados por meio do seu próprio desempenho, o BLUP teria uma significativa vantagem sobre os outros métodos de melhoramento de valores preditos.

Bernardo (1994) avaliou a utilidade do BLUP em prever cruzamentos não realizados usando dados de marcadores moleculares do tipo RFLP e também usando a informação de parentesco genético entre linhagens de milho. O autor verificou que não houve diferenças significativas entre os resultados dos dois procedimentos. Assim, o uso de marcadores moleculares pode ser uma alternativa quando não se dispõe da genealogia entre os materiais genéticos a serem utilizados no estudo; porém, esta alternativa implica em elevados custos.

Panter & Allen (1995a) compararam a precisão das predições BLUP com a precisão das estimativas BLUE em um experimento de identificação de linhagens superiores em cruzamentos de variedade de soja, simulando situações de balanceamento e desbalanceamento de dados. Além disso, os autores avaliaram o ganho na precisão do BLUP pela inclusão da informação de parentesco genético no cálculo do mesmo. Em todas as situações estudadas, as predições BLUP apresentaram menores erros padrão e maiores valores de correlação classificatória entre os valores preditos e o desempenho dos cruzamentos avaliados no campo. Devido ao alto parentesco genético entre as linhagens usadas nos cruzamentos, não houve um aumento considerável na precisão do BLUP pela inclusão do parentesco com relação à não inclusão do mesmo no cálculo do BLUP. Contudo, os autores recomendam que, se as linhagens a serem cruzadas não são altamente relacionadas, os dados de parentesco devem ser incluídos no cálculo pois eles podem aumentar a precisão das predições.

André (1999) comparou as metodologias de quadrados mínimos ordinário (OLS) e BLUP na predição das capacidades de combinação em modelos dialélicos com efeitos aleatórios, usando informações de similaridade genética entre os genitores obtidas por meio de marcadores moleculares de herança dominante. O autor utilizou dados simulados e dados experimentais de linhagens de feijão e avaliou, ainda, a eficiência do BLUP em prever o desempenho de híbridos não realizados ou perdidos na análise dialélica. O autor concluiu que o BLUP é uma alternativa eficiente e superior ao método dos quadrados mínimos ordinário para a análise dialélica, desde que haja informação disponível sobre a similaridade genética entre os genitores, com a vantagem adicional de possibilitar a previsão de cruzamentos não realizados ou perdidos.

## 2.4 Modelagens espaciais

Uma das pressuposições para a validade de uma análise de variância convencional é a independência entre os erros experimentais. Muitos pesquisadores acreditam que a prática da casualização neutraliza alguma correlação que possa haver entre os erros experimentais, resultando na independência entre eles. Acontece que, ainda assim, pode-se detectar correlação espacial entre parcelas vizinhas, principalmente em experimentos com parcelas muito pequenas e, ou, com tratamentos distribuídos sistematicamente. Vários autores, como Matheron (1963), Brownie et al. (1993), afirmam que quando os campos experimentais não são uniformes, a correlação entre observações tende a ser maior para parcelas vizinhas e diminui à medida que a distância entre as parcelas aumenta. Nestes casos, a utilização de métodos estatísticos alternativos que procuram eliminar os efeitos da correlação espacial, conhecidos como métodos de análises de vizinhança e método de análise espacial, podem aumentar a precisão das estimativas de contrastes entre tratamentos e de parâmetros genotípicos e fenotípicos.

Brownie et al. (1993) reforçam que o argumento para usar estes métodos alternativos de análise estatística é que maiores ganhos na precisão podem ser obtidos comparativamente à análise convencional. Os autores relatam ainda que tais métodos podem ser aplicados aos dados obtidos de um delineamento experimental clássico, apenas melhorando as estimativas.

Os métodos de análise de vizinhança, de interesse particular, consistem no ajuste de uma covariável para o controle local, com base nas médias observadas das parcelas vizinhas (Rickey, 1924) ou nas médias dos resíduos de parcelas vizinhas (Papadakis, 1937). Já na realização de uma análise espacial, consideram-se as coordenadas geográficas (que serão denominadas adiante de

linha e coluna) do centro de cada parcela do experimento no cálculo da matriz de variâncias e covariâncias residuais, por meio da qual é modelada a dependência espacial.

A maioria destes métodos foi desenvolvida extrapolando os conceitos de séries temporais e geoestatística para experimentação agrícola, os quais foram a base para a construção dos modelos de tendência no solo e dos erros correlacionados entre parcelas vizinhas, que são as principais fontes de heterogeneidade espacial. Tais modelos são de interesse para o melhoramento vegetal, sendo, portanto, tratados a seguir.

Tanto os modelos de análise de vizinhança, o de médias móveis (MM) e o de Papadakis (PPD), quanto o modelo de análise espacial são todos “modelos espaciais”; porém, manteve-se a denominação análise espacial (AE) para o método espacial que modela diretamente a covariância entre parcelas como função da distância. Conforme indicado por Atkinson (1969), MM e PPD são casos especiais do modelo AE em que um modelo do tipo auto-regressivo, com alcance pré-determinado, é construído.

#### **2.4.1 Análise de vizinhança: método de Papadakis**

Papadakis (1937) foi o primeiro a introduzir o método do vizinho mais próximo (Nearest-Neighbor Method-NN) com a finalidade de reduzir o erro experimental. O método de Papadakis consiste em corrigir o valor observado de cada parcela pelo efeito médio do tratamento aplicado à parcela, e então, a média dos valores corrigidos das parcelas vizinhas é utilizada como covariável, seguindo-se com uma análise de covariância. O que se deseja com a aplicação do método é eliminar parte das tendências existentes nos solos, por meio do

desempenho das parcelas vizinhas, conseqüentemente diminuindo o erro experimental.

Bartlett (1938) discutiu a validade do método sugerido por Papadakis usando dois experimentos de algodão, e concluiu que o método é aproximado, contudo, pode ser satisfatório quando o número de parcelas por bloco é grande.

Pearce & Moore (1976) aplicaram o método de ajustamento das parcelas pela média dos valores dos resíduos das parcelas vizinhas (método de Papadakis) em culturas perenes frutíferas como o chá, pêssego, abacaxi, uva e maçã, em diferentes partes do mundo, obtendo reduções nos erros experimentais na maioria dos experimentos. Os autores usaram, como covariável, dados de duas a oito árvores vizinhas, e concluíram que quanto maior o número de vizinhos utilizados, maior a redução no erro experimental.

Analisando experimentos simulados, Vivaldi (1990) comparou os métodos de análise espacial, o de Papadakis, o de médias móveis e o de blocos móveis com relação à redução do erro experimental. O autor verificou que o método de Papadakis foi o que proporcionou maior redução do erro; porém, sugeriu estudos complementares com respeito aos métodos de médias móveis e blocos móveis, por meio da modificação dos pesos dados às parcelas ou pela obtenção do número ótimo de parcelas envolvidas na criação das covariáveis.

Brownie et al. (1993) compararam a precisão da análise de tendência, do método de Papadakis e da análise com base em modelos com erros correlacionados em relação à uma análise clássica em blocos casualizados. Os autores analisaram dados de onze experimentos de milho e um experimento de soja, em que havia evidências de heterogeneidade do solo dentro dos blocos, e concluíram que houve maior eficiência da análise de tendência associada à análise de erros correlacionados em relação à análise clássica, ficando o método de Papadakis em uma situação intermediária.



Cressie & Hartfield (1996) concluíram, em seus estudos com trigo e tabaco, que as estimativas de efeitos de tratamentos obtidas por meio do método de Papadakis são equivalentes às estimativas de máxima verossimilhança e de quadrados mínimos generalizados quando um modelo Gaussiano condicional é assumido. Os autores obtiveram ganhos de 20% na eficiência das estimativas dos tratamentos quando a modelagem espacial foi adotada.

#### **2.4.2 Análise de vizinhança: método das médias móveis**

Desde 1924, Rickey, um dos pioneiros na utilização do método das médias móveis, ajustava médias de progênies de milho a partir da razão entre as médias das fileiras vizinhas e, visto que cada fileira possuía vizinhos diferentes, o autor denominou a técnica de médias móveis. O autor detectou reduções no erro padrão das médias após o ajuste por esta técnica (Rios, 1997). Assim, no método das médias móveis, o valor de cada parcela é corrigido pela média dos valores das parcelas vizinhas e não pelo efeito de tratamento, como no método de Papadakis.

Souza (1997), comparando a eficiência de diferentes delineamentos e métodos de análise espacial em experimentos de feijão e eucalipto, concluiu que os métodos de análise de vizinhança, Papadakis e o das médias móveis, mostraram-se eficientes no controle da heterogeneidade ambiental dentro dos blocos, sendo que a eficiência dos mesmos em relação aos parâmetros genéticos e fenotípicos avaliados foi comparável à do látice.

Segundo Zimmerman & Harville (1991), a maioria dos métodos de análise de vizinhança procuram remover os efeitos ambientais de forma indireta, ou seja, a partir do uso de covariáveis obtidas pela diferença ou pelos resíduos

das parcelas vizinhas. Estes autores propuseram um método em que os efeitos de tendência e as correlações entre os erros foram incluídos diretamente no modelo e utilizaram como método de estimação dos parâmetros a máxima verossimilhança. Conforme os autores, resultados promissores têm sido obtidos com esta abordagem, também denominada “modelo linear de campo aleatório” (**Random Field Linear Model - RFLM**). Estes métodos são abordagens mais gerais de modelos (lineares ou não lineares) mistos.

### **2.4.3 Análise espacial**

A variabilidade espacial, principalmente a que se refere às propriedades do solo, é estudada por pesquisadores desde o início do século. Mais recentemente, ela tem despertado uma atenção especial em várias áreas de pesquisa.

Com as facilidades dos recursos computacionais, os métodos de análise espacial de experimentos tornam-se uma alternativa aos melhoristas para maximizar o ajustamento dos efeitos dos tratamentos, removendo ao máximo os efeitos ambientais dos valores fenotípicos.

Para uma dada forma de controle local e um número de repetições, a eficiência dos estimadores dos contrastes entre tratamentos depende somente da variação residual. Na análise clássica, as posições espaciais dos tratamentos no delineamento são ignoradas. Em um ajuste espacial, a eficiência será tipicamente dependente não somente da variação residual, mas também destas posições.

Duarte (2000) afirma que nas fases iniciais de programas de melhoramento, em que normalmente há pouca disponibilidade de material genético para propagação, o melhorista se vê obrigado a utilizar poucas

repetições e parcelas de pequeno tamanho, o que pode violar a suposição de independência entre as observações baseada nos conceitos da estatística clássica. O autor reforça que o uso de blocos grandes (com mais de oito parcelas) também contribui para a heterogeneidade intrabloco. Para efeito de análise, nestes casos, considerar as posições espaciais das parcelas em que foram alocados os tratamentos poderá melhorar a eficiência das estimativas obtidas.

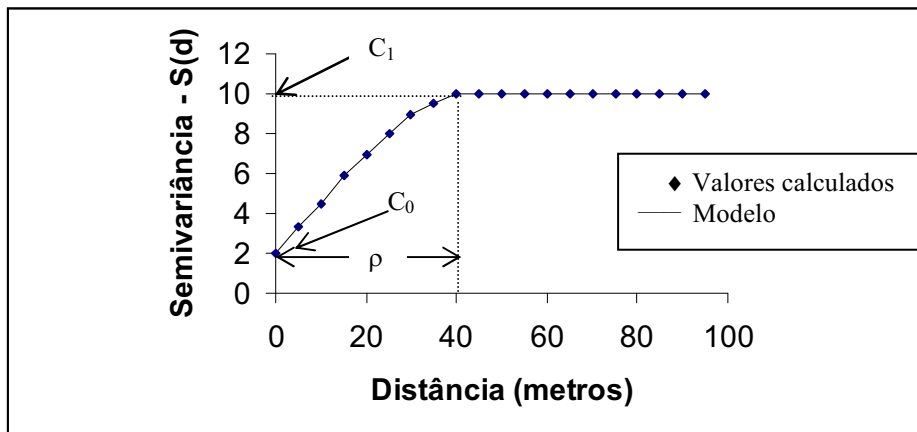
Em geral, associa-se a análise com dependência espacial à Geoestatística, que surgiu na África do sul, quando Krige (1951), trabalhando com dados de concentração de ouro, concluiu que as covariâncias de duas amostras dependiam da distância entre elas. Devido a este fato, muitos modelos espaciais são chamados modelos “geostatísticos”.

Essa dependência espacial pode ser visualizada por meio de um semivariograma, que é uma plotagem da semivariância em função da distância de duas amostras. A semivariância ( $S(d)$ ) é definida como sendo a metade da variância de diferenças entre pares de observações e pode então ser estimada pela seguinte equação:

$$\hat{S}(d) = \frac{1}{2N(d)} \sum_{i=1}^{N(d)} [Z(s+d) - Z(s)]^2 ,$$

em que  $N(d)$  é o número de diferenças  $Z(s+d) - Z(s)$  separadas por uma distância  $d$ , com  $Z(s)$  e  $Z(s+d)$  representando a observação realizada na posição  $s$  e  $s+d$ , respectivamente.

