

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

Modelos lineares mistos para explicar a variabilidade espacial  
na análise conjunta de experimentos agronômicos

Cássio Dessotti

Tese apresentada para obtenção do título de  
Doutor em Ciências. Área de concentração: Es-  
tatística e Experimentação Agronômica

Piracicaba  
2014

Cássio Dessotti  
Licenciado em Matemática

**Modelos lineares mistos para explicar a variabilidade espacial  
na análise conjunta de experimentos agronômicos**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:  
Prof<sup>ª</sup> Dr<sup>ª</sup> **SÔNIA MARIA DE STEFANO PIEDADE**

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

**Piracicaba  
2014**

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

Dessotti, Cássio

Modelos lineares mistos para explicar a variabilidade espacial na análise conjunta de experimentos agronômicos / Cássio Dessotti. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.

101 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Variabilidade espacial 2. Modelos lineares mistos 3. Análise de variância  
4. Análise de grupos de experimentos 5. Produção de cana-de-açúcar I. Título

CDD 633.61  
D475m

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

## DEDICATÓRIA

*Aos meus pais,  
**Laerte Dessotti e  
Eliamar Ventura Ribeiro Dessotti***

*pelo amor, dedicação e carinho que me  
transmitem.*

*Com amor, DEDICO.*



## AGRADECIMENTOS

Primeiramente a Deus, força maior do universo, pela saúde, paciência e perseverança para realizar este trabalho.

À minha orientadora professora Dra. Sônia Maria De Stefano Piedade, pelo apoio, por sempre acreditar em mim, e pela amizade.

Aos professores doutores Decio Barbin, Julio César Pereira, Roseli Aparecida Leandro, César Gonçalves de Lima, Renata Alcarde, Carlos Tadeu dos Santos Dias, Clarice Garcia Borges Demetrio e Silvio Sandoval Zocchi.

Aos professores da UNESP de Ilha Solteira, Dr. Evaristo Bianchini (em memória) e Walter Veriano Valério Filho, pela indicação, motivação e grande amizade.

Às secretárias Solange de Assis Paes Sabadin e Luciane Brajão pela atenção. Aos técnicos em informática Eduardo e Jorge pelo suporte.

A todos meus colegas de pós-graduação, não apenas pelos momentos de dedicação e estudos, como também pelos ótimos momentos de risos.

Agradeço em especial aos amigos de pós graduação de meu departamento: Eduardo Gomes, Ana Julia, Thiago Gentil, Edilan Quaresma, Mauricio Lordelo, Ricardo Olinda, Pedro Cerqueira, Italo Frazão, Simone Grego, Simone Sartório, Simone Werner, José Nilton, Natalia Martins, Alessandra Santos, Lucas Cunha, Everton da Rocha, Marina Mestre, Luiz Ricardo, Rodrigo Pescim e Djair Durand.

Aos amigos conquistados ao longo dos anos acadêmicos: Paula Sanches, Elton Pereira, Darcio Silva, Mariana Urbano, Michele Barbosa, Elton Araújo, Gilberto Fernandes, Vitor Moretto, Wilton Souza e Flávio Galego.

Aos amigos de infância que perduram por todos estes anos: Douglas Maioli, Patricia Peramo, Ana Mayra, Raul Sippel, Emir Stringhetta e Johnny Andrade.

Aos amigos Thiago Abreu, Alexandre Cancilieri, Tiago Santana, Luis Angeloni e Luiz Trench por todo apoio nesta caminhada.

Aos estimados amigos Guilherme Biz e Ezequiel Bautista pela forte amizade a ser levada ao longo de toda vida.

À CAPES pelo fundamental suporte financeiro concedido.

Muito obrigado a todas as pessoas que contribuíram de alguma forma na minha formação e realização deste trabalho.

**Agradecimento especial**

Aos meus pais Eliamar Ventura Ribeiro Dessotti e Laerte Dessotti e meu querido irmão Fernando Augusto Dessotti, por todo o amor e carinho depositados em mim, por toda a oração direcionada a mim, e por tudo que representam em minha vida.

Aos meus avós Sinézia, Mario, Marcília e Antônio (em memória) e à minha tia Dinha, por todos os cuidados e ensinamentos em minha formação.

À minha amada namorada Jaqueline Augusto Sena, por me apoiar, acreditar em meu potencial, compreender minhas dificuldades e ausências, e estar sempre ao meu lado, independente de distâncias geográficas, nunca permitindo que me falte carinho.

*“As invenções são, sobretudo, o resultado de um trabalho teimoso.”*

(Santos Dumont).





## SUMÁRIO

RESUMO . . . . .	11
ABSTRACT . . . . .	13
LISTA DE FIGURAS . . . . .	15
LISTA DE TABELAS . . . . .	17
1 INTRODUÇÃO . . . . .	19
2 REVISÃO BIBLIOGRÁFICA . . . . .	23
2.1 Análise de grupos de experimentos . . . . .	23
2.2 Modelos lineares . . . . .	27
2.3 Modelos lineares mistos . . . . .	28
2.3.1 Especificação dos modelos lineares mistos . . . . .	29
2.3.2 Estruturas de covariância . . . . .	30
2.3.3 Seleção de modelos . . . . .	34
2.3.4 Estimação dos componentes de variância . . . . .	35
2.3.4.1 Método ANOVA . . . . .	36
2.3.4.2 Método da Máxima Verossimilhança (ML) . . . . .	37
2.3.4.3 Método da Máxima Verossimilhança Restrita (REML) . . . . .	38
2.3.5 Estimação dos efeitos fixos e predição dos efeitos aleatórios . . . . .	41
2.3.6 Inferência para parâmetros de efeitos fixos . . . . .	44
2.3.7 Inferência para parâmetros de efeitos aleatórios . . . . .	46
2.3.8 Diagnóstico de resíduos . . . . .	48
2.4 Variabilidade espacial em ensaios agrícolas . . . . .	49
2.5 Modelos geoestatísticos . . . . .	53
2.5.1 Modelo exponencial . . . . .	57
2.5.2 Modelo gaussiano . . . . .	57
2.5.3 Modelo esférico . . . . .	57
3 MATERIAL E MÉTODOS . . . . .	59
3.1 Material . . . . .	59
3.1.1 Grupo de experimentos de cana-de-açúcar . . . . .	59
3.1.2 Grupo de experimentos simulados . . . . .	63
3.2 Métodos . . . . .	63
3.2.1 Modelos mistos para o grupo de experimentos de cana-de-açúcar . . . . .	63

3.2.2 Modelos mistos para o grupo de experimentos simulados . . . . .	68
4 RESULTADOS E DISCUSSÃO . . . . .	71
4.1 Avaliação do grupo de experimentos de cana-de-açúcar . . . . .	71
4.2 Avaliação do grupo de experimentos simulados . . . . .	81
5 CONSIDERAÇÕES FINAIS . . . . .	83
REFERÊNCIAS . . . . .	85
ANEXOS . . . . .	93

## RESUMO

### Modelos lineares mistos para explicar a variabilidade espacial na análise conjunta de experimentos agrônômicos

O objetivo deste trabalho foi avaliar a incorporação de funções geostatísticas na matriz de variâncias e covariâncias residual no estudo de modelos lineares mistos a partir de um grupo de quatro experimentos de cana-de-açúcar, conduzidos na Guatemala nos seguintes locais: fazenda Limones - usina açucareira Pantaleón (LP), fazenda Bálsamo - usina açucareira Pantaleón (BP), área 1 da fazenda Limones - usina Madre Tierra (MT1) e área 2 da fazenda Limones - usina Madre Tierra (MT2). A variável resposta de interesse foi a produção de cana-de-açúcar por hectare, o delineamento utilizado nos quatro locais foi o casualizado em blocos, com cinco repetições e os mesmos seis tratamentos referentes a diferentes dosagens de um biorregulador (estimulante de crescimento). Em princípio, foram ajustados e comparados diversos modelos alternando-se o efeito de blocos, ora considerado fixo, ora aleatório, e a estrutura da matriz de variâncias e covariâncias ( $\mathbf{R}$ ), segundo os modelos exponencial, gaussiano e esférico. Estes modelos foram comparados, e os que admitem estruturas de dependência espacial se destacaram estatisticamente como os melhores, a partir do critério de Akaike (AIC), sendo então selecionados os modelos BFExp (blocos de efeito fixo e função exponencial na matriz  $\mathbf{R}$ ) e BAExpH (blocos de efeito aleatório, função exponencial para  $\mathbf{R}$  e variâncias diferentes entre os locais). A seguir, foi realizada a estimação dos efeitos fixos e a predição dos efeitos aleatórios por meio do método da máxima verossimilhança restrita (REML) pois esta metodologia proporciona um menor viés para suas estimativas. As análises conjuntas nos dois modelos selecionados não apresentaram interação tratamentos versus locais, nem mesmo efeito de tratamentos significativos, não sendo aconselhado o desdobramento desta interação. O efeito de locais por sua vez, foi significativo apenas no modelo BAExpH, e detectou-se neste caso a superioridade do local BP em relação aos demais. Adicionalmente, os locais foram analisados individualmente, focando a comparação dos modelos e as análises de variâncias, contudo, assim como na análise conjunta, nos modelos escolhidos para cada local, os efeitos de tratamentos também não foram significativos. Gráficos de resíduos foram construídos e representaram bons ajustes para os modelos BFExp e BAExpH para descrever os dados deste grupo de experimentos. Por fim, foi realizado um estudo de simulação cujos resultados deram mais credibilidade e suporte para a importância e relevância de se verificar, por meio de comparações, a necessidade de uso de um modelo mais elaborado, que considere a possível existência de dependência espacial entre as observações.

Palavras-chave: Variabilidade espacial; Modelos lineares mistos; Análise de variância; Análise de grupos de experimentos; Produção de cana-de-açúcar



## ABSTRACT

### Linear mixed models to explain the spatial variability in joint analysis from agronomical essays

The aim of this research was to evaluate the incorporation of geostatistical functions in the residual variances and covariances matrix in linear mixed models in a group of four experiments cane sugar conducted in four sites of Guatemala: farm Limones - Pantaleon sugar mill (LP), farm Bálsamo - Pantaleon sugar mill (BP), area 1 of the farm Limones - sugar mill Madre Tierra (MT1) and area 2 of the farm Limones - sugar mill Madre Tierra (MT2). Production of sugar cane was the interest variable analyzed at all locations, using the randomized block design with five replications and the same six treatments related to different doses of a plant growth regulator. Initially the models were adjusted and compared with alternating the blocks effect, sometimes considered fixed, sometimes random, and the structure of the variance and covariance matrix ( $\mathbf{R}$ ) according to the exponential, gaussian and spherical models. The models were compared, and, among them, those with spatial dependence structures stood out as the best statistically from the Akaike information criterion (AIC), and the selected modelos were the BFExp model (block as fixed effect and exponential function to  $\mathbf{R}$ ) and the BAExpH model (block as random effect, exponential function to  $\mathbf{R}$  and different variances among the sites). After that, the estimation of fixed effects and prediction of random effects using the restricted maximum likelihood method (REML) were done, since this methodology provides a lower bias to their estimates. The joint analysis of both selected models showed no interaction between treatments and locals, even significant effect of treatments, not being advised the unfolding of this interaction. The effect of local was significant only in the BAExpH model, and detected in this case the superiority of the local BP in relation to the others. Additionally, individual sites were examined similarly to the previous case, through comparison of models and analysis of variance, however, treatment effects weren't significant too. Residual plots were constructed and represented satisfactory fit of the models to describe the data in all cases studied. Finally, a simulation study showed results with more credibility and support for the importance and relevance of verifying, through comparisons, the need to use a more structured model that considers the possible existence of spatial dependence among observations.

Keywords: Spatial variability; Linear mixed models; Analysis of variance;  
Joint analysis from agronomical essays; Sugar cane production



## LISTA DE FIGURAS

Figura 1 - Semivariogramas (a) Teórico e (b) Empírico com características de dependência espacial . . . . .	54
Figura 2 - Modelos teóricos (a) exponencial, (b) gaussiano e (c) esférico . . . . .	58
Figura 3 - Croqui de instalação do experimento com o georreferenciamento das parcelas para cada local . . . . .	61
Figura 4 - Produção média de cana-de-açúcar, estimada por tratamento (trat) e por local, segundo o modelo BFExp . . . . .	74
Figura 5 - Produção média de cana-de-açúcar, estimada por tratamento (trat) e por local, segundo o modelo BAExpH . . . . .	74
Figura 6 - Gráficos dos resíduos condicionais estudentizados: (a) em função dos valores preditos, (b) histograma e (c) quantil-quantil para o modelo BFExp	76
Figura 7 - Gráficos dos resíduos condicionais estudentizados: (a) em função dos valores preditos, (b) histograma e (c) quantil-quantil para o modelo BAExpH	76
Figura 8 - Gráficos dos semivariogramas empíricos dos resíduos condicionais estudentizados do modelo BFExp para os modelos ajustados para cada local: (a) gaussiano no local LP, (b) exponencial no local BP, (c) exponencial no local MT1 e (d) gaussiano no local MT2 . . . . .	78
Figura 9 - Gráficos dos semivariogramas empíricos dos resíduos condicionais estudentizados do modelo BAExpH para os modelos ajustados para cada local: (a) gaussiano no local LP, (b) exponencial no local BP, (c) exponencial no local MT1 e (d) gaussiano no local MT2 . . . . .	78





## LISTA DE TABELAS

Tabela 1 - Tratamentos estimulantes (Trat.) para cana-de-açúcar baseados em ácido giberélico ( $AG_3$ ) . . . . .	60
Tabela 2 - Estrutura de dados balanceados para cada um dos 4 experimentos do grupo em estudo (locais: (1) LP, (2) BP, (3) MT1 e (4) MT2) sob o delineamento em blocos ao acaso, com 5 repetições, em que são avaliados os 6 tratamentos estimulantes de crescimento de plantas de cana-de-açúcar	61
Tabela 3 - Resumo das primeiras linhas do banco de dados, detalhadas segundo o local, tratamento (Trat), bloco e produção (Prod), além das coordenadas de latitude (Lat) e longitude (Long) referentes a cada parcela . . . . .	62
Tabela 4 - Médias e desvios padrões (d.p.) da produção de cana-de-açúcar (ton/ha) nos quatro locais . . . . .	71
Tabela 5 - Valores obtidos por meio do critério de Akaike (AIC) para os modelos com blocos de efeito fixo, ajustados sem e com estruturas de correlação espacial . . . . .	72
Tabela 6 - Valores obtidos por meio do critério de Akaike (AIC) para os modelos com blocos de efeito aleatório, ajustados sem e com estruturas de correlação espacial . . . . .	72
Tabela 7 - Graus de Liberdade (GL), Estatística $F$ e respectivos níveis descritivos (valores- $p$ ) associados às fontes de variação de efeito fixo do modelo selecionado com blocos de efeito fixo (M03 - BFExp) . . . . .	73
Tabela 8 - Graus de Liberdade (GL), Estatística $F$ e respectivos níveis descritivos (valores- $p$ ) associados às fontes de variação de efeito fixo do modelo selecionado com blocos de efeito aleatório (M12 - BAExpH) . . . . .	73
Tabela 9 - Resumo do teste Tukey-Kramer para o modelo BAExpH com agrupamento dos locais segundo a produção média (Prod) de cana-de-açúcar ( $tm\ h^{-1}$ ) . . . . .	75
Tabela 10 - Estimativas de componentes de variância ( $\sigma^2$ ) e do parâmetro espacial ( $\rho$ ) do modelo BFExp . . . . .	77
Tabela 11 - Estimativas de componentes de variância ( $\sigma^2$ ) e do parâmetro espacial ( $\rho$ ) do modelo BAExpH . . . . .	77

Tabela 12 -Valores obtidos por meio do critério de Akaike (AIC) para os quatro modelos com efeito fixo para blocos, sem e com estruturas de correlação espacial, para as análises individuais em cada um dos quatro locais . . .	79
Tabela 13 -Valores obtidos por meio do critério de Akaike (AIC) para os quatro modelos com efeito aleatório para blocos, sem e com estruturas de correlação espacial, para as análises individuais em cada um dos quatro locais . . . . .	80
Tabela 14 -Estatística $F$ e significância para tratamentos nos modelos selecionados a partir de AIC, dentre os que consideraram o efeito de blocos como fixo (BF) e os que o consideraram aleatório (BA) . . . . .	80
Tabela 15 -Estudo dos 500 grupos de experimentos gerados para cada cenário, com resíduos obtidos pelas funções exponencial ( $\mathbf{R}_{exp}$ ), gaussiana ( $\mathbf{R}_{gau}$ ) e esférica ( $\mathbf{R}_{esf}$ ), na verificação da quantidade e porcentagem de modelos mais adequados sem (M1-M2) versus com (M3-M8) estrutura de dependência espacial, para os três IDE . . . . .	81
Tabela 16 -Esboço dos três modelos mais adequados em nove cenários, a partir da quantidade e porcentagem de vezes que cada um deles foi selecionado, no estudo de 500 grupos de experimentos gerados para as três configurações de ( $\mathbf{R}$ ) com resíduos obtidos por meio das funções exponencial ( $\mathbf{R}_{exp}$ ), gaussiana ( $\mathbf{R}_{gau}$ ) e esférica ( $\mathbf{R}_{esf}$ ), para os três IDE (10%, 50% e 90%) .	82

## 1 INTRODUÇÃO

A cana-de-açúcar é uma planta pertencente ao gênero *Saccharum*. As espécies de cana-de-açúcar são provenientes do sudeste asiático, e foram introduzidas nas Américas, onde são muito cultivadas atualmente. Destacam-se, entre os derivados da cana-de-açúcar, o próprio açúcar, o álcool combustível e mais recentemente, o biodiesel. Uma tonelada de cana-de-açúcar produz, em média, 80 litros de etanol sendo que um hectare de terra produz 88 toneladas de cana-de-açúcar, e assim, no total são produzidos 7040 litros de etanol por hectare (EMBRAPA, 2013).

O Brasil se destaca como o maior produtor de cana-de-açúcar do mundo há mais de 30 anos. O país produziu aproximadamente 600 milhões de toneladas desta cultura na safra 2012/2013. Atualmente, a produtividade média de cana-de-açúcar no Brasil, é de 80 toneladas por hectare, valor muito abaixo do potencial biológico devido às interferências de vários fatores, como clima, solo, pragas, entre outros (FAOSTAT, 2014).

A Guatemala vem aumentando sua produção de cana-de-açúcar nas últimas décadas, sendo que de 1981 para 2011, o país passou de 19º para 13º maior produtor do mundo (ASAZGUA, 2013). A área destinada à produção desta cultura na safra 2012/2013 no país, foi de 263 mil hectares (ha), sendo produzidas 26,7 milhões de toneladas métricas (tm) de cana, sendo que 1 tm corresponde a 1000 quilogramas (CENGICAÑA, 2013).

Atualmente com o avanço das técnicas de cultivo agrícola, é crescente a preocupação em reduzir o impacto ambiental sem prejudicar o aumento de produtividade. Assim, torna-se fundamental a utilização de novas tecnologias, que possibilitem alcançar tais benefícios. Neste sentido, destaca-se o uso dos biorreguladores: compostos responsáveis por tornar as plantas mais eficientes e adaptadas a explorar melhor o ambiente, que atuam nos vegetais minimizando o efeito de estresse e contribuindo para o desenvolvimento do sistema radicular e da parte aérea, levando a planta a novos patamares de produtividade. (STOLLER, 2014).

A alta produtividade é resultado de uma equação complexa que engloba variedades com alto potencial genético, e um ambiente favorável para as plantas produzirem. Além disso, muitas vezes e devido a inúmeros fatores, o ambiente apresenta características que facilitam a existência de dependência espacial entre as amostras estudadas. Esta dependência, embora na maioria das vezes desconhecida ou desprezada, reflete nos valores amostrais afetando diretamente os resultados provenientes das análises estatísticas.

A análise clássica de experimentos de campo assume que todas as observações, sejam elas tomadas em posições adjacentes ou não, são não correlacionadas. Assim, a matriz de variâncias e covariâncias residual é geralmente considerada como uma matriz diagonal, ou seja, os erros são admitidos independentes. Desta forma, a posição das parcelas que recebem os tratamentos no campo, ou seja, a distribuição espacial dos mesmos é ignorada. Entretanto, a dependência espacial pode existir e contribuir para o aumento da variação residual, sendo assim importante considerá-la nos modelos lineares mistos (MLM), por meio da modelagem da matriz de variâncias e covariâncias residual ( $\mathbf{R}$ ).

Visando controlar uma possível correlação espacial, a casualização é efetuada com o propósito de se realizar uma análise de variância confiável. No entanto, o controle de tal correlação é mais eficiente quando são utilizados modelos espaciais. Até mesmo formas de controle como a instalação de blocos, por exemplo, podem ser ineficazes para tratar problemas de gradientes ambientais, conforme alerta Resende (2007).

A instalação e delimitação dos blocos costumam ser realizadas a priori, desde o planejamento do experimento, de forma que se percebe e se controla muitas vezes a presença de manchas ou outros gradientes ambientais dentro do experimento. Porém, tais adversidades causadoras de dependência espacial, nem sempre são controladas pelo pesquisador devido a diversos fatores, já que é extremamente difícil o conhecimento das características e do histórico da área em estudo.

Nestes casos, apenas técnicas de análise espacial permitem contornar a questão e propiciar tomadas de decisões a respeito de modelos ou tratamentos mais adequados, a partir da construção de blocos a posteriori, ou por meio de uma escolha conveniente da matriz de variâncias e covariâncias residual, baseando-se nos dados experimentais, conforme realizado por Duarte (2000).

É evidente que a variabilidade ou heterogeneidade espacial associada à fertilidade e estrutura do solo, umidade, interceptação de luz e diversos outros possíveis fatores ambientais, contribuem de forma significativa, para o aumento da variação residual. Desta forma, é extremamente importante controlar, seja por meio de delineamento experimental adequado ou a partir de modelagem estatística, a variação residual espacial (tendência) em análises de dados oriundos de agricultura.

Adicionalmente a isto, o fator de variação correspondente a blocos, constantemente considerado de efeito fixo, pode ser considerado de efeito aleatório, integrando a matriz de variâncias e covariâncias da parte aleatória do modelo linear misto (MLM), já que desde que não se tenha conhecimento de algum fator claro de variação a ser controlado a priori, os blocos dentro de um experimento agrônômico, costumam ser uma amostra de subáreas (parcelas) de toda a área em estudo.

Uma análise alternativa foi proposta neste trabalho, no contexto dos modelos lineares mistos, porém relevando a dependência espacial residual que muitas vezes está presente nas parcelas em estudo. Tal abordagem foi realizada para avaliar um grupo de quatro experimentos no delineamento casualizado em blocos, modelando de diversas formas, a matriz de variâncias e covariâncias a partir de funções geoestatísticas, visando a melhor maneira de se comparar seis tratamentos de um biorregulador estimulante aplicado em plantas (produção) de cana-de-açúcar.