A Figura 1, a seguir, ilustra um semivariograma experimental com características típicas, ou seja, observações localizadas próximas são mais semelhantes entre si do que aquelas mais distantes.



**FIGURA 1.** Representação esquemática de um semivariograma experimental com modelo ajustado.

O ideal é que o ponto de origem do semivariograma passe pela origem do gráfico, mas isto raramente ocorre. Na prática os semivariogramas amostrais normalmente apresentam o chamado efeito pepita (nugget effect), simbolizado por  $C_0$ , que corresponde a semivariância para a menor distância observada. Na maioria dos casos, admite-se não haver mais dependência entre as observações a partir de uma distância denominada alcance ( $\rho$ ), a partir da qual a semivariância máxima passa a ser constante ( $C_1$ ), determinando-se um patamar. A diferença  $C_1 - C_0$  é outro parâmetro para ajuste de semivariogramas e é chamado de contribuição (sill).

A Geoestatística utiliza um método de interpolação chamado “krigeagem”, o qual usa a dependência espacial entre amostras vizinhas, expressa no semivariograma, para estimar valores em qualquer posição dentro do campo, sem tendência e com variância mínima. Essas duas características fazem da “krigeagem” um interpolador ótimo, conforme Burgess & Webster (1980a).

A geoestatística é um caso particular dos modelos mistos em que é facilmente modelável a matriz  $\mathbf{R}$  com base nas distâncias entre unidades experimentais, segundo algum modelo (esférico, exponencial, etc.).

No caso específico da genética e melhoramento de plantas, os tratamentos podem ser melhor representados por efeitos aleatórios, como já comentado anteriormente. Esta abordagem é essencialmente vantajosa quando se conhece a matriz de parentesco (G) entre os níveis de tratamentos, embora não haja muitos exemplos na literatura de análise espacial e de vizinhança com efeitos de tratamentos sendo considerados como efeitos aleatórios, exceção feita a Duarte (2000).

Conforme Littell et al. (1996), nos modelos de correlação espacial, a covariância é dada em função da distância (d) entre duas parcelas, (i, j) e (i', j'), no campo experimental. Assim, o modelo espacial resultante tem a forma geral:

$$\text{Cov}(e_{ij}, e_{i'j'}) = [f(d)]\sigma^2.$$

Para o ajuste da covariância, alguns dos modelos (também denominados modelos geoestatísticos) mais comumente usados são:

#### 1. Esférico

$$f(d) = [1 - 1,5(d/\rho) + 0,5(d/\rho)^3] \cdot I,$$

em que

$$I = 1 \text{ se } d < \rho,$$

$$I = 0 \text{ se } d \geq \rho.$$

## 2. Exponencial

$$f(d) = [\exp(-d/\rho)] .$$

## 3. Gaussiano

$$f(d) = [\exp(-d^2/\rho^2)] .$$

## 4. Linear

$$f(d) = (1 - \rho d).I ,$$

em que

$I = 1$  se  $\rho d < 2$ ,

$I = 0$  se  $\rho d \geq 2$ .

Nestes modelos,  $\rho$  é o alcance (range) e  $d$  é a distância entre duas parcelas. Quando  $d < \rho$ , considera-se que as observações são espacialmente correlacionadas.

Todos estes modelos, e ainda outros, estão implementados na rotina computacional “proc mixed” do SAS (Littell et al., 1996).

Na abordagem de modelos mistos, a covariância espacial é então incorporada no modelo em (1) por meio da matriz de variâncias e covariâncias residuais dada por  $\mathbf{R} = [f(d)]\sigma^2$ .

Grondona & Cressie (1991) comprovaram, em experimentos industriais, que a análise espacial forneceu estimadores de contrastes entre tratamentos 30% mais eficientes do que os da análise clássica.

Es & Es (1993), avaliando dois experimentos simulados, observaram que na presença de correlação espacial entre observações, os testes estatísticos associados a contrastes de tratamentos cujas parcelas estiveram separadas por pequenas distâncias têm maior probabilidade de erro tipo II (tratamentos diferentes apontados como estatisticamente iguais). Já os contrastes de tratamentos cujas parcelas estiveram separadas por distâncias maiores foram testados com maior probabilidade de erro tipo I (tratamentos iguais apontados como estatisticamente diferentes).

Duarte (2000) aplicou a análise espacial em um ensaio de competição de linhagens de soja delineado em blocos aumentados. Os resultados mostraram que a abordagem espacial proporcionou uma melhoria considerável na discriminação dos tratamentos, além de levar a um diferente ordenamento das linhagens em relação à análise clássica, e também a uma seleção menos influenciada por efeitos da variação local.

Portanto, as referências estudadas reforçam as seguintes hipóteses que motivaram esse estudo:

- é possível substituir o controle local do delineamento pela análise espacial;
- é possível melhorar qualquer delineamento com a análise espacial e, ou, com a análise de vizinhança;
- deve-se procurar estimadores de efeitos aleatórios com propriedades BLUP;
- considerar efeitos de tratamentos como aleatórios pode mudar as conclusões a respeito dos experimentos.

### **3 MATERIAL E MÉTODOS**

Os dados utilizados nesse estudo são provenientes de experimentos distintos de progênies de feijão e milho fornecidos pelo programa de melhoramento genético do Departamento de Biologia da Universidade Federal de Lavras (UFLA).

#### **3.1 Experimentos utilizados**

Foram analisados 38 experimentos com a cultura do feijão e 5 com a cultura do milho.

A Tabela 1 apresenta a relação dos experimentos utilizados com suas respectivas características, como época de semeadura, número de repetições, número de blocos por repetição, número de tratamentos, número total de parcelas, número de linhas por parcela e comprimento das linhas em metros referentes aos 38 experimentos, avaliados por meio de delineamentos em látice envolvendo as culturas do feijão e do milho.

A Tabela 2 apresenta algumas destas características; porém, elas são provenientes de experimentos avaliados por meio de delineamentos em blocos casualizados completos. O experimento 43 foi avaliado em esquema de parcelas subdivididas e, neste caso,  $t$  e  $t'$ , na Tabela 2, representam o número de tratamentos da parcela e da subparcela, respectivamente.



**TABELA 1.** Número do experimento (Exp.), época de semeadura, número de repetições (r), número de blocos por repetição (b), número de tratamentos (t), número total de parcelas (p), número de linhas por parcela (L) e comprimento das linhas (C) em metros referentes aos experimentos avaliados em látice com as culturas do feijão e do milho

Exp.	Época de semeadura	Cultura	r	b	t	p	Tamanho da parcela	
							L	C (m)
1	SECA/2000	Feijão	4	16	256	1024	1	1
2	SECA/2000	Feijão	3	20	400	1200	2	2
3	INVERNO/2000	Feijão	3	20	400	1200	2	2
4	INVERNO/2000	Feijão	3	13	169	507	1	2
5	INVERNO/2000	Feijão	3	10	100	300	1	2
6	INVERNO/2000	Feijão	3	6	36	108	2	3
7	INVERNO/2000	Feijão	3	6	36	108	2	2
8	INVERNO/2000	Feijão	3	5	25	75	2	5
9	INVERNO/2000	Feijão	3	5	25	75	2	3
10	INVERNO/2000	Feijão	3	5	20	60	2	4
11	SECA/2001	Feijão	2	15	226	480	1	1
12	SECA/2001	Feijão	3	9	81	243	2	2
13	SECA/2001	Feijão	3	5	25	75	2	5
14	SECA/2001	Feijão	3	5	25	75	2	5
15	SECA/2001	Feijão	3	4	16	48	2	2
16	SECA/2001	Feijão	3	4	16	48	2	2
17	SECA/2001	Feijão	2	5	25	50	1	2
18	SECA/2001	Feijão	2	5	25	50	1	2
19	SECA/2001	Feijão	2	5	25	50	1	2
20	SECA/2001	Feijão	2	5	25	50	1	2
21	SECA/2001	Feijão	2	5	25	50	1	2
22	SECA/2001	Feijão	2	5	25	50	1	2
23	SECA/2001	Feijão	2	5	25	50	1	2
24	SECA/2001	Feijão	2	5	25	50	1	2
25	SECA/2001	Feijão	2	5	25	50	1	2
26	SECA/2001	Feijão	2	5	25	50	1	2
27	SECA/2001	Feijão	2	5	25	50	1	2
28	SECA/2001	Feijão	2	5	25	50	1	2
29	SECA/2001	Feijão	2	5	25	50	1	2
30	SECA/2001	Feijão	2	5	25	50	1	2
31	SECA/2001	Feijão	2	5	25	50	1	2
32	SECA/2001	Feijão	2	5	25	50	1	2
33	SECA/2001	Feijão	2	5	25	50	1	2
34	SECA/2001	Feijão	2	5	25	50	1	2
35	SECA/2001	Feijão	2	5	25	50	1	2
36	SECA/2001	Feijão	2	5	25	50	1	2
37	NOVEMBRO/2000	Milho	2	12	144	288	1	3
38	NOVEMBRO/2000	Milho	3	9	81	243	1	3

**TABELA 2.** Número do experimento (Exp.), época de semeadura, número de blocos (b), número de tratamentos (t), número de tratamentos das subparcelas (t'), número de parcelas (p), número de linhas por parcela (L) e comprimento das linhas (C) em metros referentes aos experimentos avaliados em blocos casualizados completos com as culturas do feijão e do milho

Exp.	Época de semeadura	Cultura	b	t	t'	p	Tamanho da parcela	
							L	C (m)
39	INVERNO/2000	Feijão	5	11	-	55	2	5
40	SECA/2001	Feijão	3	86	-	258	1	2
41	NOVEMBRO/2000	Milho	4	36	-	144	1	3
42	NOVEMBRO/2000	Milho	9	9	-	81	2	5
43	NOVEMBRO/2000	Milho	6	6	2	72	4	5

### 3.2 Local e condução

Os experimentos foram instalados e conduzidos na área experimental do Departamento de Biologia da Universidade Federal de Lavras (UFLA), localizada na região sul do estado de Minas Gerais, a 910 metros de altitude, 21 ° 14' S e 45° 00' W de longitude.

As épocas de semeadura dos experimentos de feijão apresentadas nas Tabelas 1 e 2 correspondem aos meses de fevereiro e março (período da seca), e abril a julho (período de inverno).

Foram obtidos e analisados dados referentes à produção de grãos (g/parcela) para ambas as culturas.

### 3.3 Formas de análise

Os dados de cada experimento com a cultura do feijão ou do milho foram submetidos a diferentes formas de análise, considerando ora que não

houvesse o controle local (CL), ora o delineamento original de controle local (látice, no caso dos experimentos 1 a 38 e blocos casualizados completos, no caso dos experimentos 39 a 43). Para cada uma destas situações (considerando e não o CL) foram consideradas a metodologia tradicional de unidades experimentais independentes por meio do modelo linear de Gauss-Markov Normal (GMN), a análise espacial (AE) correspondente, e ainda os métodos de análise de vizinhança que modelam uma covariável, o de médias móveis (MM) e o de médias de resíduos vizinhos, também denominado método de Papadakis (PPD) por analogia ao método utilizado pela primeira vez por este autor, em 1937.

A seguir estão descritos todos os modelos estatísticos utilizados nesse estudo.

### **3.3.1 Modelo linear sem controle local: delineamento inteiramente casualizado (DIC)**

Os dados de produção de grãos dos experimentos 1 a 42 foram submetidos à análise de variância de acordo com o seguinte modelo do delineamento inteiramente casualizado:

$$y_{ij} = \mu + p_i + e_{ij} . \quad (12)$$

O experimento 43 foi analisado de acordo com o delineamento inteiramente casualizado em esquema de parcela subdividida, considerando os efeitos de tratamentos das parcelas e das subparcelas como fixos.

### **3.3.2 Modelo linear com controle local: delineamento em blocos casualizados completos (DBC)**

Os dados de produção de grãos dos experimentos 39, 40, 41 e 42 foram submetidos, além de ao modelo, conforme a equação (12), à análise de variância conforme o modelo de blocos casualizados completos:

$$y_{ij} = \mu + b_j + p_i + e_{ij} . \quad (13)$$

O experimento 43 foi também analisado em blocos casualizados completos em esquema de parcela subdividida considerando o efeito de blocos como aleatório e os efeitos de tratamentos das parcelas e das subparcelas como fixos.

### **3.3.3 Modelo linear com controle local: delineamento em látice**

Para os dados de produção de grãos dos experimentos de 1 a 38 foi realizada, além de ao modelo, conforme a equação (12), à análise de variância conforme o modelo em látice a seguir:

$$y_{ijk} = \mu + r_k + b_{j(k)} + p_i + e_{ijk} . \quad (14)$$

Os termos dos modelos apresentados de acordo com as equações (12), (13) e (14) são descritos por:

$y_{ij}$  ou  $y_{ijk}$ : valor observado na parcela experimental que recebeu a  $i$ -ésima progênie, no  $j$ -ésimo bloco e na  $k$ -ésima repetição (no caso do modelo, conforme equação 14);

$\mu$ : constante associada a todas as observações;

$p_i$ : efeito aleatório da  $i$ -ésima progênie com  $i = 1, 2, \dots, I$  e  $p_i \underset{i.i.d.}{\sim} N(0, \sigma_p^2)$ ;

$b_j$ : efeito aleatório do  $j$ -ésimo bloco com  $j = 1, 2, \dots, J$  e  $b_j \underset{i.i.d.}{\sim} N(0, \sigma_b^2)$ ;

$r_k$ : efeito fixo da  $k$ -ésima repetição com  $k = 1, 2, \dots, K$ ;

$b_{j(k)}$ : efeito aleatório do  $j$ -ésimo bloco dentro da  $k$ -ésima repetição;

$e_{ij}$  ou  $e_{ijk}$ : erro experimental associado à observação  $y_{ij}$  ou  $y_{ijk}$ , respectivamente.