Os objetivos deste trabalho foram: (i) construir modelos clássicos para os dados de cana-de-açúcar em questão, mas principalmente, modelos mistos sob diferentes estruturas de matriz de variâncias e covariâncias, modelando a dependência espacial residual a partir de três funções geoestatísticas, visando por fim, uma comparação de tratamentos embasada no modelo mais adequado e (ii) simular conjuntos de dados de grupos de experimentos, visando a comparação de MLM com diferentes funções geoestatísticas na matriz  $\mathbf{R}$ , avaliando possíveis vantagens nesta abordagem em relação à análise clássica.

Na seção 2 é realizada a revisão de literatura abordando os temas levantados nesta introdução e essenciais para as análises estatísticas. Na seção 3 são descritos os dados utilizados neste trabalho e a metodologia a ser adotada para analisá-los, além da metodologia utilizada na simulação de conjuntos de dados. Na seção 4 são apresentados e discutidos os resultados das análises dos dados reais e simulados e na seção 5 são exibidas as considerações finais deste trabalho.



## 2 REVISÃO BIBLIOGRÁFICA

Nesta seção é apresentada uma breve revisão de análise de grupos de experimentos, incluindo contexto histórico, características e aplicações, além da descrição da temática dos modelos lineares mistos (MLM), com destaque para o processo de estimação e predição dos efeitos envolvidos. É detalhada ainda, a teoria utilizada para construção e comparação de modelos com diferentes estruturas para as matrizes  $\mathbf{G}$  e  $\mathbf{R}$  de variâncias e covariâncias. Além disso, são destacados nesta seção, alguns trabalhos que contribuíram com o estudo da variabilidade espacial em ensaios agrícolas e as estruturas de diferentes funções geoestatísticas para explicar tal variabilidade.

### 2.1 Análise de grupos de experimentos

A análise de grupos de experimentos, resume-se essencialmente na instalação de um mesmo experimento em locais diferentes, realizando-se uma análise conjunta, que pode permitir a tomada de conclusões abrangentes a todos os locais ou indicar que se estude o efeito dos tratamentos separadamente em cada ambiente, realizando-se neste caso um desdobramento da interação significativa tratamentos versus locais.

As análises de grupos de experimentos, objeto de grande interesse, em especial, para os melhoristas genéticos, foram estudadas por diversos autores, destacando-se Rojas (1951), Kempthorne (1952), Cochran e Cox (1957), Campos (1984), Pimentel-Gomes (2009), Barbin (2013), dentre outros.

Dentre os pioneiros neste tipo de abordagem, Rojas (1951) desenvolveu uma análise de variância para grupos de experimentos em estudos de melhoramento de plantas.

Cochran e Cox (1957) também são responsáveis por grande contribuição nesta área. Tais autores afirmam ser comum o uso da análise de grupos de experimentos, mesmo quando não se tem interesse em fazer inferência sobre uma população específica, e sim em estudar a influência de fatores externos sobre as respostas de algum experimento, ou seja, esperando que os tratamentos apresentem o mesmo comportamento ao longo de toda a área de estudo. Pode-se investigar por exemplo, a possível interferência de calor ou umidade sobre determinados tratamentos ou genótipos.

Pimentel-Gomes e Guimarães (1958) também se destacam em um dos primeiros trabalhos com grupos de experimentos, a partir de um estudo referente à análise conjunta de experimentos em blocos completos casualizados com tratamentos comuns.



Pimentel-Gomes (1970) apresenta a análise de um grupo de experimentos em blocos casualizados com alguns tratamentos comuns, em que o número de repetições para tratamentos varia de um experimento para outro. Porém, o esquema de análise de variância e a obtenção das somas de quadrados já haviam sido desenvolvidos por Pimentel-Gomes e Guimarães (1958).

Agrupando-se os experimentos de interesse e visando facilitar a análise conjunta, Barbin (2013) orienta que se utilize o mesmo delineamento, os mesmos tratamentos, e, de preferência, que em todos os locais em estudo se tenha o mesmo número de repetições. Tais cuidados facilitam consideravelmente a análise conjunta.

Campos (1984) destaca que, para o agrupamento de locais similares, alguns critérios devem ser considerados, como: setor geográfico, ano agrícola, afinidade quanto à alguma característica de interesse, grandeza dos quadrados médios dos resíduos das análises individuais, dentre outros.

Banzatto e Kronka (2006) alertam que os ensaios individuais não precisam ter obrigatoriamente um número muito grande de parcelas. Ressaltam ainda que as análises conjuntas podem ser realizadas com os dados originais, os totais de tratamentos ou ainda com as médias de cada tratamento em cada ensaio.

Vencovsky e Barriga (1992) apontam que uma característica que costuma complicar esse tipo de análise é o caso de se dispor de ensaios com número diferente de blocos ou repetições. Tal complicação foi contornada por Cochran e Cox (1957) a partir de processos biométricos.

Segundo Cecon (1992), a influência que o ambiente exerce sobre o genótipo, propiciando a ocorrência de interação genótipo versus ambiente, é um dos maiores problemas que os melhoristas enfrentam na seleção de genótipos, pois a variância genética é afetada pelo ambiente específico em que é realizada a seleção, e o material melhorado geralmente é distribuído para ambientes diversos.

Este autor descreve ainda, que em grupos de experimentos, inicia-se fazendo a análise e interpretação dos dados de cada ensaio, para depois verificar se as diferenças entre tratamentos são consistentes em todos os locais, isto é, se elas variam da mesma forma de local para local, para finalmente realizar a estimação e comparação dos efeitos médios de tratamentos sobre o conjunto de experimentos.

Regazzi et al. (1999) afirmam que o processo tradicional de investigação das interações genótipo versus ambiente parte de uma análise de variância conjunta, e ainda, por meio desta, a magnitude de cada uma das interações é investigada baseando-se na variância dos efeitos de genótipos versus locais, versus anos, versus anos versus locais, e outros segundo o interesse do melhorista.

Pimentel-Gomes (2009) orienta que se o quociente entre o maior e o menor dos quadrados médios residuais (QMRes) for inferior a 7, a análise conjunta poderá ser realizada sem grandes problemas. Barbin (2013) reforça tal ideia, afirmando que nos casos em que esta homogeneidade não é satisfeita, pode-se eliminar o experimento cujo QMRes seja discrepante, ou ainda, agrupar os experimentos que possuam QMRes similares.

Outras alternativas para detectar o problema da falta de homogeneidade, tratam-se dos testes de Bartlett, descrito por Anderson e Bancroft (1952), além do teste  $F$  máximo, também conhecido por teste de Hartley (1950).

Nos casos em que a homogeneidade de variâncias não é satisfeita, pode-se agrupar todos os locais em uma única análise conjunta, ajustar o número de graus de liberdade do resíduo médio ( $v$ ) e o número de graus de liberdade da interação tratamentos versus locais ( $v'$ ), a partir da equação de Satterthwaite (1946), segundo Cochran (1954):

$$v = \frac{\left[ \sum_{k=1}^K \text{QMRes (Local } k) \right]^2}{\sum_{k=1}^K \frac{[\text{QMRes (Local } k)]^2}{\text{GLRes (Local } k)}}$$

com  $\text{GLRes (Local } k) < v < \sum_{k=1}^K \text{GLRes (Local } k)$ , e

$$v' = \frac{(I - 1)(K - 1)^2 V_1^2}{(K - 2)V_2 + V_1^2},$$

em que

$$V_1 = \frac{\sum_{k=1}^K \text{QMRes (Local } k)}{K}, \quad V_2 = \frac{\sum_{k=1}^K [\text{QMRes (Local } k)]^2}{K},$$

sendo  $I$  o número de tratamentos,  $K$  o número de locais, e ainda,  $\text{GLRes(Local } k)$  e  $\text{QMRes(Local } k)$ , o número de graus de liberdade e o quadrado médio residual referente ao  $k$ -ésimo local (experimento).

No entanto, com o passar dos anos, esta preocupação de Pimentel-Gomes (2009) e Barbin (2013) com os QMRes discrepantes entre locais, passou a ser contornada a partir da abordagem dos modelos lineares mistos (MLM), com a utilização de diferentes variâncias residuais para cada local, por meio da matriz de variâncias e covariâncias.

Os modelos lineares mistos (MLM) surgem então como mais uma alternativa para a análise de dados agronômicos, possuindo como grande vantagem, a possibilidade de modelagem da matriz de variâncias e covariâncias. Isso pode ser realizado com o uso de funções geoestatísticas, levando em conta uma possível correlação espacial presente nos dados. Esta abordagem tem ganhado destaque em diversos trabalhos como Jaramillo e Francisco (2005), Littell et al. (2006), Casanoves, Machiavelli e Balzarini (2007), Di Rienzo, Machiavelli e Casanoves (2011), entre outros.

Littell et al. (2006) afirmam que variâncias heterogêneas podem ser incorporadas na análise estatística dos dados, no contexto dos MLM, especificando diferentes estruturas de variâncias e covariâncias residuais para os níveis de um fator (ou uma combinação dos níveis dos fatores avaliados), por exemplo, especificando variâncias diferentes nos diferentes locais, dando menor peso às observações com grande variância.

Casanoves, Machiavelli e Balzarini (2007), a respeito do uso de MLM para grupos de experimentos de amendoim, estudados ao longo dos anos, realizaram comparações de modelos que consideram os locais com mesma variância residual versus modelos que admitem os locais com diferentes variâncias residuais. Mais recentemente, Di Rienzo, Machiavelli e Casanoves (2011) destacam que a heterogeneidade residual pode ser modelada pela matriz de variâncias e covariâncias residual no contexto dos MLM.

Superado o problema de heterogeneidade de variâncias para locais, a análise conjunta é construída e as atenções se voltam para a interação tratamentos versus locais. Sendo esta interação significativa, não se pode tirar conclusões gerais a toda área em estudo, o comportamento dos tratamentos difere de local para local, e desta forma, pode-se simplesmente aceitar os resultados das análises individuais ou, preferivelmente, desdobrar os graus de liberdade da interação, estudando-se os tratamentos dentro de cada local. Esta segunda opção tem como vantagem o uso de um único QMRes com um maior número de graus de liberdade, resultando em uma análise mais sensível que tende a apontar mais diferenças significativas entre tratamentos.

## 2.2 Modelos lineares

Consideremos como construção inicial, um estudo agrícola acerca da avaliação de  $I$  tratamentos ( $t_1, t_2, \dots, t_I$ ), para uma determinada cultura, avaliados em  $K$  locais ( $l_1, l_2, \dots, l_K$ ), a partir do delineamento casualizado em blocos, com  $J$  blocos por local ( $b_1, b_2, \dots, b_J$  repetições), na abordagem de uma certa variável resposta predefinida. Admitiremos ainda como fixos: a média, os efeitos de tratamentos, de locais e de blocos, totalizando  $p = I + J + K + 1$  parâmetros, resultando em um número de observações igual a  $n$ . O modelo linear clássico utilizado para a análise de dados desta natureza, é definido por Searle (1971), como segue:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1)$$

em que,  $\mathbf{y}$  representa o vetor de dimensões  $n \times 1$ , de dados observados,  $\mathbf{X}$ , de dimensões  $n \times p$ , é a matriz de delineamento;  $\boldsymbol{\beta}$ , de dimensões  $p \times 1$ , é um vetor de parâmetros desconhecidos de efeitos fixos e  $\boldsymbol{\epsilon}$  é o vetor de dimensões  $n \times 1$ , de erros aleatórios.

O objetivo do modelo linear clássico é modelar (descrever) a média de  $\mathbf{y}$ , a partir do vetor de parâmetros de efeitos fixos  $\boldsymbol{\beta}$ . Os componentes do vetor  $\boldsymbol{\epsilon}$  são variáveis aleatórias independentes e identicamente distribuídas com média 0 e variância  $\sigma^2$ .

Pressupondo que  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , tem-se que  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ , cuja função de verossimilhança é dada por:

$$L = L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = \frac{\exp \left[ -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \left( \frac{1}{\sigma^2} \mathbf{I} \right) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]}{(2\pi\sigma^2)^{\frac{n}{2}}}.$$

As estimativas de máxima verossimilhança dos parâmetros são obtidas resolvendo-se o sistema de equações normais:  $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ , cuja solução é dada por  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , desde que  $\mathbf{X}'\mathbf{X}$  seja não singular. Além disso, tem-se que  $Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ .

Caso  $(\mathbf{X}'\mathbf{X})^{-1}$  não exista, utiliza-se uma inversa generalizada do tipo  $(\mathbf{X}'\mathbf{X})^-$ , e assim, as estimativas dos parâmetros são dadas por:  $\hat{\boldsymbol{\beta}}^0 = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y}$  e  $Var(\hat{\boldsymbol{\beta}}^0) = (\mathbf{X}'\mathbf{X})^- (\mathbf{X}'\mathbf{X}) (\mathbf{X}'\mathbf{X})^- \sigma^2$ .

Os modelos lineares (clássicos) ao considerar  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , admitem independência residual, e desta forma, não consideram na modelagem, qualquer influência espacial que possa existir na área em estudo.

### 2.3 Modelos lineares mistos

Na estatística experimental, é muito comum buscar modelos capazes de explicar determinados fenômenos da forma mais precisa possível. É usual adotar modelos de fatores fixos, quando o pesquisador visa a estimação e realização de testes de hipóteses sobre as médias dos níveis de tais efeitos. Porém, muitas vezes o interesse está nos modelos aleatórios, especialmente quando se busca a predição de tais efeitos, e a estimação dos componentes de variância para cada efeito aleatório envolvido. Desta forma, visando unir estas duas ideias em um único modelo, surge o modelo linear misto.

A teoria dos modelos mistos foi proposta por Henderson (1949) para ser utilizada na avaliação genética de bovinos, e foi apresentada pela primeira vez em 1973, passando a ser utilizada na prática a partir da década de 80, devido aos avanços tecnológicos computacionais que facilitaram seu uso, conforme descrito por Resende (2007).

West, Welch e Gajęcki (2007) definem um modelo linear misto (MLM) como sendo paramétrico para dados agrupados, longitudinais ou provenientes de experimentos com medidas repetidas, que inclui parâmetros de efeito fixo, além da média geral e efeitos aleatórios associados com um ou mais fatores aleatórios, além do erro experimental.

Desta forma, esses modelos envolvem duas partes, uma parte descrevendo os efeitos aleatórios e a outra, os efeitos fixos. De forma geral, Pinheiro e Bates (2000) afirmam que os efeitos de um fator são considerados como fixos quando os níveis em estudo, geralmente poucos, forem escolhidos pelo pesquisador, de modo que a inferência é restrita a esses níveis.

Barbin (1993) segue a mesma linha, afirmando que os efeitos de um fator são considerados aleatórios, quando os níveis em estudo correspondem a uma amostra aleatória de uma população de referência, desta forma os níveis provêm de uma distribuição de probabilidade e a inferência é extrapolada para a população de referência, tratando-se de fatores com níveis que não podem ser controlados pelo pesquisador.

Porém, segundo Searle, Casella e McCulloch (1992), a decisão acerca de admitir certo efeito como fixo ou aleatório não é tão trivial quanto parece. Exemplificando, o fator blocos pode ser definido como de efeito aleatório se admitirmos que os blocos têm, em determinado experimento, a capacidade de representar todo o local em estudo; mas pode também ser considerado fixo, se o pesquisador acredita, desde o início, que os blocos diferem, ou seja, que existe um gradiente de variação entre os blocos, como manchas no

solo, entre outros. Este tipo de raciocínio pode ajudar a decidir por uma fonte de variação de efeito fixo ou aleatório a ser anexada ao modelo, porém vale ressaltar que tal decisão nem sempre é unânime.

Nos MLM, a análise da parte aleatória consiste na predição dos efeitos aleatórios na presença dos efeitos fixos, e na estimação dos componentes de variância, que são variâncias associadas aos efeitos aleatórios.

Além disso, a análise da parte fixa do MLM consiste na estimação e testes de hipóteses sobre funções estimáveis dos efeitos fixos. Em geral, tanto a predição dos efeitos aleatórios quanto a estimação dos efeitos fixos, dependem da estimação dos componentes de variância.

### 2.3.1 Especificação dos modelos lineares mistos

A seguir, será apresentada uma abordagem geral de estimação de efeitos fixos e predição de efeitos aleatórios, levando em consideração a incorporação de dependência espacial.

Segundo Searle, Casella e McCulloch (1992), um modelo linear misto (MLM) geral, em termos matriciais, é definido como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (2)$$

sendo:  $\mathbf{y}$ , o vetor de observações, de dimensões  $n \times 1$ ;  $\mathbf{X}$ , a matriz de delineamento dos efeitos fixos, conhecida, de dimensões  $n \times (p + 1)$ ;  $\boldsymbol{\beta}$ , o vetor de parâmetros de efeitos fixos desconhecidos, de dimensão  $(p + 1) \times 1$ ;  $\mathbf{Z} = [\mathbf{Z}_1, \dots, \mathbf{Z}_b]$ , sendo  $\mathbf{Z}_i$  de dimensões  $n \times q_i$  a matriz de delineamento para o  $i$ -ésimo efeito aleatório;  $\mathbf{u} = [\mathbf{u}_1^T, \dots, \mathbf{u}_b^T]^T$ , o vetor de parâmetros de efeitos aleatórios, desconhecido, de dimensão  $q \times 1$ , em que  $\mathbf{u}_i$  possui dimensão  $q_i$ , sendo  $q = \sum_{i=1}^b q_i$ ; e  $\boldsymbol{\epsilon}$  o vetor de erros aleatórios desconhecidos, de dimensões  $n \times 1$ , com  $E(\mathbf{u}) = \mathbf{0}$  e  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ .

Adicionalmente, Gumedze e Dunne (2011) alertam que  $\mathbf{u}$  e  $\boldsymbol{\epsilon}$  seguem distribuições normais, independentes e multivariadas, tal que:

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{G}(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}(\boldsymbol{\rho}) \end{bmatrix} \right),$$

sendo que  $\boldsymbol{\gamma}$  e  $\boldsymbol{\rho}$  possuem, respectivamente, dimensões  $r \times 1$  e  $s \times 1$  (com  $s \leq n(n+1)/2$ ), e são vetores de parâmetros de variância desconhecidos correspondentes a  $\mathbf{u}$  e  $\boldsymbol{\epsilon}$ . Caso os termos aleatórios sejam correlacionados, então a dimensão de  $\boldsymbol{\gamma}$  pode exceder  $q$ , ou seja,  $\boldsymbol{\gamma}$  pode ser de dimensão  $r \leq q(q+1)/2$ .

Em termos de modelos condicionais (ou hierárquicos), o modelo linear misto geral pode ser reescrito de tal modo que:

$$E(\mathbf{Y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u},$$

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

e

$$Var(\mathbf{Y}|\mathbf{u}) = \mathbf{R}.$$

Assim, marginalmente, tem-se que  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ , em que  $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$

e

$$Var(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}, \quad (3)$$

sendo  $\mathbf{G}$  de dimensões  $q \times q$ , a matriz de variâncias e covariâncias dos efeitos aleatórios  $\mathbf{u}$ , e  $\mathbf{R}$  de dimensões  $n \times n$ , a matriz de variâncias e covariâncias residual, que representa a variação intra-grupos. Assim, pode-se modelar a estrutura de variâncias e covariâncias dos dados, de (3), especificando a caracterização de  $\mathbf{G}$  e  $\mathbf{R}$ . Quando  $\mathbf{R} = \sigma^2 \mathbf{I}_{(n)}$  e  $\mathbf{Z} = \mathbf{0}$ , o modelo misto se reduz ao modelo linear clássico:  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ .

### 2.3.2 Estruturas de covariância

Halimi (2005) afirma que a análise de dados experimentais, sob a abordagem de modelos lineares mistos, permite a incorporação de uma possível dependência espacial no modelo, possibilitando assim o estudo de estruturas mais abrangentes e realísticas.

Em modelos mistos, a modelagem da parte aleatória se realiza por meio da inclusão de uma matriz de variâncias-covariâncias (LITTELL et al., 2006), sendo que a necessidade de se adicionarem parâmetros de variâncias e covariâncias pode surgir por várias razões, dentre elas:

- i) As unidades experimentais podem ser alocadas em grupos e os dados de um grupo comum são correlacionados. Isso pode ocorrer com dados de famílias, ninhadas, colônias e pessoas que habitam a mesma casa.
- ii) Medidas repetidas são tomadas sobre a mesma unidade experimental e são correlacionadas. A natureza dessas medidas pode ser multivariada. Exemplos comuns são os dados observados ao longo do tempo, chamados dados longitudinais.

Para a escolha da matriz de variâncias e covariâncias mais adequada, Diggle et al. (2002), aconselham considerar:

- i) a variabilidade devida aos efeitos aleatórios, quando as unidades amostrais compõem uma amostra aleatória de uma população;
- ii) a variabilidade devida a correlação serial, pois espera-se que observações medidas entre curtos espaços de tempo sejam mais correlacionadas do que as medidas entre longos espaços de tempo;
- iii) a variabilidade devida aos erros de medição.

Em MLM costuma-se admitir, para a matriz de variâncias e covariâncias do resíduo,  $\mathbf{R} = \sigma^2 \mathbf{I}$ , que é ideal para modelar a situação em que a discrepância entre uma observação e o seu valor esperado é devida apenas a fatores do acaso, sendo cada medida independente de medidas feitas em outras ocasiões. Esta é a estrutura usada no modelo linear clássico, que pressupõe homocedasticidade das variâncias.

Porém, várias outras estruturas podem ser adotadas para a matriz de variâncias-covariâncias  $\mathbf{R}$ , dependendo do conhecimento que se tenha do fenômeno que gera os dados. Algumas dessas estruturas são apresentadas a seguir, admitindo-se uma situação simples:



- 1) **não estruturada**: todas as variâncias e covariâncias podem ser desiguais. Trata-se de uma matriz completamente geral. As variâncias são restritas a valores não negativos e as covariâncias não têm restrições, possui  $w = 10$  parâmetros:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix},$$

- 2) **componentes de variância**: caracterizada por variâncias iguais e covariâncias nulas. Sob pressuposição de normalidade, isto é equivalente a assumir que as observações são independentes, possui  $w = 1$  parâmetro:

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix},$$

- 3) **componentes de variância com heterogeneidade**: a estrutura conserva a suposição de independência, porém o mesmo não ocorre com a homogeneidade de variâncias. Essa estrutura possui número de parâmetros  $w = 4$ :

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix},$$

- 4) **simetria composta** (variância comum mais diagonal): possui  $w = 2$  parâmetros e admite igualdade de variâncias e de covariâncias, ou seja, covariâncias constantes entre quaisquer observações de uma mesma unidade devido a erros igualmente correlacionados:

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix},$$

5) **autorregressiva de 1ª ordem - AR(1)**: utilizada para dados de séries temporais igualmente espaçados e correlações diminuindo exponencialmente, ou seja, a covariância entre duas observações decresce à medida em que aumenta o intervalo de tempo entre elas e se denota por  $\rho$  o parâmetro autoregressivo (alcance), de forma que, para um processo estacionário, assume-se que  $|\rho| < 1$ , e  $w = 2$  parâmetros:

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix},$$

6) **espacial (ou potência)**: todas as variâncias são iguais e a correlação entre observações separadas por uma distância de  $d$  unidades é  $\rho^d$ , sendo  $\rho$  (parâmetro de alcance) a correlação entre observações vizinhas. Seu uso é aconselhado para os casos em que as observações repetidas não são equidistantes, possui  $w = 1$  parâmetro:

$$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ \rho^{d_{21}} & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ \rho^{d_{31}} & \rho^{d_{32}} & 1 & \rho^{d_{34}} \\ \rho^{d_{41}} & \rho^{d_{42}} & \rho^{d_{43}} & 1 \end{bmatrix}.$$

A seguir, são apresentadas algumas estruturas dentre as mais utilizadas na literatura, para modelos que apresentam características de dependência espacial. Nas três estruturas que seguem, a matriz de variâncias e covariâncias possui em cada uma de suas coordenadas, um valor para  $f(d_{ij})$ , que se trata do valor que a função assume na distância específica entre as duas amostras ( $d_{ij}$ ). Além disso, o parâmetro  $\rho$  está diretamente relacionado com o alcance de cada modelo.