### 3.3.4 Modelagem de Gauss-Markov Normal (GMN), considerando erros independentes

Para a realização das análises de variâncias de acordo com a modelagem GMN, foram utilizados os modelos conforme as equações (12) e (13) ou (14). Contudo, o que definiu o modelo linear de Gauss-Markov Normal foram as pressuposições feitas com relação ao erro experimental. Considerou-se que os erros são variáveis aleatórias independentes e têm distribuição normal com média zero e variância constante  $\sigma^2$ . Assim, no modelo GMN isto é simbolizado matricialmente por:

$$e \underset{i.i.d.}{\sim} N(0, I\sigma^2).$$

### 3.3.5 Modelagem da análise espacial (AE), considerando erros dependentes

Semelhantemente a modelagem GMN, o que definiu à modelagem espacial foram as pressuposições acerca dos erros experimentais dos modelos, conforme as equações (12), (13) e (14). Porém, as pressuposições assumidas na AE foram de que os erros são dependentes e têm distribuição normal com média zero e matriz de variâncias e covariâncias residuais  $R = f(d)\sigma^2$ , em que  $f(d)$  é uma função da distância ( $d$ ) entre duas unidades experimentais e  $\sigma^2$  é a variância residual. Isto é simbolizado matricialmente por:

$$e \sim N(0, R).$$

Maiores detalhes sobre a AE podem ser vistos no item 3.4.

### 3.3.6 Modelagem das médias móveis (MM)

Nesta modelagem, o valor observado de cada parcela ( $i, j$ ) ou ( $i, j, k$ ) foi corrigido pela média dos valores observados de quatro parcelas vizinhas, uma à direita, uma à esquerda, uma acima e uma abaixo, a qual é definida como média móvel da referida parcela. Para as parcelas das extremidades, foi utilizada a média dos valores observados em uma ou duas ou três parcelas imediatamente próximas a ela. A média móvel de cada parcela foi usada como covariável e procedeu-se a análise de covariância em esquema inteiramente casualizado e em blocos casualizados completos ou em látice, conforme os modelos a seguir, respectivamente:

$$y_{ij} = \mu + p_i + \beta I_{ij} + e_{ij} ; \quad (15)$$

$$y_{ij} = \mu + b_j + p_i + \beta I_{ij} + e_{ij} ; \quad (16)$$

$$y_{ijk} = \mu + r_k + b_{j(k)} + p_i + \beta I_{ij(k)} + e_{ijk} , \quad (17)$$

em que

$y_{ij}$  ou  $y_{ijk}$ ,  $\mu$ ,  $p_i$ ,  $b_j$ ,  $b_{j(k)}$ ,  $r_k$ ,  $e_{ij}$  ou  $e_{ijk}$ : já foram descritos anteriormente após o modelo, conforme a equação 14;

$\beta$ : é o coeficiente de regressão linear entre a variável produção de grãos e a covariável índice ambiental;

$I_{ij}$ : é o índice ambiental da parcela que recebeu a  $i$ -ésima progênie no  $j$ -ésimo bloco (no caso dos modelos, conforme as equações 15 e 16);

$I_{ij(k)}$ : é o índice ambiental da parcela que recebeu a  $i$ -ésima progênie no  $j$ -ésimo bloco e na  $k$ -ésima repetição (no caso do modelo, conforme a equação 17).

### 3.3.7 Modelagem de Papadakis (PPD)

Nesta modelagem, o valor de cada parcela foi corrigido pelo efeito de tratamento (progênie). Foi estimado matricialmente um vetor de resíduos ambientais por meio da seguinte equação:

$$\hat{e} = y - \hat{y} ,$$

em que

$\hat{\mathbf{e}}$ : é o vetor das estimativas de erros aleatórios;  
 $\mathbf{y}$ : é o vetor das observações;  
 $\hat{\mathbf{y}}$ : é o vetor das estimativas das observações.

Em seguida foi calculado, para cada parcela (i, j) ou (i, j, k), um resíduo ambiental médio ( $\bar{I}_{ij}$  ou  $\bar{I}_{ijk}$ ) obtido a partir da média dos resíduos ambientais das parcelas vizinhas, sendo que foram tomadas, para este fim, quatro parcelas vizinhas, uma à direita, uma à esquerda, uma acima e uma abaixo. Para as parcelas das extremidades, foram utilizados os valores dos resíduos médios de uma ou duas ou três parcelas imediatamente próximas a ela. Posteriormente, foi realizada a análise de covariância em esquema inteiramente casualizado e em blocos casualizados completos ou em látice, utilizando-se como covariável os resíduos médios de cada parcela, de acordo com os modelos estatísticos, conforme as equações (15) e (16) ou (17); porém, substituindo  $I_{ij}$  por  $\bar{I}_{ij}$  e  $I_{ijk}$  por  $\bar{I}_{ijk}$ .

### 3.4 Modelo misto

Os modelos, conforme as equações (12), (13) e (14), podem ser matricialmente representados pelo modelo linear misto:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\rho} + \mathbf{e} . \quad (18)$$

em que



**y**: é o vetor das observações, de dimensões (n) x (1);

**X**: é a matriz de incidência dos efeitos fixos do delineamento, de dimensões (n) x (u);

**β**: é o vetor de efeitos fixos desconhecidos a serem estimados de dimensões (u) x (1);

**Z**: é a matriz de incidência dos efeitos aleatórios do delineamento, de dimensões (n) x (v);

**p**: é o vetor de efeitos aleatórios desconhecidos a serem preditos, de dimensões (v) x (1);

**e** : é o vetor de erros aleatórios, de dimensões (n) x (1),

sendo n o número de observações, u o número de efeitos fixos e v o número de efeitos aleatórios. Assumiu-se que:

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

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{R}),$$

$$E(\mathbf{y}) = \mathbf{X}\beta,$$

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R},$$

em que

**0** :é um vetor nulo;

**G** : é a matriz de variâncias e covariâncias genéticas dada por:

$$\mathbf{G} = \mathbf{A}\sigma_p^2,$$

sendo  $\mathbf{A}$  a matriz de coeficientes de parentesco ou similaridade genética entre progênies e  $\sigma_p^2$  a variância de progênies;

$\mathbf{R}$ : matriz de variâncias e covariâncias residuais.

Nos experimentos em estudo não se conhecia a genealogia das progênies; assumiu-se, então, que  $\mathbf{A}$  é igual à matriz identidade ( $\mathbf{I}$ ) e, portanto:

$$\mathbf{G} = \mathbf{I}\sigma_p^2.$$

Para a análise de variância por meio do modelo que pressupõe independência espacial entre as unidades experimentais (modelagem GMN), tem-se:

$$\mathbf{R} = \mathbf{I}\sigma^2,$$

em que  $\sigma^2$  é a variância residual. Já para a análise espacial, que pressupõe correlação espacial entre as unidades experimentais, têm-se;

$$\mathbf{R} = [f(d)]\sigma^2$$

em que  $f(d)$  é uma função da distância ( $d$ ) entre duas unidades experimentais.

Portanto, na modelagem espacial, a covariância é assumida como sendo uma função da distância entre as posições físicas das observações, sendo cada posição especificada por duas coordenadas: linha ( $l_{ij}$ ) e coluna ( $c_{ij}$ ). Assim, as distâncias entre as parcelas ( $i, j$ ) e ( $i', j'$ ) foram calculadas por:

$$d = \sqrt{(l_{ij} - l_{i'j'})^2 + (c_{ij} - c_{i'j'})^2}$$

Na abordagem de modelos mistos, tem-se então que:

$$\text{Cov}(e_{ij}, e_{i'j'}) = \begin{cases} \sigma^2, & \text{se } d = 0 \\ \sigma_{ij,i'j'} = [f(d)]\sigma^2, & \text{se } d > 0 \end{cases} .$$

Por não apresentar alterações em relação a outros modelos testados quanto as estimativas dos componentes de variância, foi escolhido o modelo esférico para o ajuste das covariâncias (elementos da matriz R), encontrado em Littell et al. (1996), dado por:

$$f(d) = [1 - 1,5(d/\rho) + 0,5(d/\rho)^3].I, \quad (19)$$

em que  $\rho$  é o alcance (range). Deste ponto  $\rho$  em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras torna-se invariante com a distância.

Assim, na equação (19), se  $d < \rho$ , então  $I = 1$  e pode-se concluir que as observações são espacialmente correlacionadas.

Ao contrário, se  $d \geq \rho$ , então  $I = 0$  e conclui-se que a correlação espacial é efetivamente zero.

O vetor de soluções do modelo misto na equação (18) é obtido matricialmente por:

$$\begin{bmatrix} \beta^o \\ \hat{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} \end{bmatrix} . \quad (20)$$

Nesse estudo, como os componentes de variância  $\sigma_p^2$  e  $\sigma^2$  (e portanto G e R) eram desconhecidos, utilizou-se o procedimento “proc mixed” do programa SAS para obtenção de suas estimativas, por meio de algoritmos iterativos que utilizam o método da máxima verossimilhança restrita (REML), de forma a obter quase simultaneamente o vetor de soluções na equação (20) composto de BLUE’s e BLUP’s.

### 3.5 Avaliação das formas de análise

As formas de análise descritas no item 3.3 foram comparadas com relação às suas respectivas estimativas das variâncias residuais ( $\hat{\sigma}^2$ ), das variâncias de progênies ( $\hat{\sigma}_p^2$ ), das semi-amplitudes relativas dos intervalos de confiança de  $\hat{\sigma}_p^2$ , (SA) e também quanto às estimativas do critério de informação de Akaike (AIC). Realizaram-se as comparações tomando como padrão a análise considerando o controle local do delineamento, usando o modelo linear Gauss-Markov Normal por meio das seguintes expressões:

1) Eficiência relativa (ER):

$$ER(\%) = \left( \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \right) . 100 , \quad (21)$$

em que

$\hat{\sigma}_1^2$ : é a estimativa da variância residual de um dos modelos avaliados (GMN ou MM ou PPD ou AE);

$\hat{\sigma}_2^2$ : é a estimativa da variância residual do modelo GMN com controle local.

2) Razões ( $R_{\hat{\sigma}_p^2}$ ) entre  $\hat{\sigma}_p^2$ 's:

$$R_{\hat{\sigma}_p^2} (\%) = \left( \frac{\hat{\sigma}_{p1}^2}{\hat{\sigma}_{p2}^2} \right) \cdot 100,$$

em que

$\hat{\sigma}_{p1}^2$ : é a estimativa da variância de progênie de um dos modelos avaliados (GMN ou MM ou PPD ou AE);

$\hat{\sigma}_{p2}^2$ : é a estimativa da variância de progênie do modelo GMN com controle local.

3) Razões ( $R_{SA}$ ) entre SA's:

$$R_{SA} (\%) = \left( \frac{SA_1}{SA_2} \right) \cdot 100,$$

em que

$SA_1$ : é a semi-amplitude relativa do intervalo de confiança de  $\sigma_p^2$  de um dos modelos avaliados (GMN ou MM ou PPD ou AE);

$SA_2$ : é a semi-amplitude relativa do intervalo de confiança de  $\sigma_p^2$  do modelo GMN com controle local.

SA pode ser obtida pela expressão:

$$SA(\%) = \left( \frac{A}{2\hat{\sigma}_p^2} \right) \cdot 100,$$

em que

A: é a amplitude do intervalo de confiança de  $\hat{\sigma}_p^2$  para determinada forma de análise;

$\hat{\sigma}_p^2$ : é a estimativa da variância de progênie para a forma de análise considerada.

Aos intervalos de confiança foi associado um coeficiente de confiança de 0,95.

4) Razões ( $R_{AIC}$ ) entre AIC's:

$$R_{AIC}(\%) = \left( \frac{AIC_1}{AIC_2} \right) \cdot 100,$$

em que

$AIC_1$ : é o critério de informação de Akaike (AIC) de um dos modelos avaliados (GMN ou MM ou PPD ou AE);

$AIC_2$ : é o critério de informação de Akaike (AIC) do modelo GMN com controle local.

O critério de informação de Akaike (AIC) pode ser usado na comparação de modelos com diferentes estruturas de covariância e faz parte da saída padrão do “proc mixed” do SAS. Os valores de AIC são sempre negativos e calculados pela expressão:

$$AIC = \ln(\text{REML}) - q,$$

em que

$\ln(\text{REML})$ : é o logaritmo neperiano da verossimilhança calculada com os **efeitos** fixos e componentes de variância estimados (com propriedades REML);

$q$ : é o número de parâmetros de covariância existentes no modelo completo.

O modelo com maior valor de AIC deve ser preferido por ser mais informativo.

Para fins de comparação da eficiência dos delineamentos em látice em relação aos blocos casualizados completos, procedeu-se o cálculo da eficiência relativa por meio da equação (21), em que

$\sigma_1^2$ : é a média das variâncias das estimativas de efeitos de progênies no delineamento em blocos casualizados;

$\sigma_2^2$ : é a média das variâncias das estimativas de efeitos de progênies no delineamento em látice.

Neste caso, para obtenção de  $\sigma_1^2$  e  $\sigma_2^2$ , foi utilizado o “proc mixed” do SAS, considerando o efeito de repetição como fixo e o efeito de bloco como aleatório; primeiramente, o efeito de progênies foi adotado como aleatório, depois, como fixo em ambos os delineamentos.

### **3.6 Programa computacional**

As rotinas computacionais do SAS para execução de todas as análises estatísticas envolvidas nesse estudo estão apresentadas no Anexo 1B.

O comando “proc mixed” ativa o procedimento, o qual processa o arquivo de dados, utilizando o método da verossimilhança restrita -REML (método default) para estimação dos componentes de variância ( $\sigma_p^2$  e  $\sigma^2$ ); “class” especifica os nomes das variáveis classificatórias do modelo, de acordo com o input; “model” informa a variável resposta e os efeitos fixos relacionados; “random” especifica os efeitos aleatórios e a opção “solution” especifica os valores preditos destes efeitos aleatórios. O comando “repeated” especifica a estrutura da matriz R (a estrutura de G é definida em “random”). A opção “type =” define a estrutura de covariância escolhida entre diversas existentes no proc mixed, “sp” indica que a estrutura escolhida foi a espacial, em duas dimensões (denominadas neste estudo de “c” de coluna e “l” de linha), cuja função de covariância é o modelo esférico representado por “sph”.

As dimensões “c-coluna” e “l-linha” correspondem as coordenadas geográficas de cada parcela, em metros, e seus valores foram obtidos a partir do centro da primeira parcela do experimento, que foi considerada o ponto inicial (0, 0), até o centro da parcela em questão.



## 4 RESULTADOS E DISCUSSÃO

Na Tabela 4 estão apresentadas as eficiências relativas da análise considerando o controle local do delineamento, usando o modelo linear Gauss - Markov Normal, em relação à todas as demais formas de análise consideradas para os distintos experimentos de feijão e milho. A eficiência de um modelo em relação à outro pode ser avaliada por meio da redução no valor da variância residual ( $\sigma^2$ ). Vale ressaltar que os resultados estão agrupados segundo dois aspectos:

- 1) Análises que não consideram (delineamento inteiramente casualizado) e que consideram o controle local (delineamento látice ou em blocos casualizados completos);
- 2) Tipos de modelagem: GMN, MM, PPD e AE, descritos a seguir.
  - 2.a) Modelo Gauss-Markov Normal (GMN): os erros experimentais das parcelas são tomados como independentes;
  - 2.b) Modelo de Médias Móveis (MM): o índice ambiental utilizado como covariável corresponde à média dos valores observados de parcelas vizinhas;
  - 2.c) Modelo de Papadakis (PPD): o índice ambiental utilizado como covariável corresponde à média dos resíduos de parcelas vizinhas;
  - 2.d) Modelo de Análise Espacial (AE): os erros experimentais das parcelas são tomados como dependentes.

O que se espera ao analisar a Tabela 4 é revelar a eficiência do controle local (comparando as colunas GMN das duas seções); verificar se as formas de análise espacial recuperaram a eficiência da análise (caso o controle local tenha sido eficiente), o que pode ser feito analisando as colunas MM, PPD e AE para a ausência de controle local; e, finalmente, verificar se as formas de análise

espacial melhoraram de alguma maneira o delineamento com controle local (comparando as colunas da segunda seção, MM, PPD e AE com a coluna GMN).

**TABELA 4.** Eficiências relativas (ER) em % da análise considerando o controle local (CL) do delineamento, usando o modelo Gauss-Markov Normal (GMN) em relação a todas as demais formas de análise para os dados de produção de grãos (g/parcela) dos experimentos de feijão e milho

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN <sup>1</sup>	MM <sup>2</sup>	PPD <sup>3</sup>	AE <sup>4</sup>	GMN	MM	PPD	AE
1	105,9	103,9	103,6	105,9	100,0 (2173,70)	100,2	100,3	100,0
2	124,1	114,4	116,1	124,1	100,0 (10573,64)	100,1	<b>99,4<sup>5</sup></b>	100,0
3	106,3	<b>99,7</b>	<b>98,4</b>	106,3	100,0 (9166,33)	<b>98,5</b>	<b>98,1</b>	100,0
4	115,8	116,4	112,5	115,8	100,0 (3238,48)	<b>96,2</b>	100,5	100,0
5	116,5	102,3	101,3	116,5	100,0 (3837,36)	<b>99,6</b>	<b>99,8</b>	100,0
6	151,8	121,5	128,3	151,8	100,0 (11120,85)	111,5	<b>97,0</b>	100,0
7	149,8	<b>97,3</b>	104,1	149,8	100,0 (10381,83)	<b>96,3</b>	102,4	100,0
8	106,9	108,1	108,5	106,9	100,0 (57039,64)	<b>93,6</b>	<b>79,3</b>	100,0
9	118,7	<b>59,1</b>	<b>65,3</b>	118,7	100,0 (22200,33)	<b>61,6</b>	<b>92,6</b>	100,0
10	132,2	<b>91,1</b>	120,1	132,2	100,0 (18449,42)	<b>95,8</b>	101,3	100,0
11	111,8	107,2	104,9	111,8	100,0 (3071,83)	100,1	100,3	100,0
12	101,2	<b>94,1</b>	<b>92,1</b>	101,2	100,0 (14141,11)	<b>94,0</b>	<b>97,2</b>	100,0
13	102,8	103,9	101,6	102,8	100,0 (25251,35)	<b>98,3</b>	101,5	100,0
14	<b>99,6</b>	<b>98,2</b>	<b>87,4</b>	<b>99,6</b>	100,0 (38852,29)	<b>99,4</b>	100,8	100,0
15	122,0	125,8	123,5	122,0	100,0 (6435,01)	<b>84,1</b>	<b>78,5</b>	100,0
16	119,2	122,6	117,6	119,2	100,0 (4631,32)	<b>88,9</b>	102,5	100,0
17	<b>98,7</b>	<b>89,5</b>	104,9	<b>98,7</b>	100,0 (3383,53)	<b>81,5</b>	<b>97,9</b>	100,0
18	127,5	125,6	<b>67,9</b>	127,5	100,0 (1967,26)	104,3	<b>95,8</b>	100,0
19	<b>96,9</b>	<b>92,7</b>	101,4	<b>96,9</b>	100,0 (5781,85)	<b>94,8</b>	<b>93,6</b>	100,0
20	102,9	<b>96,7</b>	<b>77,9</b>	102,9	100,0 (2898,94)	<b>83,1</b>	<b>36,8</b>	100,0
21	<b>97,1</b>	101,4	101,4	<b>97,1</b>	100,0 (4227,11)	104,5	103,0	100,0

...continua...

TABELA 4, Cont.

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN	MM	PPD	AE	GMN	MM	PPD	AE
22	<b>96,1</b>	<b>99,3</b>	<b>98,9</b>	<b>96,1</b>	100,0 (3936,35)	<b>98,4</b>	102,3	100,0
23	130,4	126,5	134,5	130,4	100,0 (3413,04)	105,7	<b>83,2</b>	100,0
24	<b>99,3</b>	100,5	100,4	<b>99,3</b>	100,0 (3816,05)	<b>94,4</b>	<b>95,8</b>	100,0
25	105,2	107,7	108,1	105,2	100,0 (3244,59)	<b>78,0</b>	<b>89,2</b>	100,0
26	104,8	103,0	<b>97,5</b>	104,8	100,0 (2077,78)	<b>81,1</b>	<b>56,2</b>	100,0
27	125,4	101,7	113,8	125,4	100,0 (2233,73)	<b>95,9</b>	106,5	100,0
28	<b>98,9</b>	105,1	<b>98,4</b>	<b>98,9</b>	100,0 (3337,56)	<b>86,9</b>	104,8	100,0
29	<b>98,7</b>	<b>99,2</b>	<b>99,5</b>	<b>98,7</b>	100,0 (4631,00)	100,8	102,1	100,0
30	<b>96,1</b>	100,2	<b>95,0</b>	<b>96,1</b>	100,0 (6748,50)	104,6	102,3	100,0
31	<b>97,9</b>	102,9	<b>97,3</b>	<b>97,9</b>	100,0 (1815,49)	105,0	<b>79,4</b>	100,0
32	<b>96,1</b>	<b>91,2</b>	<b>97,2</b>	<b>96,1</b>	100,0 (5933,44)	<b>94,7</b>	<b>88,1</b>	100,0
33	111,3	108,5	110,1	111,3	100,0 (3862,88)	102,2	<b>76,1</b>	100,0
34	105,6	109,2	108,3	105,6	100,0 (3612,11)	<b>84,6</b>	<b>91,9</b>	100,0
35	<b>98,8</b>	<b>78,8</b>	108,7	<b>98,8</b>	100,0 (6364,00)	<b>81,9</b>	107,1	100,0
36	154,6	129,3	101,7	154,6	100,0 (1620,50)	100,2	<b>99,0</b>	100,0
37	116,3	100,8	<b>98,3</b>	116,3	100,0 (744097,20)	<b>96,6</b>	100,2	100,0
38	115,9	109,4	103,0	115,9	100,0 (182116,67)	100,1	<b>98,4</b>	100,0
39	101,2	<b>73,1</b>	<b>80,2</b>	101,2	100,0 (79871,45)	<b>73,1</b>	<b>86,2</b>	100,0
40	103,9	<b>93,6</b>	<b>93,2</b>	103,9	100,0 (40367,04)	<b>92,4</b>	<b>92,6</b>	100,0
41	106,4	104,0	106,9	106,4	100,0 (479899,95)	100,5	100,9	100,0
42	126,4	104,1	<b>76,0</b>	126,4	100,0 (1088707,53)	104,1	<b>90,0</b>	100,0
43	108,3	<b>97,0</b>	102,3	108,3	100,0 (45304,59)	<b>97,0</b>	101,5	100,0
Média ( $\bar{X}$ )	111,8	102,7	101,6	111,8	100,0	<b>94,5</b>	<b>93,8</b>	100,0
Amplitude (A)	58,5	70,2	69,2	58,5	0,0	49,9	70,3	0,0

(.): os valores dentro de parênteses correspondem as estimativas das variâncias residuais ( $\hat{\sigma}^2$ 's) do modelo GMN com controle local.

<sup>1</sup>, <sup>2</sup>, <sup>3</sup> e <sup>4</sup>: modelo de Gauss-Markov Normal, de Médias Móveis, de Papadakis e de Análise Espacial, respectivamente.

<sup>5</sup>: os valores em negrito indicam ineficiência do modelo GMN, com o controle local do delineamento, em relação a determinado model.

Valores de eficiências relativas (ER) acima de 100% significam que houve uma redução na estimativa da variância residual ( $\hat{\sigma}^2$ ) de (ER-100)% quando se utilizou o modelo Gauss-Markov Normal (GMN) com o controle local do delineamento comparativamente a alguma outra forma de análise.

Comparando as colunas GMN das duas seções, observando os valores superiores a 100% da primeira coluna da Tabela 4, verifica-se que o controle local foi eficiente em 72% dos experimentos. O valor médio de ER, proporcionado pelo modelo GMN sem o controle local, foi de 111,8% indicando que houve uma redução média na variância residual de 11,8% quando foi utilizado o controle local do delineamento na análise.

É preciso notar que a ineficiência do controle local relativamente à análise por meio do modelo GMN, sem controle local, ocorreu em 11 dos 20 experimentos de feijão, situados entre os números 17 e 36 da Tabela 1 (todos experimentos pequenos, com 25 tratamentos, 50 parcelas e uma linha de dois metros). Estes experimentos foram plantados contiguamente no campo experimental e, neste caso, poderiam ser agrupados em um só experimento maior, desde que no planejamento a aleatorização conjunta de todos os tratamentos fosse permitida.

No caso dos experimentos grandes, uma mudança no sentido dos blocos (os blocos foram instalados perpendicularmente à direção das curvas de nível) poderia ter contribuído para aumentar ainda mais a eficiência do controle local. Como pode ser observado no Anexo 2B, apesar de não haver um padrão espacial, nos gráficos da esquerda (gráficos das observações) há grupos de cores que se desfazem nos gráficos da direita (gráficos dos resíduos), principalmente no sentido horizontal, sugerindo que a blocagem para os experimentos grandes 1, 2 e 3, por exemplo, poderia ser no sentido horizontal (sentido das linhas).

Após os resultados anteriores indicarem que o controle local utilizado foi eficiente para a maioria dos experimentos avaliados, foi verificado que as formas de análise espacial da primeira seção, MM, PPD e AE, em média não recuperaram a eficiência da análise pois apresentaram valores médios de ER superiores a 100% (102,7%; 101,6% e 111,8%, respectivamente).