7) **exponencial**:  $f(d_{ij}) = \exp\left(\frac{-d_{ij}}{\rho}\right)$ ,

8) **gaussiana**:  $f(d_{ij}) = \exp\left(\frac{-d_{ij}^2}{\rho^2}\right)$ ,

9) **esférica**:  $f(d_{ij}) = \begin{cases} \left[1 - 1,5\left(\frac{d_{ij}}{\rho}\right) + 0,5\left(\frac{d_{ij}}{\rho}\right)^3\right], & \text{se } d_{ij} < \rho, \\ 0, & \text{caso contrário.} \end{cases}$

Estas estruturas (7), (8) e (9) serão descritas detalhadamente, em termos de seus parâmetros geoestatísticos, na seção 2.5.

Barbosa (2009) alerta que as estimativas, os erros padrões de efeitos fixos, os diagnósticos e as inferências são influenciados pela escolha da estrutura de covariância. Além disso, afirma que a escolha depende de informações empíricas, da estrutura dos dados e ainda da disponibilidade computacional.

### 2.3.3 Seleção de modelos

Selecionar o melhor modelo significa, além de determinar a melhor estrutura para comparação de médias, escolher a melhor estrutura de covariâncias. A construção do modelo linear misto consta de três etapas: seleção dos efeitos fixos, escolha dos efeitos aleatórios e estimação, comparação de modelos e diagnósticos.

Littell, Pendergast e Natarajan (2002) sugerem que as comparações entre estruturas de variâncias e covariâncias sejam realizadas por meio dos critérios de informação de Akaike - AIC e Bayesiano - BIC, que são baseados no valor do logaritmo natural da função de verossimilhança do modelo, e dependem do número de observações e do número de parâmetros do modelo.

Akaike (1974) alerta que, caso dois modelos representem dados de forma satisfatória, então espera-se um melhor desempenho para a predição de novos dados com o mais simples destes modelos. Assim, o critério de Akaike penaliza o modelo de maior complexidade.

O critério AIC é baseado no estudo da função de verossimilhança e no número de parâmetros de cada modelo, sendo calculada para cada um destes modelos, a expressão segundo a expressão (4).

$$AIC = -2l(\hat{\beta}, \hat{\mathbf{u}}) + 2h, \quad (4)$$

sendo  $h$  o número total de parâmetros de efeitos fixos  $p + 1$  e aleatórios  $q + w$  estimados no modelo ( $q$  refere-se ao número de parâmetros do vetor  $\mathbf{u}$  em que o efeito de blocos é aleatório e  $w$  refere-se aos parâmetros de dependência espacial das matrizes de variâncias e covariâncias).

O critério BIC é bastante semelhante, penalizando com mais severidade os modelos que apresentam maior número de parâmetros, segundo a expressão (5).

$$BIC = -2l(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}) + h \log(n), \quad (5)$$

em que  $n$  é o número de observações utilizadas na estimação dos parâmetros do modelo, e  $h$  como definido para AIC.

Valores baixos para AIC e BIC são preferidos, indicando o modelo a ser escolhido dentro de uma comparação de modelos. Valores altos de verossimilhança implicam em baixos valores de AIC e BIC. Ambos critérios penalizam modelos com grande número de parâmetros, optando-se por modelos mais parcimoniosos. Nem sempre os dois critérios indicam um mesmo modelo, cabendo ao pesquisador decidir que critério seguir.

#### 2.3.4 Estimação dos componentes de variância

Segundo Barbin (1993), os componentes de variância são as variâncias associadas aos efeitos aleatórios de um modelo estatístico-matemático, e, de grande importância em genética e melhoramento animal e vegetal. A população e o método de melhoramento a serem utilizados numa análise dependem de algumas informações que podem ser obtidas a partir desses componentes.

No caso de modelos mistos, a solução das Equações de Modelos Mistos (EMM) que será apresentada na seção 2.3.5, é realizada a partir do conhecimento da matriz de variâncias e covariâncias  $\mathbf{V}$ , cuja estrutura é conhecida, porém, suas componentes não o são. Desse modo, torna-se necessário substituí-los por suas estimativas.

Diversos métodos têm sido propostos para estimar os componentes de variância, com destaque para o método dos momentos ou da análise de variância (ANOVA); método da máxima verossimilhança - ML (HARTLEY; RAO, 1971); método da estimação quadrática não-viesada de variância mínima - MIVQUE (RAO, 1971) e o

método da máxima verossimilhança restrita - REML. Existem ainda os métodos I, II e III de Henderson (1953) e aqueles derivados da estimação de funções quadráticas (MARCELINO; IEMMA, 2000).

Segundo Resende (2007), para obter estimativas tanto de ML quanto de REML, vários algoritmos têm sido desenvolvidos, dentre eles, o algoritmo estimação-maximização (EM), Escore de Fisher, Newton-Raphson e de Informação média (*Average Information REML*).

#### 2.3.4.1 Método ANOVA

O método da ANOVA consiste em igualar as esperanças dos quadrados médios de cada uma das fontes de variação presentes na análise da variância, aos seus respectivos quadrados médios, resolvendo assim um sistema linear.

Em geral, este método é adequado para modelos simples que envolvem dados balanceados. Os estimadores ANOVA são não-viesados e têm variância mínima, são funções de estatísticas suficientes, para as quais podem ser obtidas estimativas dos erros padrões associados, e uma aproximação dos números de graus de liberdade, por métodos como os propostos por Satterthwaite (1946), Fai e Cornelius (1996), Kenward e Roger (1997), citados por Alcarde (2012). Além disso, nenhuma suposição da distribuição dos dados, além das suposições básicas sobre as variâncias e covariâncias já mencionadas é exigida.

Porém, quando os dados são não balanceados, não existe um único modo de se obter a tabela da análise da variância, existindo diferentes estimativas para um mesmo componente. Como desvantagem pode-se citar o fato de que esse método permite a ocorrência de estimativas negativas para os componentes de variância, fato que torna a propriedade de estimador não viesado pouco interessante.

Como alternativa para contornar tal problema, pode-se utilizar a restrição do espaço paramétrico, ou seja, igualar as estimativas negativas a zero. Porém essa solução sacrifica a propriedade do estimador ser não viesado.

### 2.3.4.2 Método da Máxima Verossimilhança (ML)

O método da máxima verossimilhança (ML) foi desenvolvido por Fisher (1922) mas Hartley e Rao (1967) apresentaram a abordagem matricial de um modelo misto e a derivação das equações ML para várias classes de modelos. Além disso, Searle (1991) destacou que os trabalhos de Henderson (1953) tiveram grande impacto no desenvolvimento dos métodos de estimação de componentes de variância a partir de dados desbalanceados, estimulando principalmente os trabalhos de Hartley e Rao (1967).

Mesmo em situações de dados desbalanceados, Shaw (1987) afirma que os estimadores ML apresentam as seguintes propriedades desejáveis: suficiência (tal que o preditor condense o máximo possível a informação contida na amostra e não seja função dependente do parâmetro), consistência (relaciona o aumento da precisão da estimativa com o aumento do tamanho amostral), eficiência (o preditor apresenta variância mínima) e translação invariante (estimadores não afetados por mudanças nos efeitos fixos). Uma outra vantagem do método ML é a obtenção de estimativas não negativas dos componentes de variância.

Para a estimação ML de componentes de variância, Resende (2007) alerta que os efeitos fixos devem ser conhecidos, porém, eles não costumam ser conhecidos e são substituídos por suas estimativas obtidas por ML. Além disso, na estimação dos componentes de variância, o método ML não considera a perda de graus de liberdade devido à estimação desses efeitos fixos, causando então o viés. Tal viés conduz a subestimativas dos parâmetros de variância e, portanto podem conduzir a inferências incorretas. As estimativas são viesadas, mas possuem menores variâncias de amostragem do que aquelas obtidas a partir de estimadores não viesados.

Apesar de gerar estimativas viesadas, o procedimento ML é computacionalmente mais simples que o método REML e, em determinadas situações, apresenta eficiência satisfatória. O viés pode ser considerável se o número de equações independentes (posto de  $\mathbf{X}$ , em que  $\mathbf{X}$  é a matriz de delineamento dos efeitos fixos), para os efeitos fixos, for relativamente grande em relação ao número de observações ( $n$ ). Quando o posto de  $\mathbf{X}$  é pequeno com relação a  $n$ , os métodos ML e REML conduzem a resultados similares, conforme verificado por Resende et al. (1996).

### 2.3.4.3 Método da Máxima Verossimilhança Restrita (REML)

Segundo Resende (2007), um método alternativo, baseado na verossimilhança, para inferir sobre os componentes de variância nos modelos mistos é o método da máxima verossimilhança restrita (REML), desenvolvido por Patterson e Thompson (1971).

Os estimadores dos componentes de variância obtidos pelo método REML, têm sido amplamente adotados, porque eliminam o primeiro dos problemas encontrados no método ML, ou seja, leva em consideração os graus de liberdade envolvidos na estimação dos parâmetros fixos do modelo. Sendo assim, estimativas REML tendem a ser menos viesadas que as estimativas de ML, e permitem também a imposição de restrição de não negatividade. Dessa forma, o método REML é o procedimento ideal de estimação de componentes de variância com dados desbalanceados. Além disso, para o caso de dados balanceados, as soluções fornecidas pelo método REML em modelos mistos coincidem com as soluções fornecidas pelo método ANOVA (McCULLOCH; SEARLE; NEUHAUS, 2001).

De acordo com Resende (2007), o método REML é uma ferramenta flexível para a estimação de componentes de variância, efeitos fixos e a predição de efeitos aleatórios. Além disso, apresenta as seguintes vantagens:

- (i) Pode ser aplicado a dados desbalanceados.
- (ii) Permite modelar diversas estruturas de dados (medidas repetidas; análises em diferentes anos, locais ou delineamentos).
- (iii) Permite ajustar modelos com delineamentos que não podem ser acomodados pela ANOVA.
- (iv) Permite o ajuste de vários modelos alternativos, podendo-se escolher o que se ajusta melhor aos dados e, ao mesmo tempo, é parcimonioso (apresenta menor número de parâmetros).
- (v) Permite a correção simultânea para os efeitos ambientais, estimação de componentes de variância e predição de valores genéticos.
- (vi) Permite maior flexibilidade na modelagem, contemplando plenamente a análise de dados correlacionados devido ao parentesco, distribuição temporal e espacial.

O método REML requer que  $\mathbf{Y}$  tenha distribuição normal multivariada. Entretanto, vários autores relatam que os estimadores REML são também apropriados quando não se verifica normalidade dos dados (HARVILLE, 1977; MEYER, 1989).

Ao contrário do método ML, que maximiza a função de verossimilhança de todos os contrastes, o método REML maximiza a função de verossimilhança conjunta de todos os contrastes de erros ou resíduos,  $\mathbf{Y}^* = \mathbf{L}^T \mathbf{Y}$ , em que  $\mathbf{L}$  é uma matriz com  $[n - \text{posto}(\mathbf{X})]$  colunas, de posto completo, com colunas ortogonais às colunas da matriz  $\mathbf{X}$ , isto é,  $\mathbf{L}^T \mathbf{X} = \mathbf{0}$ . Em outras palavras, o método REML maximiza a parte da função de verossimilhança que é invariante aos efeitos fixos.

Dessa forma, considere a matriz  $\mathbf{L} = [\mathbf{L}_1 \mathbf{L}_2]$ , não singular, com  $\mathbf{L}_1$  e  $\mathbf{L}_2$  de dimensões  $(n \times (p + 1))$  e  $(n \times (n - p - 1))$ , respectivamente e satisfazem:

$$\mathbf{L}_1^T \mathbf{X} = \mathbf{I}_{p+1} \text{ e } \mathbf{L}_2^T \mathbf{X} = \mathbf{0}.$$

Assim, os dados podem ser então transformados de  $\mathbf{y}$  para  $\mathbf{L}^T \mathbf{y}$ , ou seja,

$$\mathbf{L}^T \mathbf{y} = [\mathbf{L}_1 \mathbf{L}_2]^T \mathbf{y} = \begin{bmatrix} \mathbf{L}_1^T \mathbf{y} \\ \mathbf{L}_2^T \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p+1} \boldsymbol{\beta} + \mathbf{L}_1^T \mathbf{Z} \mathbf{u} + \mathbf{L}_1^T \boldsymbol{\epsilon} \\ \mathbf{L}_2^T \mathbf{Z} \mathbf{u} + \mathbf{L}_2^T \boldsymbol{\epsilon} \end{bmatrix}.$$

A variável aleatória  $\mathbf{y}^T = [\mathbf{y}_1^* \mathbf{y}_2^*]^T$  tem distribuição com esperança e variância, respectivamente, dadas por:

$$E(\mathbf{Y}^*) = E \left[ \begin{pmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{pmatrix} \right] = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}$$

e

$$Var(\mathbf{Y}^*) = Var \left[ \begin{pmatrix} \mathbf{Y}_1^* \\ \mathbf{Y}_2^* \end{pmatrix} \right] = \begin{pmatrix} \mathbf{L}_1^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V} \mathbf{L}_2 \end{pmatrix}.$$

Logo,

$$\begin{pmatrix} \mathbf{y}_1^* \\ \mathbf{y}_2^* \end{pmatrix} \sim N \left[ \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{L}_1^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_1^T \mathbf{V} \mathbf{L}_2 \\ \mathbf{L}_2^T \mathbf{V} \mathbf{L}_1 & \mathbf{L}_2^T \mathbf{V} \mathbf{L}_2 \end{pmatrix} \right].$$

A distribuição completa de  $\mathbf{L}^T \mathbf{y}$  pode ser subdividida em uma distribuição condicional,  $\mathbf{Y}_1^* | \mathbf{Y}_2^*$  para a estimação de  $\boldsymbol{\beta}$ , e uma distribuição marginal, baseada em  $\mathbf{Y}_2^*$ , para a estimação dos componentes de variância (Alcarde, 2012). Esta distribuição marginal, por sua vez, é a base da máxima verossimilhança restrita.



Por outro lado, seja  $\boldsymbol{\kappa}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\alpha}^T)$  o vetor de componentes de variância, tal que  $\boldsymbol{\gamma}$  contém os componentes de variância associados a  $\mathbf{u}$  e  $\boldsymbol{\alpha}$  os componentes de variância associados a  $\boldsymbol{\epsilon}$ , sua estimação é baseada no logaritmo da função de verossimilhança restrita:

$$\ell_R = -\frac{1}{2} \left[ \log \det (\mathbf{L}_2^T \mathbf{V}^{-1} \mathbf{L}_2) + \mathbf{y}^{T*} (\mathbf{L}_2^T \mathbf{V} \mathbf{L}_2)^{-1} \mathbf{y} \right],$$

que pode ser expressa da seguinte forma:

$$\ell_R = -\frac{1}{2} \left[ \log \det (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}) + \log \det \mathbf{V} + \mathbf{y}^T \mathbf{P} \mathbf{y} \right], \quad (6)$$

em que,

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$$

e ainda,

$$\mathbf{y}^T \mathbf{P} \mathbf{y} = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$

Segundo Alcarde (2012), as estimativas REML de  $\kappa_l$ , tal que  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_L)$ , são obtidas igualando a função escore a zero, da seguinte forma:

$$U(\kappa_l) = \frac{\partial \ell_R}{\partial \kappa_l} = \frac{1}{2} \left[ \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \right) - \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \mathbf{y} \right].$$

Note que os elementos da matriz informação observada são:

$$\begin{aligned} -\frac{\partial^2 \ell_R}{\partial \kappa_l \partial \kappa_k} &= -\frac{1}{2} \text{tr} \left( \mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \right) - \frac{1}{2} \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right) + \\ &+ \mathbf{y}^T \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \mathbf{P} \mathbf{y} - \frac{1}{2} \mathbf{y}^T \mathbf{P} \frac{\partial^2 \mathbf{V}}{\partial \kappa_l \partial \kappa_k} \mathbf{P} \mathbf{y}, \end{aligned}$$

e, os elementos da matriz informação esperada são:

$$E \left( -\frac{\partial^2 \ell_R}{\partial \kappa_l \partial \kappa_k} \right) = \frac{1}{2} \text{tr} \left( \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_l} \mathbf{P} \frac{\partial \mathbf{V}}{\partial \kappa_k} \right).$$

No entanto, a solução para  $\mathbf{U}(\kappa_l) = 0$  exige um algoritmo iterativo que utiliza uma estimativa inicial ( $\boldsymbol{\kappa}^{(0)}$ ) para  $\boldsymbol{\kappa}$  e uma atualização  $\boldsymbol{\kappa}^{(1)}$ . Um desses algoritmos é o “Escore de Fisher”, em que:

$$\boldsymbol{\kappa}^{(1)} = \boldsymbol{\kappa}^{(0)} + \mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})^{-1} \mathbf{U}(\boldsymbol{\kappa}^{(0)}),$$

em que  $\mathbf{U}(\boldsymbol{\kappa}^{(0)})$  é o vetor escore e  $\mathbf{I}(\boldsymbol{\kappa}^{(0)}, \boldsymbol{\kappa}^{(0)})$  representa a matriz de informação esperada de  $\boldsymbol{\kappa}$ , avaliada em  $\boldsymbol{\kappa}^{(0)}$  (ALCARDE, 2012).

Conforme as dimensões da inversa da matriz de informação aumentam, a utilização do algoritmo “Escore de Fisher” tende a apresentar dificuldades computacionais, que são discutidas em Gilmour, Thompson e Cullis (1995), contudo, segundo estes autores, uma alternativa para evitar a sobrecarga computacional, é a utilização do algoritmo AI que apresenta propriedades de convergência semelhantes ao algoritmo “Escore de Fisher”.

### 2.3.5 Estimação dos efeitos fixos e predição dos efeitos aleatórios

As estimativas de  $\boldsymbol{\beta}$  e  $\mathbf{u}$  são obtidas por meio do uso de funções baseadas na função de verossimilhança dos dados. Assim, se a função densidade de probabilidade de  $\mathbf{y}$  é dada por:

$$f(\mathbf{y}) = \frac{1}{2\pi^{\frac{n}{2}} |\mathbf{ZGZ}^T + \mathbf{R}|^{\frac{1}{2}}} \exp \left\{ \frac{1}{2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{ZGZ}^T + \mathbf{R})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \right\},$$

a função densidade de probabilidade conjunta de  $\mathbf{y}$  e  $\mathbf{u}$ , por sua vez, pode ser escrita como o produto entre a função densidade de probabilidade condicional de  $\mathbf{y}$  dado  $\mathbf{u}$ , e a função densidade de probabilidade de  $\mathbf{u}$ , ou seja,

$$f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u})f(\mathbf{u}),$$

ou ainda,

$$f(\mathbf{y}, \mathbf{u}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \times \exp \left\{ \frac{1}{2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})] \right\} \times \\ \times \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{G}|^{\frac{1}{2}}} \exp \left\{ \frac{1}{2} [(\mathbf{u} - \mathbf{0})^T \mathbf{G}^{-1} (\mathbf{u} - \mathbf{0})] \right\},$$

cujo logaritmo denotado por  $\ell$  é dado por:

$$\ell = -n \log(2\pi) - \frac{1}{2}(\log |\mathbf{R}| + \log |\mathbf{G}|) + \frac{1}{2}(\mathbf{y}^T \mathbf{R}^{-1} \mathbf{y} - 2\mathbf{y}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} - 2\mathbf{y}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{u}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} \boldsymbol{\beta} + \mathbf{u}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{u} + \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u}).$$

Derivando-se  $\ell$  em relação a  $\boldsymbol{\beta}$  e  $\mathbf{u}$  e igualando-se as expressões resultantes a  $\mathbf{0}$ , obtêm-se as Equações de Modelos Mistos (EMM) propostas por Henderson (1984):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}$$

Se  $\mathbf{G}$  e  $\mathbf{R}$  são conhecidas, então o estimador de mínimos quadrados generalizados (*generalized least squares*-GLS), ou o melhor estimador linear não viesado (*best linear unbiased estimator*-BLUE) de  $\boldsymbol{\beta}$ , é dado por:

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

em que  $\mathbf{V}$  é dada segundo (3).

De forma análoga, o melhor preditor linear não viesado (*best linear unbiased predictor* -BLUP) de  $\mathbf{u}$  é dado por:

$$\hat{\mathbf{u}} = \mathbf{G} \mathbf{Z}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}).$$

Esses estimadores são denominados “melhores” por minimizarem a variância amostral, “lineares”, pois são funções lineares de  $\mathbf{y}$  e “não viesados”, porque  $E[BLUE(\boldsymbol{\beta})] = \boldsymbol{\beta}$  e  $E[BLUP(\mathbf{u})] = \mathbf{u}$ . Além disso, a matriz de variâncias e covariâncias para estes estimadores é dada por:

$$\mathbf{C} = cov \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1}, \quad (7)$$

em que  $[\ ]^{-1}$  denota inversa generalizada.

Entretanto, na maioria das vezes,  $\text{Var}(\mathbf{y}) = \mathbf{V}$  é desconhecida, e conseqüentemente, as matrizes  $\mathbf{G}$  e  $\mathbf{R}$  são desconhecidas. Neste caso, assumindo-se certa perda de eficiência, os parâmetros em  $\mathbf{G}$  e  $\mathbf{R}$  podem ser substituídos pelas estimativas

correspondentes, constituindo, respectivamente, as matrizes denotadas por  $\hat{\mathbf{G}}$  e  $\hat{\mathbf{R}}$ . Logo, substituindo-se  $\mathbf{G}$  por  $\hat{\mathbf{G}}$  e  $\mathbf{R}$  por  $\hat{\mathbf{R}}$  em (7), tem-se a matriz  $\hat{\mathbf{C}}$  dada por:

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-1},$$

como uma aproximação da matriz de variâncias e covariâncias dos estimadores em questão. Neste caso, os termos BLUE e BLUP não se aplicam, sendo apropriado substituí-los por EBLUE (melhor estimador linear não viesado empírico ou *empirical best linear unbiased estimator*) e EBLUP (melhor preditor linear não viesado empírico ou *empirical best linear unbiased predictor*), respectivamente, de acordo com Littell et al. (2006). O termo empírico é adicionado, portanto, para indicar esse tipo de aproximação.