Na presença de controle local, os modelos espaciais MM e PPD contribuíram para melhorar a eficiência do controle local em 65% e 58% dos experimentos, respectivamente. Os valores médios de ER inferiores a 100% apresentados na Tabela 4 pelos modelos MM (94,5%) e PPD (93,8%) comprovaram a melhoria proporcionada por eles. Contudo, não foi possível concluir com estes dados sobre a AE, pois as estimativas da variância residual não se modificaram (ou seja, a AE apresentou mesma ER média, de 100%, que GMN), apenas surgiu outro parâmetro estimado de covariância denominado alcance da correlação espacial ( $\rho$ ).

Na Tabela 5 encontram-se as razões entre as estimativas das variâncias de progênies ( $\hat{\sigma}_p^2$ 's) provenientes das diferentes formas de análise em relação as estimativas oriundas da análise com o controle local do delineamento, considerando o modelo GMN. O que se espera é que não haja alteração em  $\hat{\sigma}_p^2$  proveniente das diferentes modelagens espaciais comparativamente a  $\hat{\sigma}_p^2$  do modelo GMN, tanto na presença quanto na ausência de controle local.

**TABELA 5.** Razões ( $R_{\hat{\sigma}_p^2}$ ) em % entre as estimativas das variâncias de progênies ( $\hat{\sigma}_p^2$ 's) provenientes das diferentes formas de análise em relação a estimativa oriunda da análise considerando o controle local (CL) do delineamento, usando o modelo Gauss-Markov Normal (GMN) para os dados de produção de grãos (g/parcela) dos experimentos de feijão e milho

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN <sup>1</sup>	MM <sup>2</sup>	PPD <sup>3</sup>	AE <sup>4</sup>	GMN	MM	PPD	AE
1	103,5	110,3	105,8	103,5	100,0 (446,97)	<b>103,6<sup>3</sup></b>	<b>100,8</b>	100,0
2	94,8	102,9	100,4	94,8	100,0 (5524,00)	<b>101,9</b>	<b>101,0</b>	100,0
3	107,8	106,5	107,9	107,8	100,0 (6444,69)	<b>105,0</b>	<b>100,1</b>	100,0
4	97,7	96,8	103,2	97,7	100,0 (2902,89)	92,7	95,3	100,0
5	166,6	179,8	212,1	166,6	100,0 (283,25)	<b>193,9</b>	<b>108,0</b>	100,0
6	49,8	67,5	50,9	49,8	100,0 (4410,50)	77,0	<b>112,8</b>	100,0
7	101,5	44,7	93,7	101,5	100,0 (3351,76)	47,8	97,0	100,0
8	103,7	104,8	102,4	103,7	100,0 (11130,39)	95,9	<b>160,4</b>	100,1
9	0,0	144,9	231,5	0,0	100,0 (9633,48)	<b>114,8</b>	0,0	100,2
10	121,6	144,4	76,3	121,6	100,0 (7337,55)	<b>140,5</b>	<b>105,9</b>	100,1
11	90,4	103,4	99,7	90,4	100,0 (1040,89)	<b>101,9</b>	99,8	100,0
12	99,2	104,9	107,4	99,2	100,0 (6732,42)	<b>104,7</b>	<b>100,2</b>	100,0
13	95,6	98,1	93,5	95,6	100,0 (5275,51)	<b>107,8</b>	<b>100,9</b>	100,0
14	100,3	107,7	117,3	100,3	100,0 (19215,57)	<b>106,1</b>	88,9	100,0
15	78,0	75,6	78,9	78,0	100,0 (3112,26)	80,8	<b>121,7</b>	100,0
16	93,9	93,8	100,3	93,9	100,0 (6024,25)	<b>100,7</b>	98,8	100,0
17	101,1	74,5	81,2	101,1	100,0 (2105,54)	78,6	86,9	100,1
18	81,8	75,3	176,3	81,8	100,0 (1481,84)	96,7	<b>110,2</b>	100,0
19	125,0	90,6	0,0	125,0	100,0 (357,27)	73,5	70,4	100,2
20	99,3	113,7	113,7	99,3	100,0 (2497,32)	<b>118,4</b>	<b>160,2</b>	100,0
21	102,8	67,8	96,7	102,8	100,0 (2222,00)	64,3	99,6	100,0

...continua...

TABELA 5, Cont.

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN	MM	PPD	AE	GMN	MM	PPD	AE
22	122,0	99,7	75,0	122,0	100,0 (346,65)	76,2	98,6	100,0
23	98,1	94,7	96,1	98,1	100,0 (4187,15)	<b>100,3</b>	<b>120,7</b>	100,0
24	100,0	100,0	100,0	100,0	100,0 (0,00)	100,0	100,0	100,0
25	78,2	51,7	70,3	78,2	100,0 (296,63)	<b>115,6</b>	<b>204,3</b>	100,0
26	102,6	107,7	114,2	102,6	100,0 (1457,33)	<b>105,0</b>	<b>156,1</b>	100,0
27	102,8	134,8	93,6	102,8	100,0 (1824,02)	<b>125,8</b>	94,0	100,0
28	100,8	85,5	108,5	100,8	100,0 (2264,73)	90,7	89,6	100,0
29	100,0	100,0	100,0	100,0	100,0 (0,00)	100,0	100,0	100,0
30	105,5	98,0	113,0	105,5	100,0 (2427,16)	91,2	<b>101,1</b>	100,0
31	100,8	89,0	102,9	100,8	100,0 (2346,14)	87,6	<b>116,1</b>	100,0
32	105,1	120,7	103,9	105,1	100,0 (2258,19)	<b>116,2</b>	<b>136,0</b>	100,0
33	90,5	98,5	87,9	90,5	100,0 (2302,92)	<b>101,2</b>	<b>141,9</b>	100,0
34	97,5	91,3	98,6	97,5	100,0 (4095,27)	89,1	<b>105,0</b>	100,0
35	101,1	122,2	64,3	101,1	100,0 (3420,91)	<b>119,0</b>	84,2	100,0
36	87,0	104,8	100,4	87,0	100,0 (3404,99)	<b>103,3</b>	<b>100,7</b>	100,0
37	49,8	98,0	114,4	49,8	100,0 (105895,29)	<b>109,7</b>	87,1	100,0
38	104,5	113,4	115,4	104,5	100,0 (128540,74)	<b>106,1</b>	<b>103,3</b>	100,0
39	99,0	73,7	110,9	99,0	100,0 (19521,91)	73,8	<b>103,1</b>	100,0
40	100,0	100,0	100,0	100,0	100,0 (0,00)	100,0	100,0	100,0
41	93,2	101,5	84,3	93,2	100,0 (113809,23)	<b>102,7</b>	99,7	100,0
42	98,2	100,1	108,2	98,2	100,0 (1795842,77)	<b>100,1</b>	<b>111,6</b>	100,0
43	100,0	100,0	100,0	100,0	100,0 (0,00)	100,0	100,0	100,0
Média ( $\bar{X}$ )	96,5	99,8	102,6	96,5	100,0	<b>100,5</b>	<b>106,3</b>	100,0
Amplitude (A)	166,6	135,1	231,5	166,6	0,0	146,1	204,3	0,2

(.): os valores dentro de parênteses correspondem as estimativas das variâncias de progênies ( $\hat{\sigma}_p^2$ 's) do modelo GMN com controle local.

<sup>1</sup>, <sup>2</sup>, <sup>3</sup> e <sup>4</sup>: modelo de Gauss-Markov Normal, de Médias Móveis, de Papadakis e de Análise Espacial, respectivamente.

<sup>5</sup>: os valores em negrito correspondem a  $R_{\hat{\sigma}_p^2}$  superiores a 100% para presença de controle local.



Nota-se, por meio dos resultados da Tabela 4 e 5, que tanto na ausência como na presença de controle local a análise espacial (AE) não modificou  $\hat{\sigma}^2$  e  $\hat{\sigma}_p^2$  em relação à análise GMN para todos os experimentos, pois ambas as análises apresentaram os mesmos valores médios e as mesmas amplitudes de ER e de  $R_{\hat{\sigma}_p^2}$ .

Na Tabela 5, observa-se em negrito que em 56% dos experimentos analisados com o controle local do delineamento, ambas as análises, MM e PPD, proporcionaram  $\hat{\sigma}_p^2$ 's infladas ( $R_{\hat{\sigma}_p^2} > 100\%$ ) em relação às mesmas estimativas provenientes da análise GMN. O valor médio de  $R_{\hat{\sigma}_p^2}$  apresentado pela análise PPD, de 106,3%, evidencia que maior alteração em  $\hat{\sigma}_p^2$  foi proporcionada por esta análise.

Como o controle local utilizado foi eficiente para a maioria dos experimentos avaliados, a discussão anterior se restringiu apenas a situação de presença de controle local.

Pode-se verificar, na Tabela 6, as razões entre as estimativas das semi-amplitudes (SA) percentuais dos intervalos de confiança de  $\hat{\sigma}_p^2$ 's para cada forma de análise considerada em relação à análise com o controle local do delineamento, considerando o modelo GMN.

Quando se comparam as diferentes formas de análise por meio de SA, menores valores indicam maiores precisões no cálculo de  $\hat{\sigma}_p^2$ . O interesse maior, neste caso, foi comparar, somente na presença de controle local (visto que o controle local foi eficiente para a maioria dos experimentos), os valores de SA dos modelos espaciais (MM, PPD e AE) em relação à este valor proveniente do modelo GMN.

**TABELA 6.** Razões ( $R_{SA}$ ) em % entre as estimativas das semi-amplitudes (SA) percentuais dos intervalos de confiança de  $\hat{\sigma}_p^2$ 's dos dados de produção de grãos (g/parcela) dos experimentos de feijão e milho, para cada forma de análise considerada em relação a análise com controle local (CL) do delineamento, usando o modelo Gauss-Markov Normal (GMN)

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN <sup>1</sup>	MM <sup>2</sup>	PPD <sup>3</sup>	AE <sup>4</sup>	GMN	MM	PPD	AE
1	98.9	94.0	96.0	98.9	100,0 (44,9)	<b>98.7</b> <sup>5</sup>	<b>99,6</b>	100,0
2	111.2	102.4	104.4	111.2	100,0 (24,9)	<b>99.6</b>	<b>99,6</b>	100,0
3	96.8	94.9	94.0	96.8	100,0 (21,6)	<b>96.8</b>	<b>99,1</b>	100,0
4	103.3	104.6	101.6	103.3	100,0 (30,5)	103.0	102,6	100,0
5	65.9	56.0	49.1	65.9	100,0 (199,1)	<b>51.8</b>	<b>93,8</b>	100,0
6	184.6	126.3	159.8	184.6	100,0 (101,2)	126.8	102,8	100,0
7	116.7	156.0	97.7	116.7	100,0 (109,9)	148.6	105,1	100,0
8	93.6	93.7	95.4	93.6	100,0 (182,5)	103.0	<b>69,2</b>	<b>99,9</b>
9	-	40.4	29.5	-	100,0 (935,0)	<b>51.8</b>	-	<b>99,8</b>
10	99.4	77.9	125.0	99.4	100,0 (128,7)	<b>80.3</b>	100,1	100,0
11	119.4	102.1	101.8	119.4	100,0 (62,5)	<b>99.4</b>	100,6	100,0
12	101.1	94.0	91.8	101.1	100,0 (63,7)	<b>94.2</b>	<b>98,1</b>	100,0
13	123.0	119.7	127.1	123.0	100,0 (680,8)	<b>83.9</b>	117,7	100,0
14	99.1	94.2	80.2	99.1	100,0 (162,3)	<b>96.7</b>	116,1	100,0
15	134.8	181.7	135.3	134.8	100,0 (475,0)	148.3	<b>53,8</b>	100,0
16	104.1	105.8	102.2	104.1	100,0 (148,6)	<b>95.7</b>	102,0	100,0
17	96.4	134.8	168.8	96.4	100,0 (214,7)	106.2	122,4	100,0
18	181.4	208.6	59.6	181.4	100,0 (172,4)	115.0	<b>90,7</b>	100,0
19	0.0	359.4	-	0.0	100,0 (400,1.10 <sup>18</sup> )	140,1.10 <sup>12</sup>	832,8.10 <sup>13</sup>	<b>85,5</b>

...continua...

TABELA 6, Cont.