Segundo McLean e Sander (1988), a matriz  $\hat{\mathbf{C}}$  pode ser reescrita como:

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}_{21} \\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22} \end{bmatrix},$$

em que,

$$\hat{\mathbf{C}}_{11} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1},$$

$$\hat{\mathbf{C}}_{21} = -\hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} \mathbf{X} \hat{\mathbf{C}}_{11},$$

e

$$\hat{\mathbf{C}}_{22} = (\mathbf{Z}^T \hat{\mathbf{R}}^{-1} \mathbf{Z} + \hat{\mathbf{G}}^{-1})^{-1} - \hat{\mathbf{C}}_{21} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Z} \hat{\mathbf{G}}.$$

Nota-se que  $\hat{\mathbf{C}}_{11}$  é a equação para obter estimativas de mínimos quadrados generalizados dos elementos da matriz de variâncias e covariâncias de  $\hat{\boldsymbol{\beta}}$ .

Logo, utilizando a matriz  $\hat{\mathbf{C}}$ , as estimativas dos parâmetros de  $\boldsymbol{\beta}$  e as predições dos parâmetros de  $\mathbf{u}$  são obtidas como segue:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{V}}^{-1} \mathbf{y}$$

e

$$\hat{\mathbf{u}} = \hat{\mathbf{G}} \mathbf{Z}^T \hat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}),$$

em que  $\hat{\mathbf{V}}^{-1}$  é obtida substituindo-se os parâmetros em  $\mathbf{V}$ , pelas estimativas correspondentes.

Adicionalmente, objetivando a construção de intervalos de  $(1 - \alpha) \times 100\%$  de confiança (IC) para as estimativas dos parâmetros  $\beta_i$ , calculam-se os erros padrões  $\sqrt{\widehat{Var}(\hat{\beta}_i)}$ , a fim de que os ICs nos forneçam o campo de variação de cada um destes parâmetros, como segue:

$$\text{IC}(\beta_i) = \hat{\beta}_i \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\beta}_i)},$$

sendo  $i = 1, \dots, p + 1$  parâmetros de efeito fixo,  $\alpha$  o nível de significância,  $z_{\alpha/2}$  o valor tal que  $P(|Z| < z_{\alpha/2}) = 1 - \alpha$ , sendo  $Z$  uma variável com distribuição normal padronizada,  $\widehat{Var}(\hat{\beta}_i)$ , a variância associada à estimativa do parâmetro de efeito fixo  $\beta_i$ .

E ainda, os intervalos de  $(1 - \alpha) \times 100\%$  de confiança (IC) relacionados às estimativas de componentes de variância ( $\hat{\sigma}_i^2$ ), podem ser estimados usando o método delta, em que estas estimativas são linearizadas por meio da função logarítmica ( $\log$ ) e possuem distribuição normal assintótica. O IC na escala original da variável resposta é dado por:

$$\text{IC}(\sigma_i^2) = \exp\left(\log(\hat{\sigma}_i^2) \pm z_{1-\alpha/2} \hat{\sigma}_i^{-2} \sqrt{\widehat{Var}(\hat{\sigma}_i^2)}\right)$$

### 2.3.6 Inferência para parâmetros de efeitos fixos

Molenberghs e Verbeke (2005) e West, Welch e Galecki (2007) dentre outros, apresentam os testes de Wald,  $t$  e Wald- $F$ , que podem ser utilizados para verificar a significância dos termos fixos, ou de uma combinação linear dos mesmos, em modelos mistos.

O teste de Wald aproximado pode ser realizado para cada  $\beta_i$  em  $\boldsymbol{\beta}$ , com  $i=1, \dots, p$ , assim como a obtenção do respectivo intervalo de confiança. Seja  $\mathbf{L}$  uma matriz de constantes conhecidas e de posto completo  $c$  ( $c \leq p$ ), de dimensões  $c \times p$ , e considerando-se as hipóteses:

$$H_0 : \mathbf{L}\boldsymbol{\beta} = 0 \text{ contra } H_1 : \mathbf{L}\boldsymbol{\beta} \neq 0 \quad (8)$$

Tem-se que a estatística de Wald ( $W$ ) é dada por:

$$W = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T \left[ \mathbf{L} \left( \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i(\hat{\boldsymbol{\kappa}}) \mathbf{X}_i \right)^{-1} \mathbf{L}^T \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

em que  $\hat{\boldsymbol{\kappa}}$  é o vetor das estimativas dos parâmetros das matrizes de variância e covariância  $\mathbf{G}$  e  $\mathbf{R}$ , denominados componentes de variância. Sob  $H_0$  a estatística  $W$  tem distribuição assintótica  $\chi^2$  com  $r$  graus de liberdade.

Por outro lado, devido ao fato do teste de Wald não considerar a variabilidade introduzida pela estimação dos componentes de variância, podendo subestimar a variação dos efeitos fixos, Molenberghs e Verbeke (2005) sugerem a utilização dos testes  $t$  e Wald- $F$  para testar hipóteses sobre os parâmetros de efeito fixo.

O teste  $t$  é frequentemente utilizado para testar hipóteses do tipo:

$$H_0 : \beta_i = 0 \text{ contra } H_1 : \beta_i \neq 0,$$

em que a estatística  $t$  é calculada da seguinte maneira:

$$t = \frac{\hat{\beta}_i}{\sqrt{\widehat{Var}(\hat{\beta}_i)}},$$

que, sob a hipótese nula, segue uma distribuição  $t$  de Student com  $\nu$  graus de liberdade, que dependem exclusivamente dos dados e são calculados utilizando métodos como o de Satterhwaite (1946), Fai e Cornelius (1996) ou Kenward e Roger (1997).

Por outro lado, a estatística  $F$  do teste Wald- $F$  possui o número de graus de liberdade do numerador igual ao posto da matriz  $\mathbf{L}$ , e o número de graus de liberdade do denominador, também é calculado baseado nos métodos acima citados. Esta estatística é definida por:

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{L}^T \left[ \mathbf{L} \left( \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i(\hat{\boldsymbol{\kappa}}) \mathbf{X}_i \right)^{-1} \mathbf{L}^T \right]^{-1} \mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\text{posto}(\mathbf{L})}.$$

Uma alternativa é a aplicação do teste da razão de verossimilhanças (LRT) cuja estatística é dada a seguir:

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

em que  $L_1$  representa a função de verossimilhança do modelo aninhado ou restrito (o modelo mais simples, referente à hipótese nula) e  $L_2$ , a função de verossimilhança do modelo de referência (o modelo mais geral ou completo). Neste caso, os modelos aninhado e de referência devem conter os mesmos componentes de variância e mesmas estruturas para as matrizes  $\mathbf{G}$  e  $\mathbf{R}$ , porém diferentes conjuntos de parâmetros fixos. Essa prática permite verificar a importância dos termos fixos do modelo, uma vez que a diferença entre tais modelos encontra-se apenas com relação a esses termos.

Para este caso, a estatística para o teste da razão de verossimilhança segue, assintoticamente, a distribuição  $\chi^2$  com número de graus de liberdade igual à diferença entre o número de parâmetros de efeito fixo dos modelos em questão. Entretanto, para os casos em que os parâmetros encontram-se na fronteira do espaço paramétrico, a estatística do teste da razão de verossimilhança segue uma mistura de distribuições  $\chi^2$  (SELF; LIANG, 1987).

### 2.3.7 Inferência para parâmetros de efeitos aleatórios

De acordo com Resende (2007), o uso da análise de variância para a construção de testes  $F$  para os efeitos aleatórios em modelos desbalanceados é muito difícil. Isto porque é necessária a obtenção dos quadrados médios a partir dos componentes de variância e seus multiplicadores, que são muito difíceis de ser computados sob desbalanceamento. Há, no entanto, uma maneira mais formal para testar os efeitos aleatórios, ou seja, para verificar se determinado efeito aleatório necessita permanecer no modelo. Essa abordagem formal baseia-se em estatísticas fundamentadas na verossimilhança.

Segundo Pinheiro e Bates (2000), os modelos de referência e aninhado, devem ser estimados utilizando o mesmo procedimento. West, Welch e Galecki (2007), por sua vez, sugerem o uso do método REML para a estimação dos componentes de variância, já que proporciona estimativas menos viesadas, comparadas com o método ML. Além disso, quando o teste é realizado para componentes aleatórios, a especificação da parte fixa deve ser a mesma para os dois modelos.

Uma das estatísticas utilizadas para testar as hipóteses  $H_0 : \sigma_i^2 = 0$  e  $H_1 : \sigma_i^2 > 0$  (existe variabilidade entre os níveis do fator aleatório  $i$ ) é a  $Z$  de Wald, que é calculada dividindo-se a estimativa do parâmetro por seu erro padrão assintótico. Os

erros padrões assintóticos são obtidos a partir da inversa da matriz de derivada segunda da verossimilhança, em relação a cada um dos parâmetros de efeito aleatório (matriz informação de Fisher).

A estatística  $Z$  de Wald é válida para grandes amostras, mas ela pode ser incerta para pequenos conjuntos de dados e para componentes de variância, que apresentam uma distribuição assimétrica.

Segundo Resende (2007), uma melhor alternativa é utilizar o teste da razão de verossimilhança (*Likelihood Ratio Test* -LRT), e recomenda calcular previamente a relação  $\frac{\hat{\sigma}^2}{e.p.(\hat{\sigma}^2)}$ , em que  $\hat{\sigma}^2$  é a estimativa de um componente de variância de um determinado efeito aleatório e  $s(\hat{\sigma}^2)$  seu respectivo erro padrão; e aplicar o LRT apenas quando  $1 < \frac{\hat{\sigma}^2}{e.p.(\hat{\sigma}^2)} < 2$ .

Quanto à distribuição da estatística do teste da razão de verossimilhança restrita sob a hipótese de nulidade, West, Welch e Gałeczki (2007) discriminam dois casos que dependem se os valores dos componentes de variância envolvidos na hipótese estão, ou não, na fronteira do espaço paramétrico. Os dois casos são:

(i) Os parâmetros de covariância referentes à hipótese de nulidade não estão na fronteira do espaço paramétrico, sendo que, o interesse está na verificação da homogeneidade de variâncias, ou ainda, se a covariância entre dois efeitos aleatórios é igual a zero. Nesses casos, a estatística segue assintoticamente a distribuição  $\chi^2$  com número de graus de liberdade igual à diferença entre o número de parâmetros nos modelos de referência e aninhado.

(ii) Os parâmetros de covariância estão na fronteira do espaço paramétrico: são os casos em que se deseja verificar se um efeito aleatório deve, ou não, permanecer no modelo. Neste caso, Stram e Lee (1994) demonstraram que a estatística para o teste da razão de verossimilhanças para um único parâmetro de variância, que se encontra na fronteira do espaço paramétrico, segue uma mistura de distribuições  $\chi^2$ ,  $0,5\chi_0^2 + 0,5\chi_1^2$ . Para os casos em que  $k$  parâmetros se encontram na fronteira o espaço paramétrico, a estatística segue também uma mistura de distribuições  $\chi^2$ , porém, nesse caso a mistura é dada por  $0,5\chi_0^2 + 0,5\chi_k^2$ .

Self e Liang (1987) apresentaram, adicionalmente, as distribuições para outros casos, como teste simultâneo para parâmetros de variância para os quais um se encontra na fronteira do espaço paramétrico e outro não.



### 2.3.8 Diagnóstico de resíduos

Os diagnósticos devem ser parte do processo de construção de modelos e análise de conjunto de dados. Os resíduos são utilizados para examinar as suposições do modelo estatístico-matemático e detectar a presença de valores atípicos e possíveis observações influentes.

Lembrando que as médias marginal e condicional nos modelos mistos são dadas por  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$  e  $E(\mathbf{y}|\mathbf{u}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ , respectivamente, dois tipos de resíduos são apresentados a seguir:

- (i) Resíduos marginais, que consistem da diferença entre o valor observado e a média marginal estimada. Neste caso, o vetor  $\mathbf{r}_m$  de resíduos marginais é definido como:

$$\mathbf{r}_m = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

- (ii) Resíduos condicionais, que consistem da diferença entre o valor observado e o valor predito da observação. Neste caso, o vetor  $\mathbf{r}_c$  de resíduos condicionais é definido como:

$$\mathbf{r}_c = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{u}} = \mathbf{r}_m - \mathbf{Z}\hat{\mathbf{u}}.$$

Por outro lado, segundo Gregoire, Schabenberger e Barrett (1995), dadas as matrizes  $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T$  e  $\mathbf{K} = \mathbf{I} - \mathbf{Z}\hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}$ , tem-se que

$$\widehat{Var}(\mathbf{r}_m) = \hat{\mathbf{V}} - \mathbf{Q}$$

e

$$\widehat{Var}(\mathbf{r}_c) = \mathbf{K}(\hat{\mathbf{V}} - \mathbf{Q})\mathbf{K}^T.$$

Alcarde (2012) considera que os resíduos  $\mathbf{r}_m$  e  $\mathbf{r}_c$  não são adequados para diagnósticos, pois podem apresentar correlações, mesmo para dados não correlacionados, e possuem difícil interpretação quando são incorporadas variâncias distintas ao modelo. Um modo de minimizar tais problemas consiste em trabalhar com os resíduos padronizados ou os resíduos de Pearson, descritos a seguir:

(i) Resíduo marginal estudentizado

$$r_{mi}^{estudentizado} = \frac{r_{mi}}{\sqrt{\widehat{Var}(r_{mi})}};$$

(ii) Resíduo condicional estudentizado

$$r_{ci}^{estudentizado} = \frac{r_{ci}}{\sqrt{\widehat{Var}(r_{ci})}};$$

(iii) Resíduo marginal de Pearson

$$r_{mi}^{Pearson} = \frac{r_{mi}}{\sqrt{\widehat{Var}(y_i)}};$$

(iv) Resíduo condicional de Pearson

$$r_{ci}^{Pearson} = \frac{r_{ci}}{\sqrt{\widehat{Var}(y_i|\mathbf{u})}}.$$

A autora recomenda, ainda, considerar o melhor preditor linear não viesado (BLUP) de  $\mathbf{u}$ , para diagnosticar os efeitos aleatórios. West, Welch e Galecki (2007) sugerem adicionalmente, a utilização de gráficos de diagnósticos padrões, ou seja, histogramas, gráficos de quantil-quantil e gráficos de dispersão para a verificação da normalidade dos resíduos, e nesse caso, dos resíduos condicionais estudentizados.

Uma análise de resíduos é fundamental e indispensável na validação e adoção de um determinado modelo como o mais adequado, porém, os resíduos podem ser afetados caso a área em estudo apresente dependência espacial, sendo então necessária a busca por modelos alternativos.

## 2.4 Variabilidade espacial em ensaios agrícolas

Dados agrícolas são espacialmente dependentes, conforme Martínez (1994), quando a observação de uma unidade experimental fornece alguma informação para uma certa unidade adjacente (relativamente próxima), como o seu valor ou magnitude. As observações tomadas em condições similares podem muito bem ser espacialmente dependentes ou independentes de acordo com a distância entre as observações vizinhas.

A partir dos anos 90, diversos autores passaram a desenvolver trabalhos, sob diferentes metodologias, visando contornar a dependência espacial muitas vezes presente e ignorada.

Dentre os pioneiros nesta abordagem destacam-se Grondona e Cressie (1991) e Zimmerman e Harville (1991), que propõem um estudo que considera a disposição espacial das unidades experimentais, por meio da geoestatística. Os dois últimos autores levam em consideração a presença de autocorrelação espacial, a qual modela diretamente o efeito da parcela vizinha, de forma que as observações são consideradas coletivamente como uma realização parcial de um campo aleatório.

Hadarbach (1996) por sua vez, considerou a heterogeneidade espacial em ensaios que se dispõem de grandes quantidades de tratamentos, trabalhando-se então com delineamentos experimentais aumentados.

Oda (2005) fez uso de modelos com estruturas de dependência espacial para um experimento em delineamento sistemático tipo “leque”, verificando neste estudo um ganho com o uso do modelo geoestatístico.

Jaramillo e Francisco (2005) contornaram o problema da variabilidade espacial do solo em estudo de fertilização de feijão, trabalhando com os resíduos das funções esférica e exponencial na comparação de médias de tratamentos.

Pontes e Oliveira (2003) similarmente aos anteriores, superaram a existência de erros correlacionados por meio de ferramentas geoestatísticas, diminuindo as estimativas de médias de erros padrões, em um delineamento em blocos aumentados, para estudar a produtividade de soja em diversas progênies.

Nogueira, Lima e Oliveira (2013) simularam um conjunto de dados no delineamento em blocos ao acaso e foram melhorando este modelo com o auxílio de uma função de dependência espacial. Os autores concluíram que uma análise de variância a partir de modelos que levam a dependência espacial em conta pode resultar em conclusões bem diferentes da análise habitual.

Todos estes trabalhos tiveram por objetivo comum a obtenção de melhores modelos para maior confiabilidade nos resultados das análises, sendo que em todos os casos, o solo teve característica heterogênea.

Deste modo, Jaramillo e Francisco (2005) afirmam, que uma característica habitual do solo é a sua heterogeneidade, mesmo em pequenas áreas que poderiam ser consideradas homogêneas, isto se deve ao fato de que a formação do solo envolve vários processos e interações. Esta heterogeneidade de solo induz uma variabilidade nas suas propriedades que pode ser de magnitude considerável e que pode afetar as generalizações e previsões realizadas em tal solo.

Segundo Littel et al. (2006), os métodos estatísticos de correlação espacial podem ser divididos em dois grupos: caracterização e ajuste. A caracterização envolve o cálculo de estimativas de parâmetros de covariância, enquanto o ajuste corresponde à remoção dos efeitos de correlação espacial para obter mais qualidade nas estimativas ajustadas.

Pode-se interpretar um experimento espacial como um ensaio comparativo em que as unidades experimentais ocupam posições fixas distribuídas por toda uma região em uma, duas, ou três dimensões, segundo Zimmerman e Harville (1991). Especificamente, no caso da cultura de cana-de-açúcar, um experimento pode ser visto como um problema espacial. Uma característica comum em experimentos espaciais é a presença de heterogeneidade sistemática entre as unidades experimentais. Tipicamente, a natureza da presente heterogeneidade é tal que há uma correlação significativa entre as unidades vizinhas.

Para Martínez (1994), confia-se que a aleatorização seja capaz de neutralizar os efeitos da correlação espacial presentes nas unidades experimentais vizinhas, podendo-se assim validar os resultados das análises de variância realizadas com estas unidades.

Resende e Thompson (2004) afirmam que o uso de blocos, como forma de controle local, é determinado antes da implantação dos ensaios, de forma que percebe-se muitas vezes, por meio da coleta dos dados experimentais, a presença de manchas ou gradientes ambientais dentro das áreas onde estão instalados os experimentos, os quais não foram considerados adequadamente pelos blocos delineados a priori. Os referidos autores alertam ainda que a variabilidade ou heterogeneidade espacial associada à estrutura do solo, umidade e interceptação de luz dentre outros fatores ambientais contribui para o aumento da variação residual. Desta forma, é importante controlar, a partir do delineamento ou da análise, a variação residual espacial ou a tendência geográfica que os dados possam apresentar.

Duarte (2000) ressalta em seu trabalho a necessidade de utilizar técnicas de análise espacial, capazes de contornar o problema da correlação a partir da flexibilização da matriz de variâncias e covariâncias, baseando-se nos próprios dados experimentais.

Segundo Jaramillo (2005), se uma variável apresenta características de dependência espacial, esta pode violar o princípio da independência entre as amostras e os procedimentos da estatística paramétrica clássica passam a não ser adequados para a análise. Neste caso, a variabilidade espacial existe devido ao fato de que as áreas agrícolas apresentam valores diferentes dependendo da localização e/ou espaçamento entre as amostras utilizadas, sendo que desta forma, o valor atribuído a uma variável em determinada posição geográfica depende da distância e/ou direção que esta se encontra de um local vizinho.

Upchurch e Edmonds (1991) argumentam que a variabilidade espacial tem dois componentes fundamentais, um aleatório e um sistemático, levando-se em conta a fonte do erro que produz a variação. Quando a variabilidade não pode ser relacionada a causas conhecidas, ela é definida como variabilidade aleatória. A variabilidade sistemática é aquela que pode ser atribuída a causas compreensíveis e previsíveis.

Se a correlação depende exclusivamente da distância entre as medidas (magnitude), os modelos que estimam as covariâncias entre as observações são chamados estacionários. As funções de correlação para os modelos estacionários podem ser isotrópicas ou anisotrópicas. As funções isotrópicas são idênticas em qualquer direção (dependem apenas da magnitude das distâncias) enquanto as anisotrópicas permitem diferentes valores de seus parâmetros em diferentes direções, e assim dependem também da direção em que as distâncias são calculadas.

Samra, Anlauf e Weber (1990) equacionam a variabilidade sistemática em dois fatores: um de tendência e o outro de variabilidade espacial, expressando ainda a variação total como resultante da soma de três variabilidades: tendência, espacial e aleatória.

Es e Es (1993) mostraram que, dentro deste problema de correlação espacial, os testes estatísticos associados a contrastes entre médias de tratamentos, cujas parcelas estiverem separadas por pequenas distâncias, têm maior probabilidade de erro tipo II e os contrastes cujas parcelas estiverem separadas por distâncias maiores são testados com maior probabilidade de erro tipo I.

Stroup, Baenziger e Mulitze (1994) salientam que os efeitos de parcelas distribuem-se de acordo com algum modelo de correlação espacial descrevendo assim, as tendências locais, de maneira similar aos modelos geoestatísticos de predição. Sendo assim, um delineamento instalado em uma área com estas características pode ser estudado por meio de um modelo linear misto (MLM), admitindo-se os erros espacialmente correlacionados.

Di Rienzo, Macchiavelli e Casanoves (2010) alertam sobre a possibilidade de se utilizar funções geoestatísticas no contexto dos MLM para realizar a modelagem da estrutura espacial das parcelas, contemplando assim a estrutura de correlação entre estas. Além disso, esta modelagem pode lidar com a heterogeneidade de variâncias residuais entre locais.

## 2.5 Modelos geoestatísticos

A geoestatística apresenta técnicas eficazes na verificação e quantificação da dependência espacial entre amostras de uma determinada área experimental em estudo. O processo de análise de dados espacialmente referenciados e com característica de dependência espacial, abrange, a princípio, métodos gráficos exploratórios, avançando posteriormente, para técnicas mais refinadas, que possibilitam a escolha de um modelo com a maior capacidade possível de explicar a dependência espacial presente em um determinado conjunto de dados. Segundo Ribeiro Jr. e Diggle (2001), este estudo visa a realização de predições em locais de onde não se amostrou e a estimação dos parâmetros do modelo selecionado.

A metodologia mais utilizada para representar a dependência espacial na geoestatística, é o semivariograma, capaz de descrever a associação espacial dos pontos amostrais em função da distância entre eles. Segundo Faraco (2006), o semivariograma se trata de uma técnica exploratória que auxilia os métodos geoestatísticos, contribuindo na modelagem da dependência espacial, sugerindo os parâmetros básicos do modelo: alcance ( $a$ ), efeito pepita ou variância aleatória ( $C_0$ ), variância espacial ( $C$ ) e patamar ( $C_0 + C$ ).

De acordo com o mesmo autor, o semivariograma consiste na elaboração de classes de pares de pontos que possuem mesma (ou similar) distância um do outro, para a realização de uma espécie de estudo da evolução da correlação, conforme se aumenta a distância entre os pontos de um mesmo grupo.