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento				
	GMN	MM	PPD	AE	GMN	MM	PPD	AE	
20	97,3	89,3	72,7	97,3	100,0	(160,2)	<b>81,6</b>	<b>52,7</b>	100,0
21	90,7	302,0	114,8	90,7	100,0	(274,1)	415,6	106,8	100,0
22	0,0	131,1	978,5.10 <sup>6</sup>	0,0	100,0	(223,0.10 <sup>9</sup> )	328,7.10 <sup>8</sup>	1147,5	<b>99,9</b>
23	107,2	108,3	112,5	107,2	100,0	(132,9)	101,8	<b>83,7</b>	100,0
24	-	-	-	-	100,0	(-)	-	-	-
25	241,8.10 <sup>5</sup>	268,6.10 <sup>26</sup>	260,2.10 <sup>10</sup>	241,8.10 <sup>5</sup>	100,0	(578,2.10 <sup>9</sup> )	<b>0,0</b>	<b>0,0</b>	<b>99,4</b>
26	96,8	91,0	82,2	96,8	100,0	(195,8)	<b>80,7</b>	<b>53,7</b>	100,0
27	111,9	77,0	112,6	111,9	100,0	(174,2)	<b>84,7</b>	136,1	100,0
28	97,3	145,4	99,4	97,3	100,0	(193,2)	121,1	129,7	100,0
29	-	-	-	-	100,0	(-)	-	-	-
30	76,9	133,7	65,4	76,9	100,0	(617,4)	206,6	104,5	<b>99,9</b>
31	98,2	114,8	97,1	87,7	100,0	(113,9)	118,0	<b>85,5</b>	100,0
32	79,4	57,1	83,8	79,4	100,0	(530,3)	<b>66,6</b>	<b>49,0</b>	100,0
33	136,7	116,9	141,5	99,7	100,0	(227,5)	<b>73,8</b>	<b>71,8</b>	100,0
34	105,2	115,8	110,0	105,2	100,0	(123,3)	133,6	106,5	100,0
35	95,5	59,6	514,5	95,5	100,0	(264,6)	<b>63,7</b>	191,8	100,0
36	130,1	109,3	100,4	130,1	100,0	(92,0)	100,9	<b>99,5</b>	100,0
37	20801,9	102,7	66,2	20801,9	100,0	(365,6)	<b>68,9</b>	157,0	<b>99,9</b>
38	99,4	96,9	95,2	99,4	100,0	(87,5)	<b>99,8</b>	<b>99,0</b>	100,0
39	100,8	99,7	86,7	100,8	100,0	(163,4)	<b>99,7</b>	<b>92,2</b>	100,0
40	-	-	-	-	100,0	(-)	-	-	-
41	117,8	103,8	154,4	117,8	100,0	(168,9)	<b>97,9</b>	102,2	100,0
42	104,1	100,5	95,9	104,1	100,0	(184,7)	100,5	<b>97,7</b>	100,0
43	-	-	-	-	100,0	(-)	-	-	-

(.): os valores dentro de parênteses correspondem as estimativas das semi-amplitudes (SA's) do modelo GMN com controle local.

-: indica que  $\hat{\sigma}_p^2$  é nula, não havendo, portanto, estimativa de sua semi-amplitude e da  $R_{SA}$ .

<sup>1</sup>, <sup>2</sup>, <sup>3</sup> e <sup>4</sup>: modelo de Gauss-Markov Normal, de Médias Móveis, de Papadakis e de Análise Espacial, respectivamente.

<sup>5</sup>: os valores em negrito correspondem a  $R_{SA} < 100\%$  para presença de controle local.

Comparando os modelos com controle local na Tabela 6, as análises MM, PPD e AE proporcionaram, respectivamente, 51%, 44% e 16% dos experimentos com menores SA's ( $R_{SA} < 100\%$ ) em relação às semi-amplitudes da análise GMN que foi tomada como referência. Portanto, os componentes de variância  $\sigma_p^2$ 's foram em geral estimados com maior precisão pelo método MM, comparativamente aos demais métodos espaciais.

Note que os experimentos 19, 22 e 25 apresentaram alguns valores de semi-amplitudes extremamente altos. Nestes experimentos as  $\hat{\sigma}_p^2$ 's não eram nulas.

Foram verificadas, ainda, as eficiências dos delineamentos avaliados em látice em relação aos delineamentos em blocos casualizados completos; os resultados estão disponíveis na Tabela 7.

**TABELA 7.** Eficiências relativas (ER) em % dos 38 delineamentos em látice em relação ao delineamento em blocos casualizados completos correspondente, considerando o efeito de repetição como sendo fixo, o efeito de bloco como sendo aleatório e o efeito de tratamento ora sendo tratado como fixo, ora como aleatório em ambos os delineamentos

Experimento	Tratamento Fixo	Tratamento aleatório
1	<b>100,7<sup>1</sup></b>	<b>103,5</b>
2	<b>108,8</b>	<b>106,0</b>
3	<b>101,8</b>	<b>104,2</b>
4	<b>105,3</b>	<b>104,0</b>
5	100,0	<b>179,0</b>
6**	<b>119,1</b>	83,5
7	<b>110,8</b>	<b>118,6</b>
8*	<b>103,4</b>	98,6
9	<b>105,1</b>	-
10	<b>100,6</b>	<b>118,0</b>
11	<b>101,8</b>	<b>100,5</b>
12	100,0	100,0
13	99,8	100,0
14	100,0	100,0
15**	<b>110,8</b>	91,8
16	<b>104,9</b>	<b>102,5</b>
17	100,0	100,0
18	99,7	100,0
19	100,0	<b>100,2</b>
20*	<b>101,2</b>	99,9
21	100,0	100,0
22	<b>104,0</b>	100,0
23	99,7	<b>102,5</b>
24	100,0	-
25**	<b>103,4</b>	86,2
26	99,6	<b>101,0</b>
27	<b>101,0</b>	<b>106,7</b>
28	99,7	100,0
29	99,7	-
30	100,0	100,0
31	100,0	100,0
32	<b>100,7</b>	100,0
33	<b>100,3</b>	100,0
34	<b>105,7</b>	100,0
35	100,0	100,0
36	100,0	100,0
37*	<b>101,1</b>	98,9
38	<b>105,5</b>	<b>106,3</b>

\*: diferença entre as ER's provenientes de tratamento fixo e aleatório menor ou igual a 5% ;

\*\* : diferença entre as ER's provenientes de tratamento fixo e aleatório maior a 5% .

-: indica que  $\hat{\sigma}_p^2$  é nula para os delineamentos em látice ou em blocos casualizados completos, não existindo variância desta estimativa e, portanto, não sendo possível o cálculo automático da ER como delineado.

<sup>1</sup>: os valores em negrito correspondem aER's superiores 100%.

Valores de ER superiores a 100%, destacados em negrito na Tabela 7, indicam presença de variância de blocos ( $\sigma_b^2$ ), evidenciando, portanto, que o modelo em látice foi eficiente.

No estudo em questão, 17 (considerando tratamentos de efeitos fixos) e 21 (considerando tratamentos de efeitos aleatórios) dos 38 experimentos apresentaram componentes da variância nulos para blocos. Nestes casos, a análise em látice foi ineficiente.

Detectou-se uma correlação de Pearson de -0,19 entre as duas colunas de eficiências da Tabela 7. Este resultado surpreendente indica que há diferença na efetividade do látice quando se consideram tratamentos como de efeito fixo ou aleatório.

Como pode ser observado no Anexo 2A, uma análise posterior das correlações de Spearman ( $r$ ) para os efeitos estimados de tratamentos indicou que na maioria dos casos houve alterações no ordenamento das progênies para seleção. Isto implica em seleção de progênies diferentes, em termos de produção de grãos, quando seus efeitos são considerados fixos ou aleatórios. As maiores alterações ocorreram para experimentos grandes (pois apresentaram baixos valores de  $r$ ) sugerindo que, na escolha da natureza do efeito do tratamento, se fixo ou aleatório, um dos critérios que deve ser considerado pelos melhoristas é o número de progênies a serem avaliadas.

Já Duarte (2000), utilizando dados de produção de grãos de 32 experimentos de avaliação de linhagens de soja, comparou por meio de correlações de Pearson e de Spearman diferentes modelos, ora considerando efeito de linhagens como fixo, ora como aleatório. Na maioria dos experimentos os modelos selecionaram praticamente os mesmos genótipos por terem apresentado altas correlações. Contudo, o autor acrescenta que apesar da concordância entre os modelos quanto ao ordenamento dos genótipos, os valores das respostas genotípicas médias poderão diferir substancialmente entre estes modelos se a herdabilidade for

baixa (inferior a 0,20). O autor ainda aconselha que uma escolha equivocada do modelo de análise pode comprometer a eficiência do programa de melhoramento no sentido que materiais genéticos pouco promissores poderão ser avaliados e mantidos em próximos ciclos seletivos acarretando em desperdício de tempo e de recursos.

Bueno Filho (2000) também obteve praticamente o mesmo ordenamento dos genótipos pelas quatro formas de análise avaliadas em seu estudo considerando as combinações de efeitos para tratamento e bloco, respectivamente: 1) fixo, aleatório; 2) fixo, fixo; 3) aleatório, aleatório e 4) aleatório, fixo.

Nos casos dos experimentos da Tabela 7 que receberam a indicação \*\*, houve eficiência do látice quando se considerou o efeito de tratamento como fixo e ineficiência quando ele foi considerado aleatório. Isto indica confundimento parcial entre o efeito de bloco e efeitos de tratamentos que compõem o bloco, ou seja, o que se chama de efeito isolado de bloco ( $b_j$ ) na verdade é adicionado um lixo correspondente aos efeitos de tratamentos ( $p_1+p_2+\dots+p_p$ ) deste bloco. Resultados semelhantes a este não foram encontrados na literatura.

É evidente que, em média, experimentos pequenos (com poucos tratamentos) devem sofrer mais com este tipo de efeito, pois uma amostra representativa de tratamentos num bloco tende a apresentar efeito médio nulo. Assim, nos experimentos pequenos avaliados o confundimento foi mais importante que  $b_j$  isolado. Estes resultados indicam a ineficiência do látice para experimentos pequenos com a cultura do feijão, embora os valores do cálculo tradicional desta eficiência, ou seja, considerando efeito de bloco como aleatório e de tratamento como fixo, possam se apresentar como altos na Tabela 7.

Um critério bastante interessante para a comparação global entre formas de análise distintas é o Critério de Informação de Akaike (AIC). O modelo que apresentar maior valor de AIC deve ser preferido em relação ao outro. Duarte

(2000) comenta que modelos com estrutura de covariância complexa (com maior número de parâmetros) tendem a exibir uma melhor qualidade de ajustamento do que os modelos de estrutura mais simples. Na verdade, isto depende da precisão dos próprios componentes da variância.

Na Tabela 9 encontram-se as razões entre os valores de AIC para todas as formas de análise avaliadas em relação à análise com o controle local do delineamento, considerando o modelo GMN.



**TABELA 8.** Razões ( $R_{AIC}$ ) em % entre as estimativas do Critério de Informação de Akaike (AIC) dos dados de produção de grãos (g/parcela) dos experimentos de feijão e milho, para cada forma de análise considerada em relação à análise com controle local (CL) do delineamento, usando o modelo Gauss-Markov Normal (GMN)

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN	MM	PPD	AE	GMN	MM	PPD	AE
1	100,3	100,3	100,2	100,4	100,0 (-5473,91)	100,0	100,0	100,0
2	100,9	100,5	100,6	100,9	100,0 (-7479,66)	100,0	100,0	100,0
3	100,3	99,9	99,8	100,3	100,0 (-7421,47)	<b>99,8</b>	99,9	100,0
4	100,6	100,7	100,6	100,7	100,0 (-2886,66)	<b>99,9</b>	<b>99,9</b>	100,0
5	101,2	100,3	100,4	101,3	100,0 (-1678,45)	<b>99,8</b>	100,0	100,1
6	102,3	101,1	101,2	102,5	100,0 (-667,20)	<b>99,5</b>	100,0	100,1
7	102,7	99,0	100,2	102,9	100,0 (-663,34)	<b>97,8</b>	100,0	100,2
8	101,9	102,0	102,0	102,1	100,0 (-512,45)	100,0	<b>99,5</b>	100,2
9	102,4	97,6	98,5	102,6	100,0 (-474,48)	<b>96,0</b>	99,1	100,2
10	103,2	101,2	102,0	103,4	100,0 (-380,70)	<b>99,0</b>	<b>99,9</b>	100,3
11	100,4	100,4	100,1	100,5	100,0 (-2673,30)	100,1	100,0	100,0
12	100,6	100,2	100,1	100,6	100,0 (-1532,44)	<b>99,7</b>	99,8	100,1
13	102,1	102,2	101,9	102,3	100,0 (-480,75)	100,1	100,0	100,2
14	101,9	101,9	101,4	102,1	100,0 (-501,34)	100,0	<b>99,8</b>	100,2
15	103,3	103,5	103,3	103,6	100,0 (-277,27)	99,8	<b>98,9</b>	100,4
16	103,0	103,3	103,0	103,4	100,0 (-275,31)	100,1	100,0	100,4
17	101,1	99,8	100,8	101,5	100,0 (-279,07)	<b>98,2</b>	99,4	100,4
18	102,2	102,0	100,6	102,6	100,0 (-267,38)	100,2	<b>99,7</b>	100,4
19	101,1	100,6	100,7	101,4	100,0 (-283,63)	99,5	<b>99,1</b>	100,4
20	101,1	101,2	99,7	101,4	100,0 (-278,27)	100,0	<b>96,9</b>	100,4
21	101,0	100,5	101,0	101,4	100,0 (-283,33)	<b>99,4</b>	100,0	100,4

...continua...

TABELA 8, Cont.

Experimento	Modelos sem o CL do delineamento				Modelos com o CL do delineamento			
	GMN	MM	PPD	AE	GMN	MM	PPD	AE
22	101,0	101,2	101,2	101,4	100,0 (-274,95)	100,1	100,0	100,4
23	102,1	102,0	102,1	102,4	100,0 (-285,96)	100,3	<b>99,4</b>	100,3
24	101,0	101,2	101,0	101,4	100,0 (-272,56)	100,1	<b>99,5</b>	100,4
25	101,2	101,3	101,4	101,6	100,0 (-270,95)	99,4	<b>99,3</b>	100,4
26	101,0	101,1	100,8	101,4	100,0 (-269,43)	99,8	<b>98,3</b>	100,4
27	101,6	101,1	100,6	102,0	100,0 (-273,79)	<b>99,9</b>	<b>99,9</b>	100,4
28	101,1	101,2	101,2	101,5	100,0 (-279,33)	100,0	<b>99,9</b>	100,4
29	101,1	101,2	101,1	101,5	100,0 (-276,93)	100,1	100,0	100,4
30	101,0	101,2	100,9	101,4	100,0 (-292,44)	100,2	100,0	100,3
31	101,0	101,0	100,7	101,4	100,0 (-269,75)	100,0	<b>98,9</b>	100,4
32	101,0	100,9	101,0	101,4	100,0 (-289,64)	99,8	<b>99,6</b>	100,3
33	101,6	101,7	101,4	102,0	100,0 (-281,97)	100,2	<b>99,5</b>	100,4
34	101,4	101,5	101,4	101,7	100,0 (-285,15)	99,8	<b>99,6</b>	100,4
35	101,2	100,3	100,8	101,5	100,0 (-293,29)	<b>99,2</b>	99,9	100,3
36	103,1	102,8	101,2	103,4	100,0 (-271,49)	100,3	<b>99,6</b>	100,4
37	100,6	100,2	100,1	100,7	100,0 (-2367,58)	<b>99,8</b>	99,9	100,0
38	100,8	100,7	100,4	100,9	100,0 (-1859,00)	100,0	100,0	100,1
39	99,7	97,6	98,4	100,0	100,0 (-390,65)	<b>97,8</b>	99,0	100,3
40	100,1	99,4	99,4	100,2	100,0 (-1735,08)	<b>99,4</b>	99,5	100,1
41	100,0	100,0	99,3	100,1	100,0 (-1157,18)	100,1	100,1	100,1
42	100,4	99,5	97,8	100,5	100,0 (-690,66)	99,6	<b>99,2</b>	100,1
43	99,7	98,9	99,3	99,9	100,0 (-422,25)	100,1	100,1	100,2
Média ( $\bar{X}$ )	101,3	100,8	100,7	101,5	100,0	<b>99,7</b>	<b>99,6</b>	100,2
Amplitude (A)	3,6	6,0	5,4	3,7	0,0	4,4	3,2	0,4

(.): os valores dentro de parênteses correspondem as estimativas dos AIC's do modelo GMN com controle local.

<sup>1</sup>: os valores em negrito correspondem a  $R_{AIC}$  inferiores a 100%.

Como pode ser observado pelos valores médios apresentados pelos diferentes modelos na Tabela 8, praticamente não houve alterações nos valores de AIC ( $R_{AIC}$  próximos de 100%) por meio do modelo GMN com controle local em relação a qualquer outro modelo sem controle local, confirmando a eficiência do delineamento utilizado na maioria dos experimentos.

Em todos os experimentos, considerando primeiramente a situação de ausência de controle local e, posteriormente presença, a AE forneceu praticamente os mesmos valores de AIC (valores médios de  $R_{AIC}$  semelhantes) comparativamente ao modelo GMN. Este resultado, adicionado a outro já comentado, de que não houve alteração na  $\hat{\sigma}^2$  dos experimentos quando do uso da AE, sugere que o modelo espacial não melhora a análise, apenas aumenta o número de parâmetros a serem estimados.

Nenhum experimento apresentou maior valor de AIC (menor  $R_{AIC}$ ) por meio do modelo de AE na ausência de controle local comparativamente ao modelo GMN com controle local. Isto também confirma que a estrutura de covariância espacial ( $\mathbf{R} = [f(d)] \sigma^2$ ) não garantiu ao modelo espacial um ajuste de qualidade superior em relação ao modelo de erros independentes ( $\mathbf{R} = I \sigma^2$ ). Assim, em nenhum caso estudado é aconselhável desprezar o controle local do delineamento e avaliá-lo de acordo com o modelo espacial.

Na presença de controle local, pode-se verificar, pelos valores médios de  $R_{AIC}$  da Tabela 8 (99,7%, 99,6% e 100,2% para MM, PPD e AE, respectivamente), que não houve grandes diferenças com relação a este critério entre os modelos espaciais e o modelo GMN (100%).

Resultados contrários a estes foram obtidos por Duarte (2000) para dados de produção de grãos em soja, em que a parametrização do modelo espacial foi satisfatória.

Contudo, o fato dos modelos PPD e MM terem apresentado menores

$\hat{\sigma}^2$ 's (como pode ser visto na Tabela 4 pelos menores valores de ER), sugerem que modelar a média via covariável (por meio de PPD ou MM) é, em geral, mais eficiente que modelar a própria covariável (por meio da AE).

Nota-se que nos casos estudados a AE não trouxe vantagens sobre os delineamentos planejados. Já os modelos PPD e MM apesar de terem contribuído para melhorias, recomenda-se adaptações aos métodos espaciais (por exemplo quanto ao número de vizinhos) pois qualquer ganho no melhoramento é de grande valia no sentido de detectar diferenças genéticas e reduzir os custos.

Também é de se esperar que, no caso de haver maior heterogeneidade intrabloco, o método PPD passe a ser o mais eficiente por utilizar diretamente os resíduos como covariável. Contudo, pelos resultados da Tabela 5, os métodos PPD e MM afetaram as estimativas dos componentes das variâncias de progênies ( $\hat{\sigma}_p^2$ 's). Pode-se suspeitar, então, que os modelos de Papadakis e de médias móveis são menos robustos que o da análise GMN para o ajuste aos dados, pois o modelo GMN é justificado pela aleatorização do experimento, já MM e PPD são formas particulares de análise espacial em modelos do tipo auto-regressivo, conforme elucidado por Atkinson (1969).

Os modelos MM e PPD são, na verdade, modelos não lineares complexos, e pode também estar havendo um “overfitting”, situação em que o controle local extremamente eficiente realizado com os vizinhos é, na verdade, parcialmente devido a diferenças genotípicas (vale lembrar que foram considerados apenas 4 vizinhos por parcela).

Como pode ser visto no Anexo 1A, com exceção do experimento 43 que sintomaticamente é bastante diferente dos demais, sendo uma situação de efeitos fixos de tratamentos, em parcelas subdivididas, o alcance ( $\rho$ ), que é o valor de distância para a independência entre as parcelas, não variou entre os modelos

espaciais sem e com controle local. Contudo, os valores dos alcances variaram entre 96 e 2400 metros de acordo com o experimento. Estes altos valores encontrados indicam que o modelo espacial não consegue melhorar a análise, pois embora sugira que quaisquer pares de parcelas do experimento sejam dependentes, o grau desta dependência deve ser baixo (próximo de zero). Seria dispensável fazer um modelo da covariância entre parcelas a distâncias dadas.

Outra alternativa é que isto talvez indique que as covariâncias sejam praticamente constantes para todos os pares de parcelas e, assim sendo, chegar-se-ia a uma situação de análise semelhante à clássica. Somente no caso do experimento 43, o baixo valor de alcance ( $\rho=2$ ) obtido por meio do modelo espacial sem controle local sugere que tal modelagem melhora a eficiência da análise.

## 5 CONCLUSÕES

Com base nos resultados dos experimentos de milho e feijão avaliados neste estudo, conclui-se que:

- i) Em geral não é eficiente a substituição do plano experimental pelas análises espaciais para realizar o controle local.
- ii) O controle local foi, em geral, eficiente e poderia ser, ainda mais, nos experimentos grandes, com a mudança na orientação dos blocos (sentido da blocagem).
- iii) Os métodos de análise de vizinhança, o de Papadakis e o das médias móveis, melhoraram o delineamento com controle local. Porém, o mesmo não ocorreu quando do uso da análise espacial.
- iv) Representar progênies por efeitos fixos ou aleatórios nos processos de estimação provocou alterações no ordenamento das progênies para seleção.
- v) A análise espacial clássica não altera estimativas dos componentes da variância de progênies ( $\hat{\sigma}_p^2$ ) em relação ao delineamento planejado, o que leva a uma vantagem conceitual deste modelo. Se for ineficiente, a análise espacial não altera a análise planejada. Já os modelos de médias móveis e de Papadakis, que são casos especiais de modelagens espaciais, são muito promissores para melhorar a eficiência do controle local, no entanto há dúvidas quanto à adequação destas modelagens, especialmente em ensaios pequenos e com poucos vizinhos para compor o índice ambiental.
- vi) O modelo GMN pode continuar sendo utilizado em delineamentos com controle local sem perda significativa de precisão em comparação aos modelos em que são utilizadas as análises de vizinhança e a análise espacial.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

ANDRÉ, C. M. G. **Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica.** 1999. 101 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) Universidade Federal de Lavras, Lavras, MG.

ATKINSON, A.C. The use of residuals as a concomitant variable. **Biometrika**, London, v.56, n. 1, p. 33-41, 1969.

BARTLETT, M. S. The approximate recovery of information from replicated field experiments with large blocks. **Journal of Agricultural Science**, Madison, v.28, p. 418-427, 1938.

BARTLETT, M. S. Nearest neighbour models in the analysis of field experiments. **Journal of the Royal Statistical Society – Serie B**, London, v. 40, n. 2, p. 147-174, 1978.

BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, n. 1, p. 20-25, Jan. 1994.

BOSE, R. C.; NAIR, K. R. Partially balanced incomplete block designs. **Sankhya**, Calcutta, v. 4, p. 337-372, 1939.

BROWNIE, C.; BOWMAN, D. T.; BURTON, J. W. Estimating spatial variation in analysis of data from yield trials: a comparison of methods. **Agronomy Journal**, Madison, v. 85, n. 6, p. 1244-1253, Nov./Dec. 1993.

BUENO FILHO, J. S. de S. **Modelos mistos na predição de valores genéticos aditivos em testes de progênies florestais.** 1997. 118 p. Tese (Doutorado em Agronomia/Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz se Queiroz, Piracicaba, SP.

BUENO FILHO, J. S. de S.; VENCOVSKY, R. Alternativas de ensaios em látice no melhoramento vegetal. **Pesquisa Agropecuária Brasileira**, v. 35, n. 2, p. 259-296, Fev. 2000.

BURGUESS, T. M.; WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. **The**

**Journal of Soil Science**, Oxford, v. 31, n. 2, p. 315-331, June 1980a.

COCHRAN, W. G.; COX, G. M. **Experimental designs**. 2. ed. New York: John Wiley, 1957. 611 p.

CRESSIE, N.; HARTFIELD, M. N. Conditionally specified gaussian models for spatial statistical analysis of field trials. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 1, n. 1, p. 60-77, Mar. 1996.

DAVIS, J. C. **Statistics and data analysis in geology**. 2. ed. New York: John Wiley, 1986. 550 p.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 p. Tese (Doutorado em Agronomia/Genética e Melhoramento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

EISENHART, C. The assumptions underlying the analysis of variance. **Biometrics**, Raleigh, v. 3, n. 1, p. 1-21, Mar. 1947.

ES, H. M. van; ES, C. L. van. Spatial nature of randomization and its effect on the outcome of field experiments. **Agronomy Journal**, Madison, v. 85, n. 2, p. 420-428, Mar./Apr. 1993.

FEDERER, W. T. Augmented (or hoonuiaku) designs. **Hawaiian Planter's Records**, Aiea, v. 55, p. 191-208, 1956.

FISHER, R. A. **The design of experiments**. Edinburgh: London: Oliver and Boyd, 1935.

GRONDONA, M. O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, Alexandria, v. 33, n. 4, p. 381-392, Nov. 1991.

HARSHBARGER, B. Triple rectangular lattices. **Biometrics**, Raleigh, v. 5, n. 1, p. 1-13, Mar. 1949.

HARVILLE, D. A. Extension of the Gauss-Markov theorem to include the estimation of random effects. **Annals Statistics**, Bethesda, v. 4, n.2, p. 384-395, Mar. 1976.

HARVILLE, D. A. BLUP (Best Linear Unbiased Prediction) and Beyond. In: GIANOLA, D.; HAMOND, K. (Ed.). **Advances in statistical methods for genetic improvement of livestock**. New York: Springer, 1990. p. 239-276.



HENDERSON, C. R. Estimation of genetic parameters (abstract). **Annals Mathematics Statistics**, Baltimore, v. 21, n. 3, p. 309-310, Sept. 1950.

HENDERSON, C. R.; KEMPTHORNE, O.; SEARLE, S. R.; von KROSICK, C. N. Estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Raleigh, v. 15, n. 1/4, p. 192-218, Dec. 1959.

HENDERSON, C. R. Selection index and expected genetic advance. **Statistical Genetics and Plant Breeding**, Washington, p. 141-163, 1963.