Assim, a construção de um semivariograma empírico pode ser realizada a partir do cálculo das semivariâncias  $\gamma(h)$  (9), obtendo-se pares de pontos, que plotados, formam um gráfico de formas similares às da Figura 1, sendo que  $N(h)$  representa o número de pares amostrados separados por uma distância  $h$  (ou aproximadamente  $h$  em casos em que não há regularidade da malha de amostras em estudo), e ainda  $y(x_i)$  e  $y(x_i + h)$  são valores medidos nas posições  $x_i$  e  $x_i + h$ , respectivamente.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [y(x_i + h) - y(x_i)]^2 \quad (9)$$

Na Figura 1 são apresentados o semivariograma teórico e o empírico, ambos com características de dependência espacial. O semivariograma teórico exibe uma noção intuitiva de comportamento ideal para uma variável com tal característica, enquanto o semivariograma empírico, nos fornece um esboço legítimo, verificado a partir de dados reais, com pontos distribuídos de forma desordenada, mas permitindo utilizar uma função geoestatística que melhor se adequa a tais pontos.

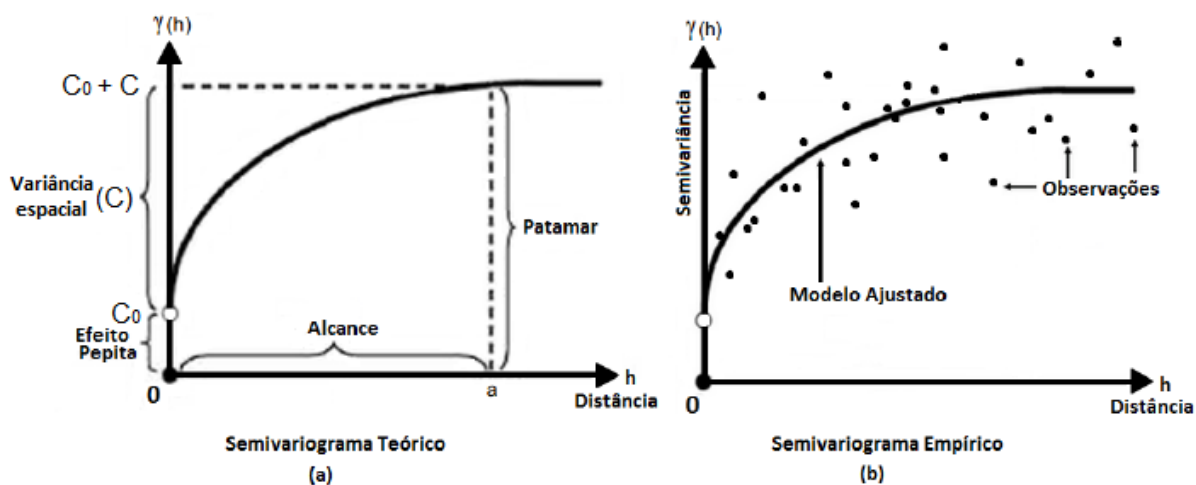


Figura 1 – Semivariogramas (a) Teórico e (b) Empírico com características de dependência espacial

Sob este enfoque, Resende (2007) destaca que para valores grandes de  $h$ , a covariância diminuirá ao passo que a variância aumentará, pois amostras mais próximas entre si tendem a ser mais parecidas do que aquelas separadas por distâncias maiores. Além disso, geralmente,  $\gamma(h)$  aumenta até estabilizar, conforme  $h$  cresce. O valor de  $h$

no momento em que a curva se estabiliza é denominado alcance (*range*) que é o raio de dependência espacial, enquanto aliado a ele, o valor de  $\gamma(h)$  representa a variância dos dados e é denominado patamar (*sill*).

Segundo Cressie (1993), embora  $\gamma(h)$  seja igual a zero para  $h$  igual a zero, na prática é comum que à medida que  $h$  tende para zero,  $\gamma(h)$  se aproxime de um valor positivo, valor este denotado por  $C_0$  e denominado por efeito pepita (*nugget effect*).

Yamamoto e Landim (2013) afirmam que o efeito pepita por ser a variância aleatória dos dados, pode ser atribuído a erros de medição ou ao fato dos dados não terem sido coletados a intervalos suficientemente pequenos, para exibir o comportamento espacial do fenômeno estudado. Valores calculados de efeito pepita relativamente grandes, denotam uma grande variabilidade nas medidas feitas à distâncias extremamente pequenas. Adicionalmente,  $C$  é denominado o componente estrutural (variância espacial), ou ainda, a variabilidade que é explicada pelo modelo.

Com base em um modelo geoestatístico pode-se ter interesse em quantificar a força, ou o grau de dependência da variável em estudo. Para isto foi criado o Índice de Dependência Espacial (*IDE*), que mede o quanto o valor medido em uma amostra pode influenciar em suas parcelas vizinhas. O *IDE* relaciona os efeitos pepita ( $C_0$ ) e patamar ( $C_0 + C$ ), classificando a dependência espacial como forte, moderada ou fraca. Dentre diversas propostas para o cálculo do *IDE*, Zimback (2001) sugeriu o quociente:

$$IDE = \frac{C}{C_0 + C} \times 100. \quad (10)$$

Assim, a classificação quanto ao grau de dependência espacial da variável em estudo, é efetuada da seguinte forma:

- i) variável com fraca dependência espacial: se  $IDE \leq 0,25$ ,
- ii) variável com moderada dependência espacial: se  $0,25 < IDE < 0,75$ ,
- iii) variável com forte dependência espacial: se  $0,75 \leq IDE$ .

Vale destacar como situação extrema, conforme alertam Silva et al. (2011), no caso em que  $IDE = 0$ , está sendo representado um modelo de efeito pepita puro, ou ainda, um fenômeno não necessariamente espacial aleatório, não existindo correlação alguma entre os valores, e deste modo a análise por meio de semivariograma não se aplica.



A partir do momento em que foi verificada a existência de dependência espacial da variável em estudo, segundo os autores acima, pode-se considerar  $h$  como um vetor que depende de sua magnitude e direção, assim, quando o variograma for idêntico para qualquer direção de  $h$ , ele é considerado isotrópico, e assim, as matrizes de variâncias e covariâncias dadas na seção 2.3.2 assumem o mesmo valor para  $f(d_{ij})$  ( $i$ -ésima linha e  $j$ -ésima coluna) para todos os pares de pontos que são equidistantes, independentemente de direção (ao longo da linha, coluna ou diagonal).

Para a modelagem da dependência espacial em questão, segundo Littell et al. (2006), é necessária a definição dos modelos de correlação espacial como segue:

$$\text{Var}[\epsilon_i] = \sigma_i^2,$$

$$\text{Cov}[\epsilon_i, \epsilon_j] = \sigma_{ij}$$

Assim, assume-se a covariância como uma função de distâncias entre posições. Sendo  $d_{ij}$  a distância euclidiana entre as amostras  $s_i$  e  $s_j$ , e os modelos de covariância são do tipo:

$$\text{Cov}[\epsilon_i, \epsilon_j] = \sigma^2[f(d_{ij})]$$

Em relação a esta abordagem, o parâmetro  $\sigma^2$  corresponde ao patamar (*sill*), e  $\rho$  caracteriza o alcance do processo (*range*).

Resende (2007) destaca que para a descrição da dependência espacial, a construção do semivariograma é suficiente, porém, para a estimação de parâmetros e predição de valores, o ajuste de modelos indicados pelo semivariograma é necessário.

A partir da elaboração do semivariograma e definição de seus parâmetros, busca-se definir modelos que melhor descrevam a dependência presente no conjunto de dados em estudo. Costuma-se ajustar mais de um modelo para um semivariograma, e os modelos ajustados podem ser comparados a partir dos critérios AIC ou BIC, por exemplo, para seleção do mais adequado.

Segundo Mello (2004), existem diversos modelos, em especial na geoestatística, aplicáveis a diferentes fenômenos com continuidade espacial. Dentre os principais modelos de correlação destacam-se os modelos exponencial, gaussiano e esférico.

Estes modelos foram apresentados na seção 2.3.2 como opções para a composição das matrizes de covariâncias e serão a seguir detalhados no conceito geoestatístico.

### 2.5.1 Modelo exponencial

Segundo Pannatier (1996), o modelo exponencial (11), é muito utilizado no ajuste de semivariogramas para descrever uma variável com dependência espacial. Isaaks e Srivastava (1989) afirmam, que este modelo converge assintoticamente para o patamar ( $C_0 + C$ ) com o alcance prático ( $a$ ), no momento em que a distância ( $h$ ) do modelo corresponde a aproximadamente 95% de  $C$  (Figura 2). Caso o efeito pepita ( $C_0$ ) seja muito pequeno e a estrutura de variabilidade cresça de maneira bastante suave, o semivariograma pode ser melhor ajustado pelo modelo gaussiano.

$$\gamma(h) = C_0 + C \left[ 1 - e^{-\frac{3h}{a}} \right], \text{ se } h \geq 0; \quad (11)$$

### 2.5.2 Modelo gaussiano

De acordo com Pannatier (1996), o modelo gaussiano (12), é utilizado para modelar fenômenos contínuos. Isaaks e Srivastava (1989) destacam que este modelo é caracterizado pelo comportamento parabólico próximo de sua origem, sendo este o único modelo que apresenta um ponto de inflexão. Semelhante ao modelo exponencial, o modelo gaussiano atinge o patamar assintoticamente e o parâmetro de alcance é definido para a distância ( $h$ ) que corresponde a 95% do patamar  $C$  (Figura 2).

$$\gamma(h) = C_0 + C \left[ 1 - e^{-3\left(\frac{h}{a}\right)^2} \right] \text{ se } h \geq 0; \quad (12)$$

### 2.5.3 Modelo esférico

O modelo esférico (13) chama atenção em sua característica, segundo Pannatier (1996), devido ao comportamento “quase linear” deste modelo de semivariograma, para pequenos valores de  $h$ . O autor observa ainda que a tangente ao gráfico na origem deste modelo, atinge o patamar ( $C$ ) a 2/3 do alcance. O modelo esférico é aproximadamente linear até cerca de 1/3 do alcance (Figura 2).

$$\gamma(h) = \begin{cases} C_0, & \text{se } h = 0 \\ C_0 + C \left[ 1,5\left(\frac{h}{a}\right) - 0,5\left(\frac{h}{a}\right)^3 \right], & \text{se } 0 < h \leq a \\ C_0 + C, & \text{se } h > a \end{cases} \quad (13)$$

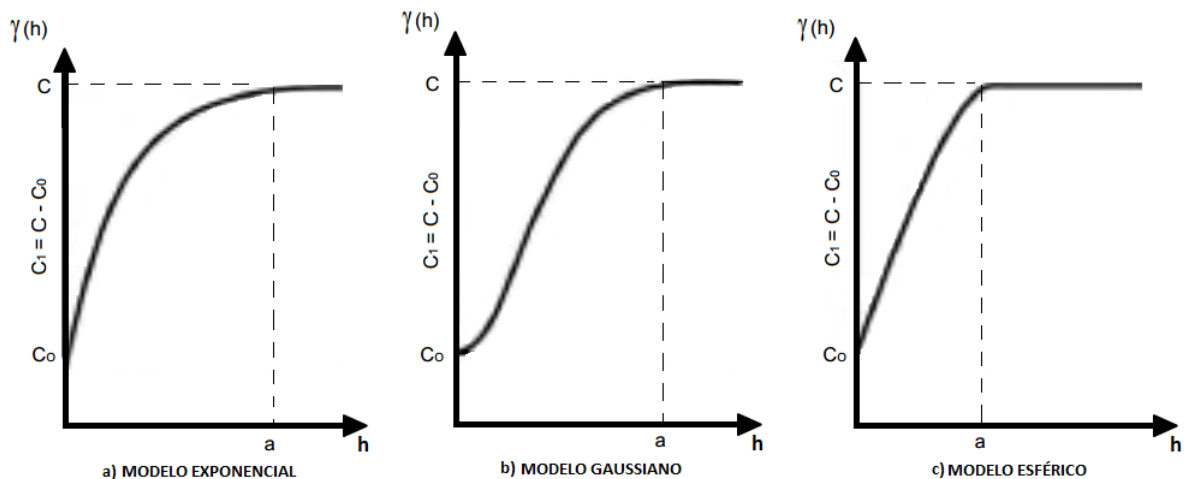


Figura 2 – Modelos teóricos (a) exponencial, (b) gaussiano e (c) esférico

Para fazer uma analogia entre os modelos 7, 8 e 9 da seção 2.3.2, e os modelos 11, 12 e 13 desta seção, sob os conceitos e parâmetros geoestatísticos, pode-se dizer que o parâmetro  $\rho$  dos modelos 7, 8 e 9, em todos os casos está diretamente relacionado ao alcance ( $a$ ) do processo. No modelo exponencial,  $3\rho$  corresponde ao valor do alcance, no modelo esférico,  $\rho$  coincide com o valor do alcance, por fim, no modelo gaussiano,  $\rho\sqrt{3}$  representa o alcance.

Os três modelos espaciais exponencial, gaussiano e esférico utilizados neste estudo, são apenas alguns dentre os modelos geoestatísticos existentes, porém, são indiscutivelmente os mais utilizados na literatura.

### 3 MATERIAL E MÉTODOS

Nesta seção é abordado, em termos de dados reais no enfoque de modelos lineares mistos, um estudo de um grupo de quatro experimentos agronômicos conduzidos na Guatemala, em que a variável resposta de interesse é a produção de cana-de-açúcar (toneladas de cana-de-açúcar por hectare). Utilizando o critério de Akaike (AIC), foram comparados diversos modelos construídos a partir de diferentes estruturas de variâncias e covariâncias, visando analisar a relevância da modelagem da matriz residual por meio de funções geoestatísticas.

Além disso, foram simulados diversos conjuntos de dados experimentais objetivando a comparação de diferentes estruturas nas matrizes de variâncias e covariâncias, com o intuito de observar a proporção de situações em que se considera mais adequado o uso da abordagem de modelos mistos, modelando-se a matriz residual, em comparação aos métodos tradicionais, em diferentes cenários.

#### 3.1 Material

##### 3.1.1 Grupo de experimentos de cana-de-açúcar

O conjunto de dados reais utilizados neste estudo foi disponibilizado pelo coordenador da área de Malezas y Madurantes do Programa de Agronomía do Centro Guatemalteco de Investigación y Capacitación de la Caña de Azúcar (CENGICANÑA).

Neste trabalho foi considerado um grupo de quatro experimentos conduzidos nos locais: (1) Fazenda Limones (usina açucareira Pantaleón) - LP, (2) Fazenda Bálsamo (usina açucareira Pantaleón) - BP, (3) área 1 da Fazenda Limones (usina Madre Tierra) - MT1 e (4) área 2 da Fazenda Limones (usina Madre Tierra) - MT2. Em todos os locais foi utilizado o mesmo delineamento experimental, o casualizado em blocos, com cinco repetições e com os mesmos seis tratamentos (Tabela 1). A estrutura dos dados pode ser visualizada na Tabela 2. Os experimentos foram conduzidos de maio a dezembro de 2013, utilizando a variedade CP 88-1165 (CP = Canal Point, Florida; 88 = ano da seleção - ano de 1988; 1165 = número ordenado da seleção).

Cada unidade experimental foi constituída de 5 fileiras de cana, de 10 metros de comprimento e com espaçamento de 1,5 metros, totalizando uma área de 75 m<sup>2</sup>.

Tabela 1 – Tratamentos estimulantes (Trat.) para cana-de-açúcar baseados em ácido giberélico ( $AG_3$ )

Trat.	Descrição
T1	Dose de 0,37g de ácido giberélico ( $AG_3$ ) aplicada nas plantas aos 40-45, 90-120 e 150-180 dias depois do plantio
T2	Dose de 0,37g de ácido giberélico ( $AG_3$ ) aplicada nas plantas aos 90-120 e 150-180 dias depois do plantio
T3	Dose de 0,74g de ácido giberélico ( $AG_3$ ) aplicada nas plantas aos 90-120 e 150-180 dias depois do plantio
T4	Dose de 1,11g de ácido giberélico ( $AG_3$ ) aplicada nas plantas aos 90-120 e 150-180 dias depois do plantio
T5	Dose de 0,40g de ácido giberélico ( $AG_3$ ) + 5kg de sulfato de amônio + 5kg de melaço + 15kg de ureia aplicada aos 140 a 170 dias depois do plantio
T6	Testemunha: Tratamento sem estimulante ( $AG_3$ )

A respeito dos tratamentos, segundo Cardoso (2007), a aplicação dos reguladores vegetais compostos pelo ácido giberélico ( $AG_3$ ) já ocorre em várias espécies hortícolas, frutíferas e ornamentais, visando a obtenção de plantas floridas fora de época. Adicionalmente, de acordo com Taiz e Zeiger (2006), estes reguladores promovem o crescimento das células a partir da degradação de substâncias da parede celular por meio de ativação de enzimas hidrolíticas.

Cardoso (2007) afirma ainda que a aplicação de  $AG_3$ , na maioria das vezes, é feita via pulverização foliar nas plantas, o que torna esta técnica ainda mais viável, já que o produtor de cana-de-açúcar está habituado a aplicar fertilizantes nas plantas.

Com relação ao tratamento 5 (T5), é importante considerar que o melaço da cana é resultante do processo de fabricação do açúcar e utilizado na fermentação para produção de álcool, em especial o etanol. Além disso, o melaço é utilizado amplamente em rações animais, no caso deste experimento, a função do melaço foi de atuar como coadjuvante. Adicionalmente, o sulfato de amônio contém 21% de nitrogênio, e é preparado por reação da amônia com o ácido sulfúrico, tratando-se de um composto inorgânico usado como fertilizante. A ureia contém 45% de nitrogênio, e trata-se de um fertilizante adicional, possibilitando a complementação da quantidade necessária de nitrogênio no solo, para que se obtenha melhor produtividade nas culturas.

Tabela 2 – Estrutura de dados balanceados para cada um dos 4 experimentos do grupo em estudo (locais: (1) LP, (2) BP, (3) MT1 e (4) MT2) sob o delineamento em blocos ao acaso, com 5 repetições, em que são avaliados os 6 tratamentos estimulantes de crescimento de plantas de cana-de-açúcar

Local	Trat.	Blocos				
		1	2	3	4	5
1	1	$y_{111}$	$y_{121}$	$y_{131}$	$y_{141}$	$y_{151}$
	2	$y_{211}$	$y_{221}$	$y_{231}$	$y_{241}$	$y_{251}$
	3	$y_{311}$	$y_{321}$	$y_{331}$	$y_{341}$	$y_{351}$
	4	$y_{411}$	$y_{421}$	$y_{431}$	$y_{441}$	$y_{451}$
	5	$y_{511}$	$y_{521}$	$y_{531}$	$y_{541}$	$y_{551}$
	6	$y_{611}$	$y_{621}$	$y_{631}$	$y_{641}$	$y_{651}$

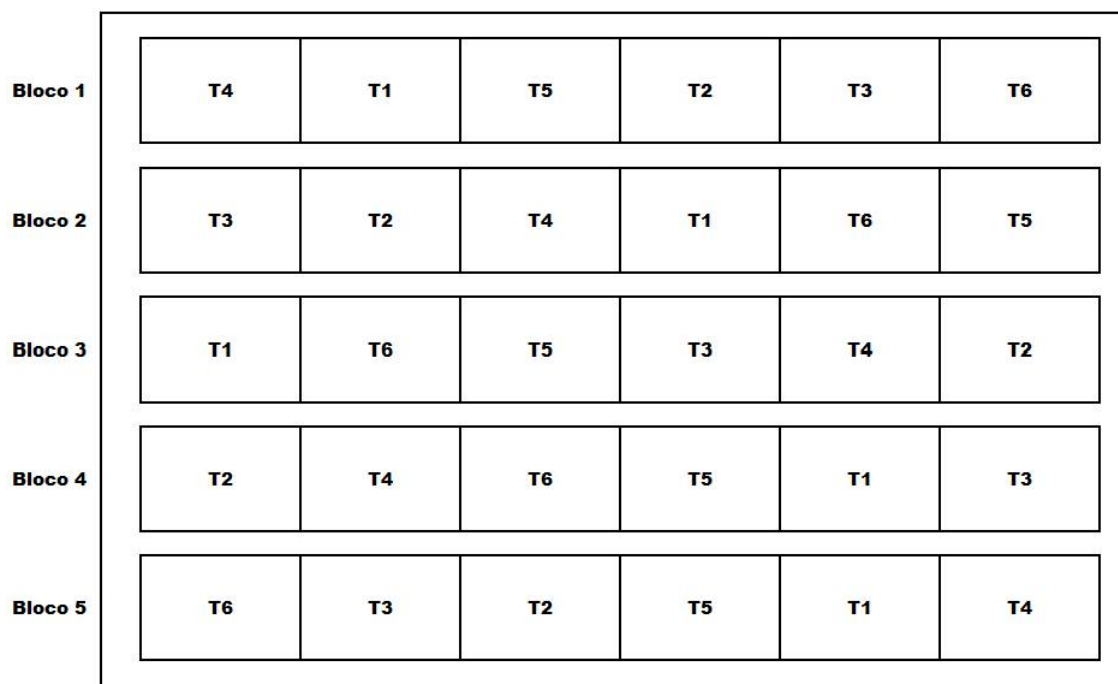


Figura 3 – Croqui de instalação do experimento com o georreferenciamento das parcelas para cada local

A Figura 3 apresenta um esboço do croqui utilizado nos quatro experimentos, representando o resultado do sorteio (aleatorização) dos tratamentos dentro de cada bloco para cada local.

A Tabela 3 contém um resumo das primeiras sete parcelas segundo suas características principais a serem relevadas na construção dos modelos, em especial, vale destacar as colunas Lat e Long referentes às posições geográficas de cada uma das amostras conforme o sorteio definiu.

Tabela 3 – Resumo das primeiras linhas do banco de dados, detalhadas segundo o local, tratamento (Trat), bloco e produção (Prod), além das coordenadas de latitude (Lat) e longitude (Long) referentes a cada parcela

Local	Trat	Bloco	Lat	Long	Prod
1	4	1	1	1	116,7
1	1	1	1	2	103,4
1	5	1	1	3	90,1
1	2	1	1	4	100,1
1	3	1	1	5	92,7
1	6	1	1	6	88,9
1	3	2	2	1	122,7
⋮	⋮	⋮	⋮	⋮	⋮

Característica fundamental para esta abordagem, o georreferenciamento dos dados possibilitou o estudo da dependência espacial residual possivelmente existente, a partir da construção de modelos lineares mistos, utilizando funções geoestatísticas (seções 2.3.2 e 2.5). O croqui dos experimentos (Figura 3), representando o resultado dos sorteios (aleatorização) dos tratamentos dentro de cada bloco, serviram de suporte para o conhecimento das posições de cada uma das parcelas, e suas respectivas vizinhas, visando a construção dos semivariogramas e utilização das funções geoestatísticas de interesse.

Sob estas circunstâncias, a variável resposta de interesse foi a produção de cana-de-açúcar em toneladas por hectare ( $\text{tm ha}^{-1}$ ), sendo considerados dois objetivos principais para este estudo: (i) avaliar uma possível existência de diferenças significativas entre as diferentes combinações de doses e períodos de aplicação do ácido giberélico ( $AG_3$ ), além de um tratamento testemunha; (ii) comparar modelos lineares mistos (MLM) que possibilitam a incorporação de diferentes estruturas de matrizes de variâncias e covariâncias contra modelos tradicionais que admitem que os resíduos são independentes.

### 3.1.2 Grupo de experimentos simulados

Com o objetivo de obtenção de resultados mais abrangentes, foram gerados conjuntos de dados simulando características e fenômenos reais, para que, por meio do estudo do comportamento de diversos modelos, segundo o critério de Akaike (AIC), seja possível comparar e avaliar a importância de se utilizar uma modelagem que considere a existência de dependência espacial, a partir de funções geoestatísticas para a matriz  $\mathbf{R}$ , em cenários em que o grau de dependência espacial vai de fraco a forte. O software R (2014) possibilitou este estudo de simulação em diversos cenários diferentes.

## 3.2 Métodos

Nesta seção serão propostos e apresentados os modelos de interesse para solucionar o problema da possível dependência espacial para os dados de produção de cana-de-açúcar recém citados. Os mesmos modelos serão também utilizados no estudo de simulação de dados, visando verificar o comportamento de modelos de diferentes estruturas em relação a dependência espacial.

### 3.2.1 Modelos mistos para o grupo de experimentos de cana-de-açúcar

Para o estudo deste grupo de experimentos, foi considerado o modelo estatístico-matemático a seguir, com algumas variações:

$$y_{ijk} = m + t_i + b_{j(k)} + l_k + tl_{ik} + \varepsilon_{ijk}, \quad (14)$$

admitindo-se  $i = 1, \dots, I$  tratamentos,  $j = 1, \dots, J$  blocos e  $k = 1, \dots, K$  locais,  $y_{ijk}$  é o valor observado referente ao  $i$ -ésimo tratamento, no  $j$ -ésimo bloco do  $k$ -ésimo local;  $m$  é uma constante inerente a todas as observações;  $t_i$  é o efeito do  $i$ -ésimo tratamento (efeito fixo);  $b_{j(k)}$  é o efeito do  $j$ -ésimo bloco no  $k$ -ésimo local, que será considerado fixo em alguns casos e aleatório com  $b_{j(k)} \sim N(0, \sigma_b^2)$  em outros;  $l_k$  é o efeito fixo do  $k$ -ésimo local;  $tl_{ik}$  representa o efeito da interação entre o  $i$ -ésimo tratamento e o  $k$ -ésimo local, e  $\varepsilon_{ijk}$  é o erro experimental aleatório associado às observações  $y_{ijk}$ , com  $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ .

Considerando-se  $b_{j(k)}$  aleatório e  $\varepsilon_{ijk}$  como independentes, a variância de uma observação é dada por:  $\sigma_b^2 + \sigma_\varepsilon^2$ .



Os fatores de variação do modelo 14 referentes a tratamentos e locais foram sempre considerados fixos por representarem apenas seus casos particulares, não constituindo uma amostra representativa de fatores similares.

Restringe-se a comparação dos modelos a suas estruturas de parcelas, sendo que diferentes estruturas de parcelas induzem a distintas estruturas de correlação entre as observações. Esta comparação de estruturas pode ser realizada a partir dos modelos lineares mistos, incluindo funções geoestatísticas para a explicação da variabilidade espacial.

A seguir, serão apresentados os 16 modelos a serem comparados. Em cada modelo foram combinadas diferentes estruturas de matrizes de variâncias e covariâncias residuais ( $\mathbf{R}$ ) com três funções geoestatísticas, visando um melhor ajuste para os dados de cana-de-açúcar:

- **M01 - Modelo BF**: o efeito de blocos foi considerado fixo, os erros são independentes para a matriz  $\mathbf{R}$  e a variância constante nos diversos locais.
- **M02 - Modelo BFH**: foi considerado o efeito de blocos como fixo, os erros como sendo independentes para a matriz  $\mathbf{R}$ , porém, foram admitidas diferentes variâncias entre os locais.
- **M03 - Modelo BFExp**: correlação espacial exponencial (item 7, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e a variância constante nos diversos locais.
- **M04 - Modelo BFExpH**: foi considerada uma correlação espacial exponencial (item 7, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e as variâncias diferentes entre os locais.
- **M05 - Modelo BFGau**: foi adotada uma correlação espacial gaussiana (item 8, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e a variância constante nos diversos locais.
- **M06 - Modelo BFGauH**: foi considerada uma correlação espacial gaussiana (item 8, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e as variâncias diferentes entre os locais.

- **M07 - Modelo BF<sub>Esf</sub>**: foi adotada uma correlação espacial esférica (item 9, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e a variância constante nos diversos locais.
- **M08 - Modelo BF<sub>EsfH</sub>**: foi considerada uma correlação espacial esférica (item 9, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado fixo e as variâncias diferentes entre os locais.
- **M09 - Modelo BA**: foi considerado o efeito de blocos aleatório para a matriz  $\mathbf{G}$ , os erros independentes para a matriz  $\mathbf{R}$  e a variância constante nos diversos locais.
- **M10 - Modelo BAH**: foi considerado o efeito de blocos como aleatório para a matriz  $\mathbf{G}$ , os erros independentes para a matriz  $\mathbf{R}$  e as variâncias diferentes entre os locais.
- **M11 - Modelo BA<sub>Exp</sub>**: foi admitida correlação espacial exponencial (item 7, seção 2.3.2) para a matriz  $\mathbf{R}$ , o efeito de blocos foi considerado aleatório para a matriz  $\mathbf{G}$ , além de variância constante nos diversos locais.
- **M12 - Modelo BA<sub>ExpH</sub>**: a correlação espacial para a matriz  $\mathbf{R}$  foi dada pela função exponencial (item 7, seção 2.3.2), o efeito de blocos foi considerado aleatório para a matriz  $\mathbf{G}$ , e as variâncias diferentes entre os locais.
- **M13 - Modelo BA<sub>Gau</sub>**: foi admitida estrutura de correlação espacial gaussiana (item 8, seção 2.3.2) para a matriz  $\mathbf{R}$ , possui efeito de blocos aleatório para a matriz  $\mathbf{G}$ , além de variância constante nos diversos locais.
- **M14 - Modelo BA<sub>GauH</sub>**: a correlação espacial para a matriz  $\mathbf{R}$  foi dada pela função gaussiana (item 8, seção 2.3.2), o efeito de blocos foi considerado aleatório para a matriz  $\mathbf{G}$ , e as variâncias diferentes entre os locais.
- **M15 - Modelo BA<sub>Esf</sub>**: admite correlação espacial esférica (item 9, seção 2.3.2) para a matriz  $\mathbf{R}$ , apresentou efeito de blocos aleatório para a matriz  $\mathbf{G}$ , além de variância constante nos diversos locais.
- **M16 - Modelo BA<sub>EsfH</sub>**: a correlação espacial para a matriz  $\mathbf{R}$  também foi dada pela função esférica (item 9, seção 2.3.2), o efeito de blocos foi considerado aleatório para a matriz  $\mathbf{G}$ , e as variâncias diferentes entre os locais.

O método da máxima verossimilhança restrita - REML (PATTERSON; THOMPSON, 1971) foi adotado neste estudo visando a obtenção de estimativas com menor viés para o vetor de componentes de variância, e, além disso, o método dos mínimos quadrados generalizados, foi utilizado para a estimação do vetor de parâmetros de efeito fixo. Adicionalmente, foi utilizada a estrutura de componentes de variância para modelar a matriz  $\mathbf{G}$  quando o efeito de blocos é considerado aleatório, e por fim, a matriz  $\mathbf{R}$  foi construída a partir das funções geoestatísticas.

Para a comparação de modelos, foi necessária a definição de dois grupos de modelos. O primeiro grupo contém os modelos em que o efeito de blocos foi considerado fixo, ou seja, modelos que possuem os mesmos fatores de efeitos fixos, enquanto por outro lado, o segundo grupo foi composto por modelos com efeito aleatório para blocos. Os modelos foram confrontados, dentro de seus grupos, e selecionados a partir do critério de informação de Akaike (AIC), conforme (4), visando a escolha do mais adequado para descrever os dados em cada situação. Tal comparação foi possível a partir do software estatístico R (2014) e encontra-se disponível para consulta no Anexo A.

Após selecionados os dois modelos mais adequados, segundo critério de Akaike, foram construídos quadros de análise de variância para estes modelos, com o intuito de avaliar a significância dos fatores de efeito fixo e da interação tratamentos  $\times$  locais, utilizando para isto, o teste Wald- $F$ , como descrito na seção 2.3.6. Tais procedimentos foram realizados a partir do software estatístico SAS (SAS INSTITUTE, 2011) e estão disponíveis para consulta no Anexo B.

Para testar os efeitos de tratamentos e da interação entre tratamentos e locais, as estatísticas segundo o teste Wald- $F$  utilizadas nesta abordagem, tanto no caso de efeito de blocos fixo, quanto no caso aleatório, por meio dos quadrados médios (QM) dos fatores de variação, são dadas segundo as equações a seguir:

$$F_T = \frac{QM_{Trat.}}{QM_{Res}},$$

$$F_{T \times L} = \frac{QM_{T \times L}}{QM_{Res}},$$

$$F_L = \frac{QM_{Local}}{QM_{Res}}.$$

Por outro lado, no caso de blocos de efeito aleatório, o teste Wald- $F$  para locais é dado por:

$$F_L = \frac{QMLocal}{QMBloco(Local)}.$$

De forma complementar, quando o fator de blocos foi admitido aleatório, foram testadas as hipóteses referentes aos componentes de variância,  $H_0 : \sigma_i^2 = 0$  e  $H_a : \sigma_i^2 > 0$  (existe variabilidade entre os níveis do fator aleatório  $i$ ). Para este propósito, foi utilizado o teste  $Z$  de Wald, descrito na seção 2.3.7.

Em seguida, os diagnósticos foram abordados observando-se os gráficos de dispersão para os resíduos condicionais estudentizados, conforme descrito na seção 2.3.8. A verificação da normalidade tanto para o vetor de parâmetros de efeito aleatório, quanto para os resíduos, foi realizada utilizando os gráficos de quantil-quantil, e a homocedasticidade por meio do gráfico de resíduos contra valores preditos.

Como abordagem adicional, devido ao fato da não existência de significância da interação entre tratamentos e locais, não pode-se tirar conclusões generalizadas, e assim, foram realizadas as análises de variância individuais segundo o modelo estatístico-matemático dado como segue:

$$y_{ij} = m + t_i + b_j + \varepsilon_{ij}, \quad (15)$$

O modelo dado em (15) de forma similar ao modelo (14), baseia-se no uso do efeito de blocos ora como fixo, ora como aleatório, enquanto o efeito de tratamentos é sempre considerado fixo.

Com relação à definição e construção dos modelos a serem comparados, foram utilizados os modelos ímpares descritos nesta seção (M01, M03, M05, M07, M09, M11, M13 e M15), desconsiderando as informações a respeito de locais, e comparando as estruturas segundo as funções geoestatísticas responsáveis pela modelagem de uma possível dependência espacial.

Desta forma, para cada um dos quatro locais em estudo, foram novamente construídos dois grupos de interesse, um abrangendo os modelos que consideram o efeito

de blocos como fixo e o outro, como efeito aleatório. O primeiro grupo contém os modelos M01, M03, M05 e M07 e o segundo grupo contém os modelos M09, M11, M13 e M15. Assim, para cada local foram selecionados dois modelos: um de efeito fixo para blocos e o outro de efeito aleatório para blocos, possibilitando verificar a necessidade do uso de modelos com dependência espacial, e observando possíveis diferenças entre médias de tratamentos para cada local, segundo os modelos selecionados.

### 3.2.2 Modelos mistos para o grupo de experimentos simulados

De forma oposta ao habitual, visando a verificação da importância e validade do uso de modelos lineares mistos (MLM), cujas estruturas das matrizes de variâncias e covariâncias foram definidas por meio de funções geoestatísticas, para explicar uma possível dependência espacial presente no local em estudo, optou-se por realizar um estudo de simulação.

Para a simulação de conjuntos de dados de grupos de experimentos caracterizados pelo delineamento em blocos ao acaso, cada grupo foi constituído por determinados número de locais (4), de repetições (5) e de tratamentos (5) (os mesmos tratamentos nos quatro locais). Foram adotados apenas modelos fixos nesta abordagem. O processo de simulação foi realizado por meio do software estatístico R (2014).

Neste estudo, foram considerados apenas modelos com o efeito fixo para blocos (14), com o intuito de investigar o implemento da estrutura de dependência espacial a partir do critério de Akaike (AIC).

Além disso, foram abordadas diferentes situações a partir da modelagem da matriz  $\mathbf{R}$  por meio das funções geoestatísticas exponencial, esférica e gaussiana, além de distintos Índices de Dependência Espacial - IDE (10), sendo eles alto (90%), moderado (50%) e baixo (10%). O cruzamento das 3 estruturas de dependência espacial com os 3 níveis de IDE constituem os 9 cenários de simulação de interesse neste trabalho.

Foram verificados dentro de cada cenário de simulação, conforme as descrições da seção 3.2.1 o número de vezes que cada um dos 8 modelos (M01-M08) foi o mais vantajoso, sendo simulados 500 grupos de 4 experimentos, para serem construídos e comparados os 8 modelos em cada um destes grupos de experimentos.

Para todos os modelos simulados, foi considerado o efeito pepita igual a 1 ( $C_0 = 1$ ). Para os modelos com fraco IDE (10%), a variância espacial considerada foi de 0,1111 ( $C = 0,1111$ ), para os modelos com moderado IDE (50%), a variância espacial utilizada foi de 1 ( $C = 1$ ), e, para modelos com forte IDE (90%), utilizou-se  $C = 9$ .

Adicionalmente, para todos os locais gerados neste estudo, foram simuladas 25 amostras em uma malha regular  $5 \times 5$ , com base nas coordenadas de cada parcela, a partir dos três modelos geoestatísticos de interesse, considerando-se um alcance espacial  $a = 7$ . Considerou-se ainda para cada local, uma média geral  $m = 10$ , além de variâncias predeterminadas para gerar os efeitos fixos de tratamentos (5) e de blocos (5). O programa referente ao estudo de simulação, e os valores iniciais para uma possível reprodução dos resultados aqui obtidos, estão disponíveis no Anexo C.



## 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados das análises realizadas para grupos de experimentos referentes a produção de plantas de cana-de-açúcar, sendo este grupo composto por 4 locais. Os 16 modelos lineares mistos (MLM) descritos na seção 3.2.1, com diferentes estruturas no que diz respeito ao fator blocos, e às matrizes de variâncias e covariâncias, foram construídos e comparados sendo utilizadas em alguns casos funções geoestatísticas para modelar a dependência espacial dos resíduos ( $\mathbf{R}$ ). A partir de selecionados o modelo com efeito fixo para blocos e o modelo com efeito aleatório para blocos, foram realizadas análises de resíduos visando verificar a qualidade dos ajustes destes modelos.

### 4.1 Avaliação do grupo de experimentos de cana-de-açúcar

Inicialmente, foi realizada uma análise estatística descritiva que forneceu um esboço inicial a respeito do comportamento dos dados de produção de cana-de-açúcar em estudo (Tabela 4), considerando os valores médios e respectivos desvios padrões para cada local.

Tabela 4 – Médias e desvios padrões (d.p.) da produção de cana-de-açúcar (ton/ha) nos quatro locais

Trat.	LP		BP		MT1		MT2		Total	
	Média	d.p.	Média	d.p.	Média	d.p.	Média	d.p.	Média	d.p.
T1	95,34	24,24	192,30	10,31	137,74	15,28	122,70	15,08	137,02	39,47
T2	112,20	19,17	190,74	7,56	134,84	12,63	118,14	10,33	138,98	34,03
T3	106,06	25,54	190,50	8,99	125,88	11,57	126,34	12,51	137,19	35,82
T4	103,74	25,16	186,38	6,10	134,50	10,90	124,20	10,04	137,20	34,13
T5	90,18	20,03	186,48	6,17	141,16	12,25	121,78	15,09	134,90	38,13
T6	110,94	17,78	192,18	7,04	131,82	7,38	128,90	5,49	140,96	32,90

A Tabela 4 permitiu verificar que existe um grande equilíbrio a respeito dos valores médios dos tratamentos, independente do local em que se encontram, pois não existe em princípio, um tratamento que se destaca em todos os locais.

A seguir, foram construídos e avaliados separadamente, os oito modelos em que o efeito de blocos foi admitido fixo (M01-M08), e os oito em que foi considerado aleatório (M09-M16) para os valores de produção das plantas de cana-de-açúcar nos quatro locais.



Na comparação dos modelos com efeito fixo para blocos (Tabela 5), o modelo M03 - BFExp, por apresentar um valor de AIC de 735,90, ou seja, o menor dentre os avaliados, é o mais adequado segundo sua verossimilhança para modelar este conjunto de dados. Este resultado implica na indicação de um modelo com efeito fixo para blocos, com estrutura de dependência espacial dada pela função exponencial e com a mesma variância para todos os locais. O modelo BFExp superou os modelos clássicos M01-BF e M02-BFH, o que representa a existência de dependência entre as parcelas, exigindo uma abordagem mais complexa que a convencional.

Tabela 5 – Valores obtidos por meio do critério de Akaike (AIC) para os modelos com blocos de efeito fixo, ajustados sem e com estruturas de correlação espacial

Modelo	BF	BFH	BFExp	BFExpH	BFGau	BFGauH	BFESf	BFESfH
AIC	785,80	761,31	735,90	740,66	754,74	751,85	736,08	752,18

Similarmente, foram construídos e analisados os demais modelos, em que o efeito de blocos foi considerado aleatório. Baseados nesta comparação (Tabela 6), o modelo M12 - BAExpH, apresentou um valor de AIC de 837,90, ou seja, o menor dentre os avaliados, sendo então o mais adequado nesta abordagem, segundo as funções de verossimilhança. Este fato implica novamente na indicação de um modelo misto com estrutura de dependência espacial dada pela função exponencial, porém desta vez, com variâncias diferentes para cada local, além de efeito aleatório para blocos. O modelo BAExpH também superou os modelos clássicos M01-BA e M02-BAH, o que representa mais uma vez a existência de dependência entre as parcelas exigindo uma abordagem diferenciada.

Tabela 6 – Valores obtidos por meio do critério de Akaike (AIC) para os modelos com blocos de efeito aleatório, ajustados sem e com estruturas de correlação espacial

Modelo	BA	BAH	BAExp	BAExpH	BAGau	BAGauH	BAESf	BAESfH
AIC	872,76	847,17	842,38	837,90	847,24	841,80	838,57	839,01

A respeito das Tabelas 5 e 6 pode-se dizer ainda que os modelos clássicos, que não consideram funções espaciais foram menos indicados em relação a maioria dos modelos mais elaborados.

Definidos os 16 modelos, e selecionados os mais adequados para os dados em estudo (M03 e M12), foram construídos os quadros de análise de variância para os fatores de variação de efeitos fixos, com o propósito de verificar a significância de tratamentos (T), de locais (L) e da interação  $T \times L$ . As estatísticas para o teste Wald- $F$  nesta abordagem, foram utilizadas com um nível de significância de 5%, e os resultados estão apresentados nas Tabelas 7 e 8.

Tabela 7 – Graus de Liberdade (GL), Estatística  $F$  e respectivos níveis descritivos (valores- $p$ ) associados às fontes de variação de efeito fixo do modelo selecionado com blocos de efeito fixo (M03 - BFExp)

Fontes de variação	GL	F	valor- $p$
Tratamento (T)	5	0,38	0,8628
Local (L)	3	6,15	0,2862
$T \times L$	15	1,60	0,0971
Bloco (Local)	16	2,54*	0,0211

(\*) valor- $p < 0,05$

Baseados na Tabela 7, nota-se que para o modelo M03 - BFExp, apenas o efeito de blocos dentro de locais foi considerado significativo ao nível de 5% de significância. Neste modelo não se fez necessário o desdobramento do número de graus de liberdade da interação  $T \times L$  por esta não ter se apresentado significativa. Além disso, também não foram significativos os efeitos de tratamentos e locais estudados individualmente.

Tabela 8 – Graus de Liberdade (GL), Estatística  $F$  e respectivos níveis descritivos (valores- $p$ ) associados às fontes de variação de efeito fixo do modelo selecionado com blocos de efeito aleatório (M12 - BAExpH)

Fontes de variação	GL	F	valor- $p$
Tratamento (T)	5	0,30	0,9111
Local (L)	3	51,01*	0,0028
$T \times L$	15	1,62	0,0993

(\*) valor- $p < 0,05$

Por outro lado, considerando-se o modelo M12 - BAExpH, segundo a Tabela 8, foi verificado que mesmo quando foi feita a opção pelo modelo com efeito aleatório para blocos, a interação  $T \times L$  não foi significativa, tão pouco o efeito de tratamentos foi significativo. Porém, diferentemente do modelo M03, este modelo detectou diferenças significativas entre locais.

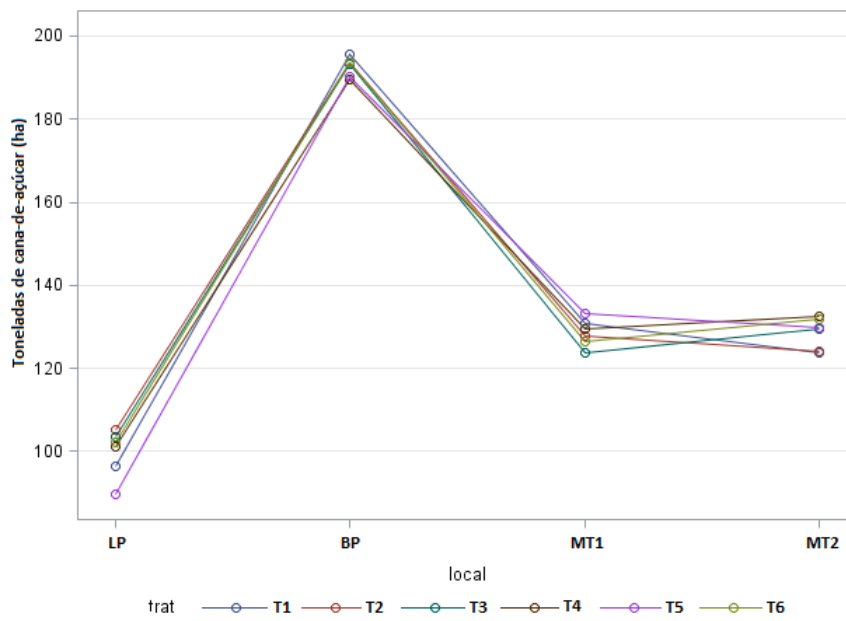


Figura 4 – Produção média de cana-de-açúcar, estimada por tratamento (trat) e por local, segundo o modelo BFExp

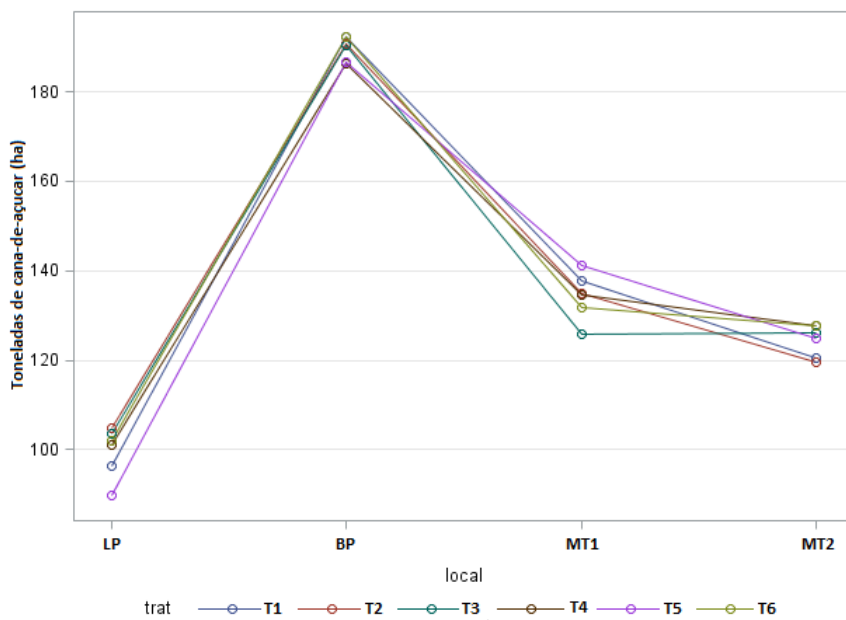


Figura 5 – Produção média de cana-de-açúcar, estimada por tratamento (trat) e por local, segundo o modelo BAExpH

Em seguida, foi utilizado o teste de Tukey-Kramer para comparar os contrastes entre os locais, estudados dois a dois, no modelo M12 - BAExpH. Os resultados deste teste estão dispostos de maneira resumida na Tabela 9, ficando evidente a superioridade do local BP diante dos demais, segundo os valores médios estimados para toneladas métricas de cana-de-açúcar por hectare ( $\text{tm h}^{-1}$ ). Vale ressaltar que o modelo M03 - BFExp não detectou esta diferença entre locais.