HENDERSON, C. R. Sire evaluation and genetic trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM, 1973, Champaign, Ill. **Proceedings...** Champaign: American society for animal Science, 1973. p. 10-41.

HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. New York: J. Wiley, 1994. 631 p.

KRIGE, D. G. A statistical approach to some basic mine evaluation problems on the witwatersrand. **Journal Chemical Metallurgical Mining Society of South Africa**, Johannesburg, v. 52, p. 119-139, 1951.

LITTEL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. **SAS<sup>®</sup> system for mixed models**. Cary, N. C: SAS Institute, 1996. 633 p.

MATHERON, G. Principles of geostatistics. **Economic Geology**, Lancaster, v. 58, p. 1246-1266, 1963.

MEAD, R.; CURNOW, R. N. **Statistical methods in agriculture and experimental biology**. New York: Chapman and Hall, 1983. 335 p.

PANTER, D. M.; ALLEN, F. L. Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. **Crop Science**, Madison, v. 35, n. 2, p. 397-405, Mar./Apr.1995a.

PANTER, D. M.; ALLEN, F. L. Using best linear unbiased predictions to enhance breeding for yield in soybean: II. Selection of superior crosses from a limited number of yield trials. **Crop Science**, Madison, v. 35, n. 2, p. 405-410, Mar./Apr. 1995b.

PAPADAKIS, J. S. Méthod statistique pou des experiences sur champ. **Bulletin Institute de L'Ameloration des Plantes à Salonique**, Paris, n. 23, 1937.

PATTERSON, H. D.; WILLIAMS, E. R. A new class of resolvable block designs. **Biometrika**, London, v. 63, n. 1, p. 83-92, July 1976.

PEARCE, S. C.; MOORE, C. S. Reduction of experimental error in perennial crops, using adjustment by neighbouring plots. **Experimental Agriculture**, Cambridge, v. 12, n. 3, p. 267-272, July 1976.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 13. ed. Piracicaba: Nobel/USP-ESALQ, 1990. 468 p.

RAMALHO, M. A.; FERREIRA, D. F.; OLIVEIRA, A. C. de. **A experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2000. 326 p.

RICKEY, F. D. Adjusting yields to their regression on a moving average, as a means of correcting for soil heterogeneity. **Journal of Agricultural Research**, Washinton, v. 27, n. 1, p. 79-90, Jan. 1924.

RIOS, M. C. D. **Alternativas de procedimento estatísticos para avaliação de genótipos em programas de melhoramento genético vegetal**. 1997. 59 p. (Dissertação-Mestrado em Estatística e Experimentação Agrônômica ) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

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

SEARLE, S. R. **Linear models**. New York: J. Wiley, 1971. 532 p.

STEEL, R. G. D.; TORRIE, J. H; DICKEY, D. A. **Principles and procedures of statistics**. 3. ed. New York: McGraw-Hill, 1997. 666 p.

SOH, A. C. Ranking parents by best linear unbiased prediction (BLUP) breeding values in oil palm. **Euphytica**, Dordrecht, v. 76, n. 1/2, p. 13-21, 1994.

SOUZA, E. A. de. **Alternativas experimentais na avaliação de progênies em programas de melhoramento genético vegetal**. 1997. 122 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, SP.

VIVALDI, L. J. Comparação entre métodos de análise espacial de experimentos de campo. **Pesquisa Agropecuária Brasileira**, Brasília, v. 25, n. 1, p. 77-84, jan. 1990.

WHITE, T. L.; HODGE, G. R.; DELORENZO, M. A. **Best linear prediction of breeding values in forest tree improvement.** Gainesville: University of Florida, 1986. p. 99-122. (Workshop of the Genetics and Breeding of southern forest trees, southern region information Exchange group 40. )

WHITE, T. L.; HODGE, G. R. **Predicting breeding values with applications in forest tree improvement.** Dordrecht: Kluwer Academic Publishers, 1989. 367 p.

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. 424-455, July 1936.

ZIMMERMAN, D. L.; HARVILLE, D. A. A random field approach to the analysis of field-plot experiments and other spatial experiments. **Biometrics**, Alexandria, v. 47, n.1, p. 223-239, Mar. 1991.

## ANEXOS

		Página
ANEXOS A		
ANEXO 1A	Estimativas dos alcances ( $\rho$ ) dos modelos espaciais sem ou com o controle local (CL) do delineamento e maiores distâncias (d) em metros entre duas parcelas referentes aos dados de produção de grãos (g/parcela) dos experimentos (Exp.) avaliados.....	77
ANEXO 2A	Correlações de Spearman (r) entre as estimativas de efeitos de tratamentos considerados ora como fixos, ora como aleatórios para os experimentos (Exp.) em látice.....	78
ANEXOS B		
ANEXO 1B	Principais comandos e rotinas do SAS empregados nas análises.....	79
ANEXO 2B	Gráficos de dispersão das observações e dos resíduos dos três maiores experimentos (1, 2 e 3) no campo experimental, com Coord Y representando as Linhas (m) e Coord X representando as Colunas (m).....	81

## ANEXOS A

**ANEXO 1A** Estimativas dos alcances ( $\rho$ ) dos modelos espaciais sem ou com o controle local (CL) do delineamento e maiores distâncias (d) em metros entre duas parcelas referentes aos dados de produção de grãos (g/parcela) dos experimentos (Exp.) avaliados

Exp.	$\rho$ (m)	d (m)
1	2048	46,803
2	2400	114,435
3	2400	122,424
4	1014	32,898
5	600	28,653
6	216	27,226
7	216	22,137
8	150	25,011
9	150	75,538
10	120	58,696
11	960	31,325
12	486	33,631
13	150	27,166
14	150	29,851
15	96	23,797
16	96	28,037
17	100	22,051
18	100	19,500
19	100	17,066
20	100	14,500
21	100	12,093
22	100	17,000
23	100	19,558
24	100	22,000
25	100	24,500
26	100	24,500
27	100	22,051
28	100	19,500
29	100	17,066
30	100	14,500
31	100	12,093
32	100	14,577
33	100	17,000
34	100	19,558
35	100	14,577
36	100	24,500
37	576	40,361
38	486	37,832
39	110	36,346
40	516	29,411
41	288	29,753
42	162	35,179
43	2 (sem CL) e 144 (com CL)	91,687

**ANEXO 2A** Correlações de Spearman ( $r$ ) entre as estimativas de efeitos de tratamentos considerados ora como fixos, ora como aleatórios para os experimentos (Exp.) em látice

Exp.	$r$
1	0,15
2	0,11
3	0,20
4	0,19
5	0,12
6	0,21
7	0,50
8	0,94
9	0,61
10	0,78
11	0,11
12	0,67
13	0,58
14	1,00
15	0,32
16	1,00
17	1,00
18	0,95
19	1,00
20	0,57
21	1,00
22	0,44
23	0,58
24	-0,01
25	0,04
26	0,79
27	0,52
28	0,77
29	0,01
30	1,00
31	1,00
32	0,48
33	0,10
34	0,14
35	1,00
36	1,00
37	0,34
38	0,30

## ANEXOS B

### ANEXO 1B Principais comandos e rotinas do SAS empregados nas análises

```
options ps=10000;

/* Produção de feijão do inverno de 2000 */

data arquivo;
input rep bloco trat parc c l y resDIC resLAT my mrDIC mrLAT;
cards;

1    1    2    1    1    17    330    20    11.8    335    17.5    -6.7
:    :    :    :    :    :    :    :    :    :    :    :
3    8    76    300    15    22.5    310    -50    -30.8    240    -25    10.2
;
run;

proc mixed cl;
title 'DIC normal – trat aleatorio';
class trat;
model y = ;
random trat/solution;
run;

proc mixed cl;
title 'DIC – Vizinhança – MM – trat aleatorio';
class trat;
model y = my;
random trat/solution;
run;

proc mixed cl;
title 'DIC – Vizinhança – PPD – trat aleatorio';
class trat;
model y = mrDIC;
random trat/solution;
run;

...continua...
```

**“ANEXO 1B, Cont.”**

```
Proc mixed cl;  
title 'Espacial – DIC';  
class trat;  
model y = ;  
random trat/solution;  
repeated / type=sp(sph)(c l);  
run;
```

```
proc mixed cl;  
title 'Latice normal – trat aleatorio';  
class rep bloco trat;  
model y = rep;  
random trat bloco(rep)/solution;  
run;
```

```
proc mixed cl;  
title 'Latice – Vizinhança – MM – trat aleatorio';  
class rep bloco trat;  
model y = rep my;  
random trat bloco(rep)/solution;  
run;
```

```
proc mixed cl;  
title 'Latice – Vizinhança – PPD – trata aleatorio';  
class rep bloco trat;  
model y = rep mrLAT;  
random trat bloco(rep)/solution;  
run;
```

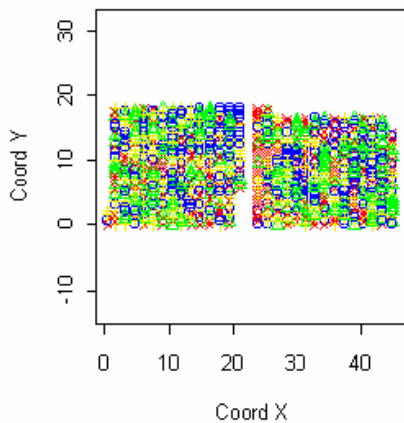
```
proc mixed cl;  
title 'Espacial - Latice';  
class rep bloco trat;  
model y = rep;  
random trat bloco(rep)/solution;  
repeated / type=sp(sph)(c l);  
run;
```

```
quit;  
*/
```

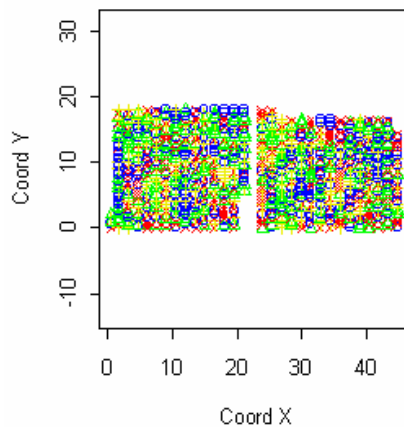


**ANEXO 2B** Gráficos de dispersão das observações e dos resíduos dos três maiores experimentos (1, 2 e 3) no campo experimental, com Coord Y representando as Linhas (m) e Coord X representando as Colunas (m)

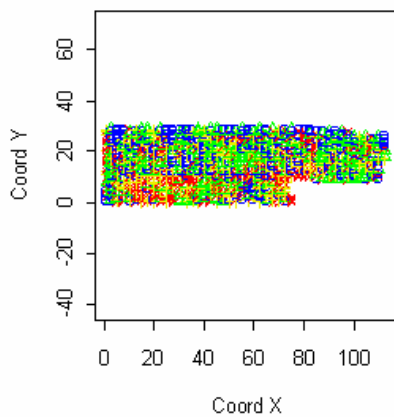
Experimento 1 - Observações



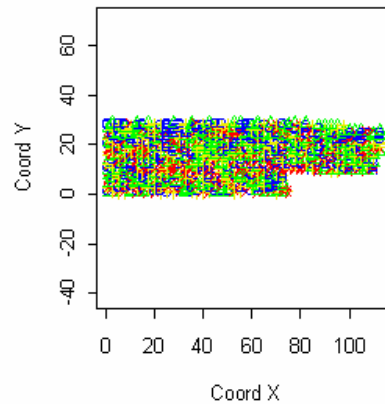
Experimento 1 - Resíduos



Experimento 2 - Observações



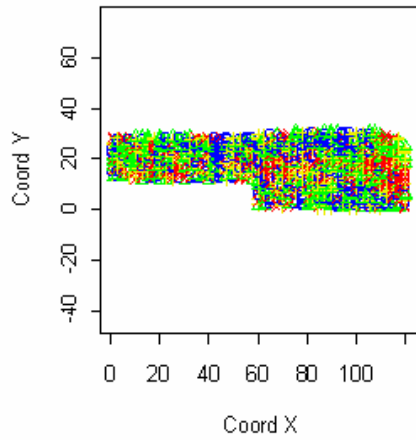
Experimento 2 - Resíduos



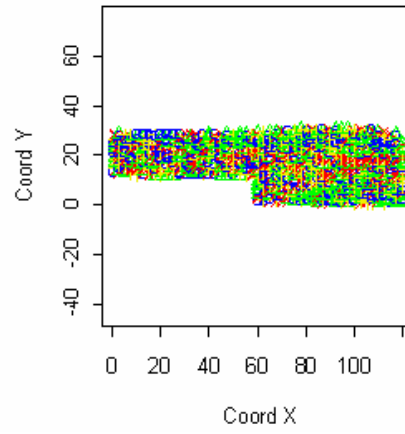
...continua...

“ANEXO 2B, Cont.”

Experimento 3 - Observações



Experimento 3 - Resíduos



- X: Parcelas com maiores produções (g/parcela),
- Y: Parcelas com segunda maiores produções (g/parcela),
- Z: Parcelas com terceira maiores produções (g/parcela),
- X: Parcelas com menores produções (g/parcela).