A partir das Figuras 4 e 5 foram avaliados os comportamentos em relação à produção de cana-de-açúcar de um local para outro com as médias estimadas para  $\text{tm h}^{-1}$  segundo seus respectivos modelos. Os gráficos evidenciaram a ausência de interação entre tratamentos e locais, confirmando os resultados expostos nas Tabelas 4 e 5. Vale ressaltar que os tratamentos dentro de cada local possuem comportamentos muito parecidos.

Tabela 9 – Resumo do teste Tukey-Kramer para o modelo BAExpH com agrupamento dos locais segundo a produção média (Prod) de cana-de-açúcar ( $\text{tm h}^{-1}$ )

Locais	Prod	Grupos
BP	189,76	a
MT1	134,32	b
MT2	124,45	b
LP	99,64	b

Médias seguidas por letras iguais não diferem significativamente pelo teste Tukey-Kramer (5%)

A seguir, visando verificar a qualidade dos dois modelos ajustados, foram apresentadas as Figuras 6 e 7 representando os valores preditos contra os resíduos condicionais estudentizados (a), histogramas (b) e gráficos quantil-quantil (c) para os modelos selecionados M03 - BFExp e M12 - BAExpH, respectivamente, para a variável resposta produção de cana-de-açúcar ( $\text{tm h}^{-1}$ ).

Nitidamente, a partir das Figuras 6(a) e 7(a), verificou-se que os resíduos (eixo y) dos modelos M03 e M12 encontram-se entre -2 e 2, com distribuição aleatória em torno da média zero, sem apresentar qualquer tendência, e assim, satisfazendo a condição de independência, reforçando ainda mais que estes modelos são adequados para descrever o comportamento dos dados. Em relação aos valores preditos, os dois gráficos apresentaram no eixo x, uma lacuna vazia em torno de 160, exibindo um salto que existe entre os valores observados nos dados originais.

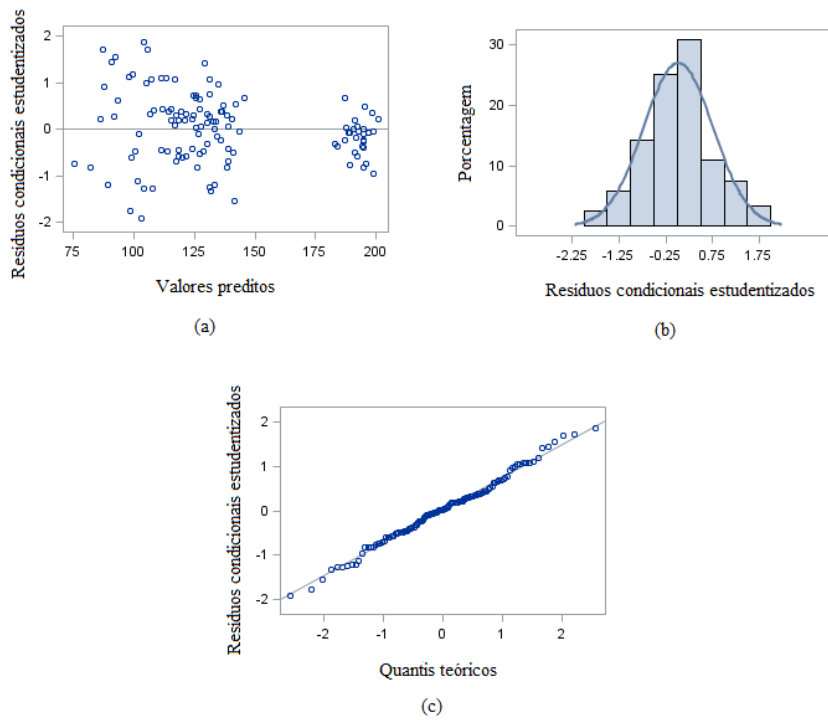


Figura 6 – Gráficos dos resíduos condicionais estudentizados: (a) em função dos valores preditos, (b) histograma e (c) quantil-quantil para o modelo BFExp

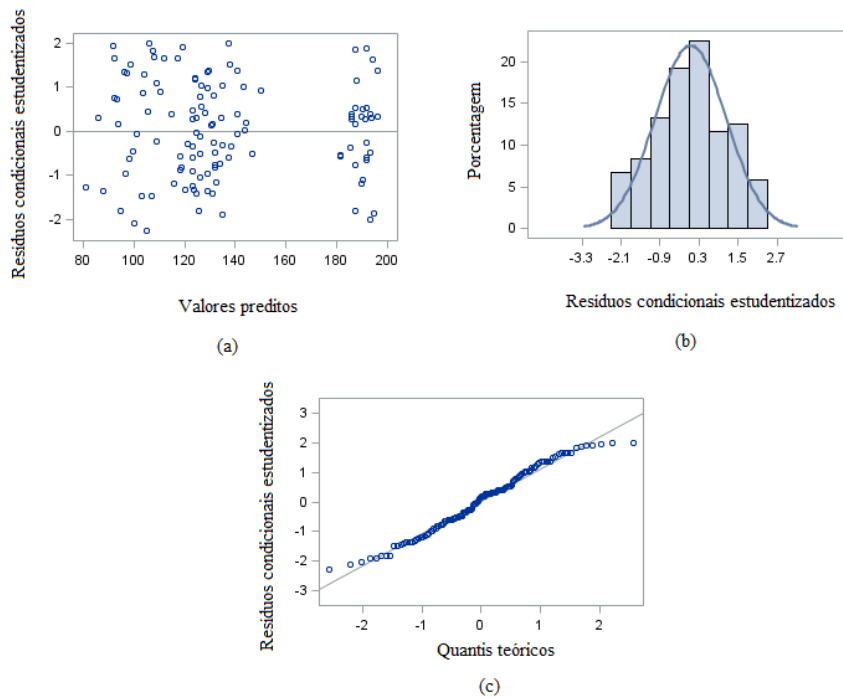


Figura 7 – Gráficos dos resíduos condicionais estudentizados: (a) em função dos valores preditos, (b) histograma e (c) quantil-quantil para o modelo BAExpH

Adicionalmente, nas Figuras 6(b) e 7(b), pode-se dizer que os histogramas construídos para a variável em estudo, apresentaram um comportamento satisfatório seguindo uma distribuição normal.

Além disso, as Figuras 6(c) e 7(c) representam os gráficos quantil-quantil dos resíduos condicionais estudentizados contra os quantis teóricos da distribuição normal relativos aos modelos M03 e M12. Segundo estes gráficos, os resíduos apresentaram-se bem ajustados e também satisfizeram as pressuposições de normalidade. Por fim, pode-se desconsiderar a presença de valores extremos (outliers).

Tabela 10 – Estimativas de componentes de variância ( $\sigma^2$ ) e do parâmetro espacial ( $\rho$ ) do modelo BFExp

CompVar	Estimativas
$\rho$	7,31
$\sigma_\varepsilon^2$	634,90

Tabela 11 – Estimativas de componentes de variância ( $\sigma^2$ ) e do parâmetro espacial ( $\rho$ ) do modelo BAExpH

Local	CompVar	Estimativas
	$\sigma_{b(l)}^2$	39,40
LP	$\sigma_\varepsilon^2$	480,07
LP	$\rho$	5,28
BP	$\sigma_\varepsilon^2$	53,22
BP	$\rho$	$2,22 \times 10^{-17}$
MT1	$\sigma_\varepsilon^2$	80,53
MT1	$\rho$	$9,5 \times 10^{-19}$
MT2	$\sigma_\varepsilon^2$	147,42
MT2	$\rho$	124,85

A respeito das estimativas de componentes de variância, na Tabela 10 são apresentados as estimativas obtidas no modelo M03 - BFExp para o resíduo ( $\sigma_\varepsilon^2$ ) e para o parâmetro do modelo exponencial ( $\rho$ ), e desta forma temos o alcance (a) deste modelo igual a  $3\rho$ , ou seja, igual a 21,93, enquanto na Tabela 11, por sua vez, foram apresentadas as estimativas do componente de blocos dentro de locais ( $\sigma_{b(l)}^2$ ), dos componentes residuais para cada local,  $\sigma_\varepsilon^2$ , além do parâmetro espacial  $\rho$ . Nota-se que os valores de  $\rho$  foram praticamente nulos nos locais BP e MT1 indicando uma dependência espacial extremamente fraca nestes dois locais.

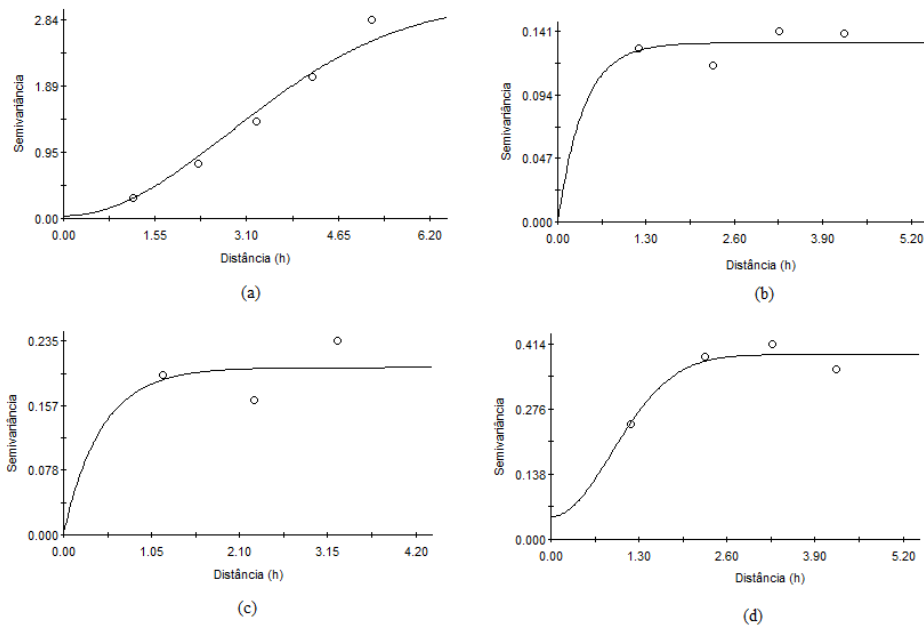


Figura 8 – Gráficos dos semivariogramas empíricos dos resíduos condicionais estudentizados do modelo BFExp para os modelos ajustados para cada local: (a) gaussiano no local LP, (b) exponencial no local BP, (c) exponencial no local MT1 e (d) gaussiano no local MT2

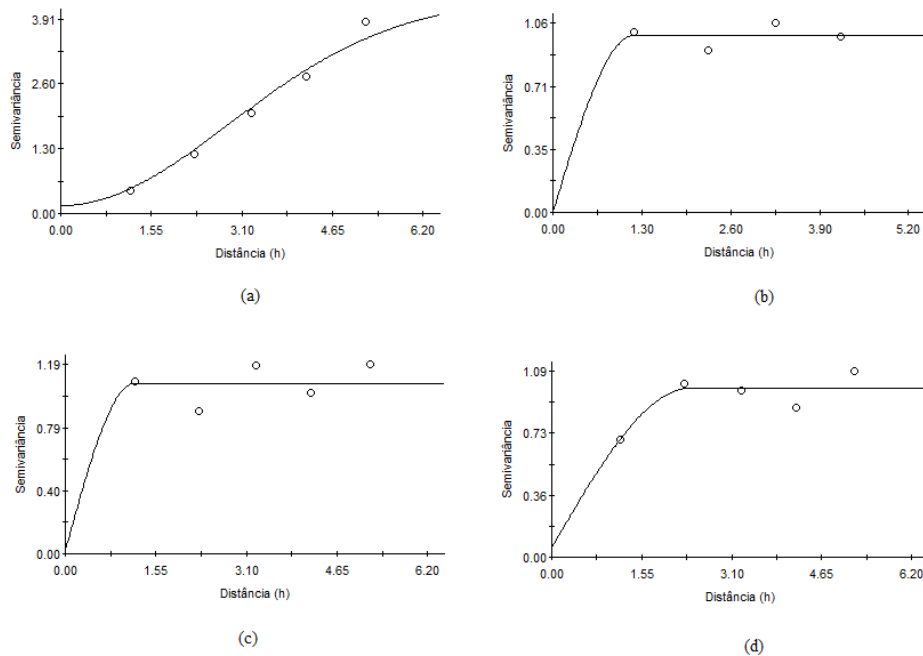


Figura 9 – Gráficos dos semivariogramas empíricos dos resíduos condicionais estudentizados do modelo BAExpH para os modelos ajustados para cada local: (a) gaussiano no local LP, (b) exponencial no local BP, (c) exponencial no local MT1 e (d) gaussiano no local MT2

Em seguida, nas Figuras 8 e 9 são apresentados os semivariogramas construídos a partir dos resíduos condicionais estudentizados obtidos quando analisado o grupo de experimentos, para cada um dos locais em estudo, considerando os modelos M03 - BFExp e M12 - BAExpH. Nota-se, que apesar de terem sido ajustados modelos geoestatísticos nas oito situações ilustradas nestas duas Figuras, apenas os locais LP e MT2 possuem comportamentos segundo suas semivariâncias, semelhantes ao semivariograma teórico (Figura 1). Este fato se assemelha com o anteriormente verificado na Tabela 11 em que o alcance estimado no modelo M12 - BAExpH foi praticamente nulo.

Os dois modelos selecionados foram bem ajustados e forneceram resultados bastante conclusivos, porém, devido ao fato de a interação tratamentos  $\times$  locais não ter se apresentado significativa em nenhum destes modelos, não foi possível extrair conclusões gerais a respeito do grupo de experimentos. Desta forma, optou-se por realizar adicionalmente, as análises individuais, ajustando modelos específicos para cada um dos quatro locais em questão, verificando, para cada caso, a necessidade de incorporação de funções geoestatísticas. Para este propósito, foram comparados os modelos ímpares também em dois grupos: (i) os que apresentam efeito fixo para blocos (M01, M03, M05 e M07), e (ii) os que apresentam tal efeito aleatório (M09, M11, M13 e M15).

Tabela 12 – Valores obtidos por meio do critério de Akaike (AIC) para os quatro modelos com efeito fixo para blocos, sem e com estruturas de correlação espacial, para as análises individuais em cada um dos quatro locais

Local	Modelo			
	BF	BFExp	BFGau	BFEsf
LP	217,31	187,94	200,46	187,94
BP	173,95	175,94	175,90	175,85
MT1	181,29	183,29	183,29	183,29
MT2	188,76	186,97	183,79	184,69

A Tabela 12 apresenta os valores calculados segundo Akaike (AIC) para os quatro modelos de interesse. A partir desta Tabela, foi possível observar que nos locais LP (modelo BFExp) e MT2 (modelo BFGau), os modelos associados a funções geoestatísticas foram os mais adequados, enquanto nos locais BP (modelo BF) e MT1 (modelo BF), tal estrutura de dependência espacial não foi necessária. De maneira similar, na Tabela 13 para modelos que admitiram o efeito de blocos como aleatório, os locais BP (modelo BA)



e MT1 (modelo BA) foram bem ajustados com os modelos mais simples, enquanto os locais LP (modelo BAGau) e MT2 (modelo BAGau) exigiram o uso de modelos lineares mistos com estrutura de dependência espacial para a matriz  $\mathbf{R}$ .

Tabela 13 – Valores obtidos por meio do critério de Akaike (AIC) para os quatro modelos com efeito aleatório para blocos, sem e com estruturas de correlação espacial, para as análises individuais em cada um dos quatro locais

Local	Modelo			
	BA	BAExp	BAGau	BAEsf
LP	242,52	227,72	224,12	224,28
BP	192,02	194,02	194,02	194,02
MT1	205,58	207,54	207,56	207,56
MT2	211,06	210,25	206,69	208,37

Posteriormente, com base nos oito modelos selecionados anteriormente, foram construídos quadros de análise de variância, e estudada, em cada caso, a estatística  $F$  para verificar a significância do efeito de tratamentos em cada modelo, conforme descrito na Tabela 14. Similarmente ao ocorrido nas Tabelas 7 e 8, na maioria dos modelos (sete deles), o efeito de tratamentos não foi significativo, ou seja, não existem diferenças significativas entre os seis estimulantes de cana-de-açúcar em estudo, da mesma forma como ocorreu com os dois modelos de análise conjunta.

Tabela 14 – Estatística  $F$  e significância para tratamentos nos modelos selecionados a partir de AIC, dentre os que consideraram o efeito de blocos como fixo (BF) e os que o consideraram aleatório (BA)

Local	Modelo	
	BF	BA
LP	2,15	3,63*
BP	0,66	0,66
MT1	1,73	1,73
MT2	1,55	1,66

(\*) valor-p < 0,05

Diante destes resultados, não existe vantagem em se trabalhar com as análises individuais, já que não são mais sensíveis na obtenção de diferenças significativas entre tratamentos. Assim, o modelo mais adequado para estimação dos efeitos fixos e predição dos efeitos aleatórios em estudo, é o modelo M12 - BAExpH, pois trata-se de

um modelo bem ajustado segundo a análise de resíduos, e capaz de detectar diferenças significativas entre locais, capacidade não verificada pelo modelo M03 - BFExp.

## 4.2 Avaliação do grupo de experimentos simulados

Para o estudo de simulação, foram definidos nove cenários (seção 3.2.2). A meta neste estudo, é verificar a quantidade e a porcentagem de casos em que os modelos com funções geoestatísticas na matriz  $\mathbf{R}$  (M03-M08) são estatisticamente superiores aos modelos clássicos (M01 - BF e M02 - BFH).

Tabela 15 – Estudo dos 500 grupos de experimentos gerados para cada cenário, com resíduos obtidos pelas funções exponencial ( $\mathbf{R}_{exp}$ ), gaussiana ( $\mathbf{R}_{gau}$ ) e esférica ( $\mathbf{R}_{esf}$ ), na verificação da quantidade e porcentagem de modelos mais adequados sem (M1-M2) versus com (M3-M8) estrutura de dependência espacial, para os três IDE

Matriz $\mathbf{R}$	Modelos	IDE		
		10%	50%	90%
$\mathbf{R}_{exp}$	M1-M2	126 (25,2%)	80 (16,0%)	33 (6,6%)
	M3-M8	374 (74,8%)	420 (84,0%)	467 (93,4%)
$\mathbf{R}_{gau}$	M1-M2	124 (24,8%)	69 (13,8%)	12 (2,4%)
	M3-M8	376 (75,2%)	431 (86,2%)	488 (97,6%)
$\mathbf{R}_{esf}$	M1-M2	126 (25,2%)	75 (15,0%)	61 (12,2%)
	M3-M8	374 (74,8%)	425 (85,0%)	439 (87,8%)

Segundo a Tabela 15, para a simulação de 500 grupos de experimentos que foi realizada, conforme aumenta-se o IDE, aumenta-se a porcentagem de casos em que se apresentaram vantajosos os modelos que incorporam a dependência espacial por meio de funções geoestatísticas na matriz  $\mathbf{R}$  (M03-M12). Neste estudo verificou-se também, que independentemente da função espacial escolhida para a matriz  $\mathbf{R}$ , os comportamentos são bem parecidos no confronto entre M01-M02 versus M03-M08.

Tabela 16 – Esboço dos três modelos mais adequados em nove cenários, a partir da quantidade e porcentagem de vezes que cada um deles foi selecionado, no estudo de 500 grupos de experimentos gerados para as três configurações de ( $\mathbf{R}$ ) com resíduos obtidos por meio das funções exponencial ( $\mathbf{R}_{exp}$ ), gaussiana ( $\mathbf{R}_{gau}$ ) e esférica ( $\mathbf{R}_{esf}$ ), para os três IDE (10%, 50% e 90%)

Matriz $\mathbf{R}$	IDE					
	10%		50%		90%	
$\mathbf{R}_{exp}$	BFExpH	139 (27,8%)	BFExpH	259 (51,8%)	BFExpH	339 (67,8%)
	BFH	97 (19,4%)	BFESfH	88 (17,6%)	BFESfH	84 (16,8%)
	BFExp	69 (13,8%)	BFH	52 (10,4%)	BFGauH	44 (8,8%)
$\mathbf{R}_{gau}$	BFExpH	142 (28,4%)	BFExpH	269 (53,8%)	BFExpH	408 (81,6%)
	BFH	94 (18,8%)	BFESfH	85 (17,0%)	BFESfH	53 (10,6%)
	BFESfH	72 (14,4%)	BFH	60 (12,0%)	BFGauH	21 (4,2%)
$\mathbf{R}_{esf}$	BFExpH	141 (28,2%)	BFExpH	255 (51,0%)	BFExpH	338 (67,6%)
	BFH	91 (18,2%)	BFESfH	91 (18,2%)	BFESfH	68 (13,6%)
	BFESfH	68 (13,6%)	BFH	65 (13,0%)	BFH	61 (12,2%)

Por fim, na Tabela 16, o modelo M04 - BFExpH apresentou-se o mais adequado em todos os cenários, mesmo quando a matriz  $\mathbf{R}$  foi obtida por meio dos modelos de dependência espacial gaussiano ou esférico. Adicionalmente, o modelo M04 não apresentou destaque quando obtido por meio de função geoestatística, em relação às demais funções.

## 5 CONSIDERAÇÕES FINAIS

Com o intuito de realizar uma modelagem alternativa, que seja capaz de explicar uma possível dependência espacial entre as amostras, foram analisados dados de produção de cana-de-açúcar em toneladas métricas por hectare, a partir da elaboração e construção de diversos modelos, alternando-se entre efeito de blocos como fixo ou aleatório, e entre funções geoestatísticas exponencial, gaussiana e esférica na matriz de variâncias e covariâncias residual.

A partir do critério de Akaike (AIC), foram selecionados o melhor modelo dentre os que possuem efeito de blocos fixo, e o melhor dentre os modelos em que blocos foi considerado de efeito aleatório. Nestes dois modelos, a interação entre tratamentos e locais não foi significativa, e assim, de maneira similar ao grupo de experimentos, foram selecionados modelos e realizadas as análises individuais para cada caso (oito análises), sendo que na grande maioria delas, não foi verificada diferença significativa entre tratamentos.

A respeito das análises de grupos de experimentos, pode-se afirmar que os modelos construídos com a incorporação de estruturas mais complexas para a matriz  $\mathbf{R}$  residual a partir de funções geoestatísticas, forneceram resultados estatisticamente mais vantajosos em comparação aos obtidos pelas análises realizadas com modelos clássicos. A análise de resíduos foi satisfatória dando maior credibilidade aos resultados. O modelo M12 - BAExpH foi apontado como o mais indicado para realizar a estimação de efeitos fixos e predição de efeitos aleatórios, por ter sido mais sensível ao detectar diferenças significativas entre locais.

Por fim, o estudo de simulação deu um suporte a mais aos resultados obtidos via dados reais, possibilitando a visualização de diversos cenários envolvendo dependência espacial forte e fraca em que os modelos alternativos propostos, com estruturas de dependência espacial, se destacaram em relação aos clássicos, sendo os mais indicados na maioria das situações.



## REFERÊNCIAS

- AKAIKE, H. A New Look at the Statistical Model Identification. **IEEE Transactions on automatic control**, New York, v. 19, n. 6, p. 716-723, Dec. 1974.
- ALCARDE, R. **Modelos lineares mistos em dados longitudinais com o uso do pacote ASReml-R**. 2012. 156 p. Tese (Doutorado em Ciências) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2012.
- ANDERSON, R.L.; BANCROFT, T.A. **Statistical theory in research**. New York: McGraw-Hill, 1952. 399 p.
- ASOCIACIÓN DE AZUCAREROS DE GUATEMALA. Disponível em: <<http://www.azucar.com.gt/economia.html>>. Acesso em: 27 out. 2013.
- BANZATTO, D.A.; KRONKA, S.N. **Experimentação agrícola**. Jaboticabal: FUNEP, 2006. 247 p.
- BARBIN, D. **Componentes de variância: teoria e aplicações**. Piracicaba: FEALQ, 1993. 120 p.
- BARBIN, D. **Planejamento e análise estatística de experimentos agronômicos**. Londrina: Mecenas, 2013. 214 p.
- BARBOSA, M. **Uma abordagem para análise de dados com medidas repetidas utilizando modelos lineares mistos**. 2009. 118 p. Dissertação (Mestrado em Agronomia) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2009.
- CAMPOS, H. **Estatística aplicada à experimentação com cana-de-açúcar**. Piracicaba: FEALQ, 1984. 292 p.
- CARDOSO, J.C. **Ácido giberélico ( $GA_3$ ) na indução do florescimento de orquídeas** 2007. 50 p. Dissertação (Mestrado em Ciências Biológicas (Botânica)) -Universidade Estadual Paulista, Botucatu, 2007.
- CASANOVA, F.; MACHIAVELLI, R.; BALZARINI, M. Models for multi-environment yield trials with fixed and random block effects and homogeneous and heterogeneous residual variances. **Journal of Agriculture of the University of Puerto Rico**, Mayaguez, v. 91, p. 117 - 131, 2007.
- CECON, P.R. **Alternativas de análise de experimentos em látice e aplicações no melhoramento vegetal**. 1992. 109 p. Tese (Doutorado em Agronomia) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1992.

CENTRO GUATEMALTECO DE INVESTIGACIÓN Y CAPACITACIÓN DE LA CAÑA DE AZÚCAR. Disponível em: <<http://www.cengicana.org/es/publicaciones/otras-publicaciones/boletines-estadisticos>>. Acesso em: 5 dez. 2013.

COCHRAN, W.G. The combination of estimates from different experiments. **Biometrics**, Arlington, v. 10, p. 101-129, 1954.

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

CRESSIE, N.A.C. **Statistics for spatial data analysis**. New York: J.Wiley, 1993, 900 p.

DIGGLE, P.; HEAGERTY, P.; LIANG, K.; ZEGER, S. **Analysis of longitudinal data**. 2. ed. New York: Oxford University Press, 2002. 396 p.

Di RIENZO J.A.; MACCHIAVELLI, R.; CASANOVES, F. **Modelos Mixtos en InfoStat**. Disponível em: <<http://www.infostat.com.ar>>. Acesso em: 17 out. 2013.

DUARTE, J.B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento vegetal**. 2000. 293 p. Tese (Doutorado em Agronomia) -Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2000.

EMBRAPA. Disponível em: <[http://www.cpafrro.embrapa.br/media/arquivos/publicacoes/livro\\_plantastropicais-2.pdf](http://www.cpafrro.embrapa.br/media/arquivos/publicacoes/livro_plantastropicais-2.pdf)>. Acesso em: 3 abr. 2014.

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, p. 420-428, 1993.

FAI, A.H.T; CORNELIUS, P.I. Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. **Journal of Statistical Computation and Simulation**, London, v. 54, n. 4, p. 363-378, 1996.

FAO. Disponível em: <[http://faostat3.fao.org/faostat-gateway/go/to/download/Q/\\*/S](http://faostat3.fao.org/faostat-gateway/go/to/download/Q/*/S)>. Acesso em: 10 jan. 2014.

FARACO, M.A. **Qualidade do ajuste de modelos geoestatísticos utilizados na agricultura de precisão**. 2006. 93 p. Dissertação (Mestrado em Engenharia Agrícola) -Universidade Estadual do Oeste do Paraná, Cascavel, 2006.

FISHER, R.A. On the mathematical foundations of theoretical statistics. **Philosophical Transactions of the Royal Society of London**. Series A, London, v. 222, p. 309-368, 1922.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B.R. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. **Biometrics**, Arlington, v. 51, n. 4, p. 1140-1450, Dec. 1995.

GREGOIRE, T.G.; SCHABENBERGER, O.; BARRETT, J.P. Linear Modelling of Irregularly Spaced, Unbalanced, Longitudinal Data from Permanent Plot Measurements. **Canadian Journal of Forest Research**, Ottawa, v. 25, n. 1, p. 137-156, Jan. 1995.

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

GUMEDZE, F.N.; DUNNE, T.T. Parameter estimation and inference in the linear mixed model. **Linear Algebra and its Applications**, Cape Town, v. 435, p. 1920-1944, 2011.

HADARBACH, D. **The design and analysis of agronomic field experiments with large numbers of treatments in spatially heterogeneous environments**. 1996. 164 p. Thesis (Ph.D.) University of Nebraska, Lincoln, 1996.

HALIMI, R. **Nonlinear mixed-effects models and nonparametric inference, a method basea on bootstrap for the analysis of non-normal repeated measures data in bios-tatistica**. 2005, 247 p. Tese (Estadística) -Universitat de Barcelona, Barcelona, 2005.

HARTLEY, H.O. The Maximum F-Ratio as Short-Cut Test for Heterogeneity of Variances. **Biometrika**, Oxford, v. 37, p. 308-312, 1950.

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

HARVILLE, D.A. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. **Journal of the American Statistical Association**, Alexandria, v. 72, n. 2, p. 320-338, June 1977.

HENDERSON, C.R. Estimation of Variance and Covariance Components. **Biometrics**, Arlington, v. 9, n. 2, p. 226-252, June 1953.

HENDERSON, C.R. **Applications of Linear Models in Animal Breeding**. Guelph: University of Guelph, 1984. 462 p.

KEMPTHORNE, O. **The Design and Analysis of Experiments**. New York: Wiley, 1952. 631 p.

ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 561 p.

JARAMILLO, D.F.J.; FRANCISCO, D. Efecto de la variabilidad sistemática en experimentos de fertilización con fríjol. **Revista Facultad Nacional de Agronomía**, Medellín, v. 58, p. 2717-2732, 2005.

KENWARD, M.G.; ROGER, J.H. Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. **Biometrics**, Arlington, v. 53, n. 3, p. 983-997, Sept. 1997.



LITTELL R., PENDERGAST J., NATARAJAN R. Modelling Covariance Structure in the Analysis of Repeated Measures Data. **Statistics in Medicine**, New York, v. 19, p. 1793-1819, 2002.

LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D.; SCHABENBERGER, O. **SAS ® System for Mixed Models**. 2. ed. Cari: SAS Institute Inc., 2006. 834 p.

MARCELINO, S.D.R.; IEMMA, A.F. Métodos de estimação de componentes de variância em modelos mistos desbalanceados. **Scientia Agricola**, Piracicaba, v. 56, n. 4, p. 643-652, Out./Dez. 2000.

MARTINEZ, B.R. Control de la correlacion espacial en experimentos de campo en el sector agricola. **Agronomía colombiana**, Bogotá, v. 11, n. 1, p. 83-89, 1994.

McCULLOCH, C.E.; SEARLE, S.R.; NEUHAUS, J.M. **Generalized, linear, and mixed models**. 2. ed. New York: J. Wiley, 2001. 384 p.

McLEAN, R.A.; SANDERS, W.L. Approximating degrees of freedom for standard errors in mixed linear models. **Proceedings of the Statistical Computing Section, American Statistical Association**, New Orleans, p. 50-59, 1988.

MELLO, J.M. **Geoestatística aplicada ao inventário florestal**. 2004. 111 p. Tese (Doutorado em Recursos Florestais) -Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.

MEYER, K. Estimation of genetic parameters. In: HILLS, W.G.; MACKAY, T.F.C. **Evolution and animal breeding**. Wallingford: CAB International, 1989. p. 161-167.

MOLENBERGHS, G.; VERBEKE, G. **Linear Mixed Models for Discrete Longitudinal Data**. New York: Springer-Verlag, 2005. 706 p.

NOGUEIRA, C.H.; LIMA, R.R. de; OLIVEIRA, M.S. de. Aprimoramento da análise de variância: a influência da proximidade espacial. **Revista Brasileira de Biometria**, São Paulo, v. 31, n. 3, p. 408-422, 2013.

ODA, M.L. **Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um experimento em delineamento sistemático tipo "leque"**. 2005. 72 p. Dissertação (Mestrado em Agronomia) -Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.

PANNATIER, Y. **Variowin 2.2 Software for special data analysis in 2D**. New York: Springer-Verlag, 1996. 96p.

PATTERSON, H.D.; THOMPSON, R. Recovery of inter-block information when blocks sizes are unequal. **Biometrika**, Oxford, v. 58, n. 3, p. 545-554, Dec. 1971.

PIMENTEL-GOMES, F.; GUIMARÃES, R.F. Joint analysis of experiments in complete randomized blocks with some common treatments. **Biometrics**, Arlington, v. 14, p. 521-526, 1958.

PIMENTEL-GOMES, F. An extension of the method of joint analysis of experiments in complete randomized blocks. **Biometrics**, Arlington, v. 26, p. 331-336, 1970.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15 ed. Piracicaba: FEALQ, 2009. 451 p.

PINHEIRO, J.C.; BATES, D.M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. 528 p.

PONTES, J.M.; OLIVEIRA, M.S. de. Uma proposta alternativa para a análise de experimentos de campo utilizando a geoestatística. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 1, p. 135-141, jan./fev., 2004

R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org>>. Acesso em: 18 out. 2013.

RAO, C.R. Minimum variance quadratic unbiased estimation of variance components. **Journal of Multivariate Analysis**, Amsterdam, v. 1, p. 445-456, 1971.

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

RESENDE, M.D.V. de; PRATES, D.F.; JESUS, A.; YAMADA, C.K. Estimaco de componentes de varincia e predico de valores genticos pelo mtodo da mxima verossimilhana restrita (REML) e melhor predico linear no viciada (BLUP) em Pinus. **Boletim de Pesquisa Florestal**, Colombo, n. 32/33, p. 18-45, jan./dez. 1996.

RESENDE, M.D.V. de; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Revista de Matemtica e Estatística**, Jaboticabal, v. 22, n. 2, p. 1-122, 2004.

RESENDE, M.D.V. de **Anlise estatística de modelos mistos via REML/BLUP na experimentaco em melhoramento de plantas perenes**. Colombo: Embrapa Florestas, 2007. 561 p.

RIBEIRO JR. P.J.; DIGGLE, P.J. **GeoR: A package for geostatistical analysis**. Disponível em: <<http://www.R-project.org>>. Acesso em: 27 maio 2013.

ROJAS, B.A. da **Analysis of a group of experiments in combining ability in corn**. 1951. 120 p. Dissertation (M.S.) Iowa State College of Agriculture and Mechanic Arts, Ames, 1951.

SAMRA, J.S.; ANLAUF, R.; WEBER, W.E. Spatial dependence of growth attributes and local control in wheat and oat breeding experiments. **Crop Science**, Madison, v. 30, p. 1200-1205, 1990.

SAS INSTITUTE. **SAS STAT 9.3 user's guide**. Cary, 2011. Disponível em: <<http://support.sas.com/documentation/onlinedoc/stat/indexproc.html#stat131>>. Acesso em: 4 set. 2013.

SATTERTHWAITE, F.E. Na Approximate Distribution of Estimates of Variance Components. **Biometrics Bulletin**, Washington, v. 2, n. 6, p. 110-114, Dec. 1946.

SEARLE, S. R. C. R. Henderson, the Statistician; and His Contributions to Variance Components Estimation. **Journal of Dairy Science**, Philadelphia, v. 74, n. 11, p. 4035-4044. 1991.

SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance components**. New York: John Wiley, 1992. 536 p.

SELF, S.G.; LIANG, K-Y. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Test Under Nonstandard Conditions. **Journal of the American Statistical Association**, Alexandria, v. 82, n. 398, p. 605-610, 1987.

SHAW, R.G. Maximum-likelihood approaches to quantitative genetics of natural populations. **Evolution**, Lancaster, v. 41, p. 812-826, 1987.

SILVA, A.F. da; QUARTEZANI, W.Z.; ZIMBACK, C. R. L; LANDIM, P.M.B. Aplicação da Geoestatística em Ciências Agrárias. In: SIMPÓSIO DE GEOESTATÍSTICA APLICADA EM CIÊNCIAS AGRÁRIAS, 2., 2011, Botucatu. **Minicurso...** Botucatu: FEPAF, 2011. 136 p.

STOLLER. Disponível em: <<http://www.stoller.com.br/stoller-do-brasil/publicacoes>>. Acesso em: 13 fev. 2014.

STRAM, D.O.; LEE, J.W. Variance components testing in the longitudinal mixed effects model. **Biometrics**, Arlington, v.50, p. 1171-1177, 1994.

STROUP, W.W.; BAENZIGER, P.S.; MULITZE, D.K. Removing spatial variation from wheat yield trials: a comparison of methods. **Crop Science**, Madison, v. 86, p. 62-66, 1994.

TAIZ, L., ZEIGER, E. **Plant Physiology**. 4 ed. Sunderland: Sinauer Associates. 2006. 764 p.

UPCHURCH, D.R.; EDMONDS, W.J. Statistical procedures for specific objectives. In: MUSBACH, M.J.; WILDING, L.P. (Ed.) **Spatial variabilities of soil and landforms**. Madison: Soil Science Society of America, 1991. p. 49-71.

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

WEST, B.T.; WELCH, K.B.;GALECKI, A.T. **Linear mixed models: A practical guide using statistical software**. New York: Chapman & Hall, 2007. 376 p.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013. 215 p.

YATES, F.; COCHRAN, W.G. The analysis of groups of experiments. **Journal of Agricultural Science**, Harpenden, v. 28, p. 556-580, 1938.

ZIMBACK, C.R.L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade**. 2001. 114 p. Tese de Livre-Docência (Livre-Docência em Levantamento de Solo e fotopedologia) -FCA/UNESP, Botucatu, 2001.

ZIMMERMAN, D.L.; HARVILLE, D.A. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. **Biometrics**, Arlington, v. 47, p. 223-239, Mar. 1991.



## ANEXOS



**ANEXO A - Códigos no software R para o ajuste e comparação de modelos lineares mistos, sob diferentes configurações a respeito de suas estruturas de dependência espacial, para um grupo de experimentos de produção de cana-de-açúcar.**

```

require(lattice)
require(nlme)

dados <- read.table("C:/Users/Cassio/Desktop/dados.txt", head=T)
trat <- as.factor(dados$trat)
blo <- as.factor(dados$blo)
local <- as.factor(dados$local)
la <- as.numeric(dados$la)
lon <- as.numeric(dados$lon)
resp <- as.numeric(dados$resp)
blo_local <- interaction(blo, local)
dados <- data.frame(local, trat, blo, la, lon, blo_local, resp)
str(dados)
tapply(resp, trat, mean)

##### Modelos com diferentes estruturas #####

### MODELO BF - 01
M01 <- gls(resp ~ 1+local+trat+local:trat+local/blo,
           method="REML", na.action=na.omit, data=dados)
### MODELO BFH - 02
M02 <- gls(resp~1+local+trat+local:trat+local/blo,
           weight=varComb(varIdent(form=~1|local)),
           method="REML", na.action=na.omit, data=dados)
### MODELO BFExp - 03
M03 <- gls(resp~1+trat+local+local:trat+local/blo,
           correlation=corExp(form=~la+lon|local,
                               metric="euclidean", nugget=FALSE),
           method="REML", na.action=na.omit, data=dados)
### MODELO BFExpH - 04
M04 <- gls(resp~1+trat+local+local:trat+local/blo,
           weight=varComb(varIdent(form=~1|local)),
           correlation=corExp(form=~la+lon|local,
                               metric="euclidean", nugget=FALSE),
           method="REML", na.action=na.omit, data=dados)
### MODELO BFGau - 05
M05 <- gls(resp~1+trat+local+local:trat+local/blo,
           correlation=corGaus(form=~la+lon|local,
                               metric="euclidean", nugget=FALSE),
           method="REML", na.action=na.omit, data=dados)
### MODELO BFGauH - 06
M06 <- gls(resp~1+trat+local+local:trat+local/blo,
           weight=varComb(varIdent(form=~1|local)),
           correlation=corGaus(form=~la+lon|local,
                               metric="euclidean", nugget=FALSE),
           method="REML", na.action=na.omit, data=dados)
### MODELO BFEsf - 07

```



```

M07 <- gls(resp~1+trat+local+local:trat+local/blo,
           correlation=corSpher(form=~la+lon|local,
                                metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BFesfH - 08
M08 <- gls(resp~1+trat+local+local:trat+local/blo,
           weight=varComb(varIdent(form=~1|local)),
           correlation=corSpher(form=~la+lon|local,
                                metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BA - 09
M09 <- lme(resp~1+local+trat+local:trat,
           random=list(blo_local=pdIdent(~1)),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAH - 10
M10 <- lme(resp~1+local+trat+local:trat,
           random=list(blo_local=pdIdent(~1)),
           weight=varComb(varIdent(form=~1|local)),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAExp - 11
M11 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           correlation=corExp(form=~la+lon,
                              metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAExpH - 12
M12 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           weight=varComb(varIdent(form=~1|local)),
           correlation=corExp(form=~la+lon,
                              metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAGau - 13
M13 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           correlation=corGaus(form=~la+lon,
                              metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAGauH - 14
M14 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           weight=varComb(varIdent(form=~1|local)),
           correlation=corGaus(form=~la+lon,
                              metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)
### MODELO BAesf - 15
M15 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           correlation=corSpher(form=~la+lon,
                              metric="euclidean",nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)

```

```
### MODELO BAEsfH - 16
M16 <- lme(resp~1+trat+local+local:trat,
           random=list(blo_local=pdIdent(~1)),
           weight=varComb(varIdent(form=~1|local)),
           correlation=corSpher(form=~la+lon,
                                 metric="euclidean", nugget=FALSE),
           method="REML",na.action=na.omit,data=dados)

crit_aic <- c(summary(M01)$AIC,summary(M02)$AIC,summary(M03)$AIC,
              summary(M04)$AIC,summary(M05)$AIC,summary(M06)$AIC,summary(M07)$AIC,
              summary(M08)$AIC,summary(M09)$AIC,summary(M10)$AIC,summary(M11)$AIC,
              summary(M12)$AIC,summary(M13)$AIC,summary(M14)$AIC,summary(M15)$AIC,
              summary(M16)$AIC)

modelos <- c("BF","BFH","BFExp","BFExpH","BFGau","BFGauH","BFESf",
             "BFESfH","BA","BAH","BAExp","BAExpH","BAGau","BAGauH","BAEsf","BAEsfH")
t_aic <- matrix(c(modelos,round(crit_aic,2)),ncol=2)
t_aic
```

**ANEXO B - Códigos no software SAS para o estudo dos dois modelos escolhidos (BFExp e BAExpH) para a análise conjunta dos quatro locais referentes à produção de cana-de-açúcar.**

```
*Modelo 03 - BFExp;
proc glimmix data=cana plots=(residualpanel pearsonpanel studentpanel);
class trat local rep;
model resp=trat local trat*local rep(local)/ddfm=kenwardroger;
random_residual_/subject=local type=sp(exp)(lat lon);
lsmeans local*trat/slice=local;
lsmeans local*trat/ plot=meanplot(sliceby=trat join);
run;
```

```
*Modelo 12 - BAExpH;
proc glimmix data=cana plots=(residualpanel pearsonpanel studentpanel);
class trat local rep;
model resp=trat local trat*local /ddfm=kenwardroger;
random rep(local);
random_residual_/subject=local type=sp(exp)(lat lon) group=local;
lsmeans local*trat/slice=local;
lsmeans local*trat/ plot=meanplot(sliceby=trat join);
run;
```

**ANEXO C - Códigos no software R para o estudo de simulação, visando a comparação de modelos alternando-se as funções de dependência espacial e o grau desta dependência.**

```

require(geoR)
require(nlme)
set.seed(3947)
phi <- 7; nug <- 1
p_sill <- 9
ide <- (p_sill)/(nug+p_sill)
t <- 5; b <- 5; m <- 10
vt <- 5; vb <- 5
blo <- c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5))
la <- c(rep(1,5),rep(2,5),rep(3,5),rep(4,5),rep(5,5))
lon <- rep(seq(1:5),5)
coord <- data.frame(la, lon)
modelos_bf <- c("BF","BFH","BFExp","BFExpH","BFGau","BFGauH","BFESf","BFESfH")

analise<-function(){

### LOCAL 1 ###
eft1 <- rnorm(t,0,sqrt(vt))
efb1 <- c(rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep
(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t))
## sorteando os tratamentos dentro dos blocos
trat <- c(sample(seq(1:5)),sample(seq(1:5)),sample(seq(1:5)),
sample(seq(1:5)),sample(seq(1:5)))
trat
eft1 <- eft1[trat]
## simulando resíduos com dependência espacial exponencial
simu1 <- grf(25, grid="reg", xlims = c(1, 5), ylims = c(1, 5),
cov.pars=c(p_sill,phi), lambda = 0, nugget=nug, cov.model="spherical", nsim=1)
resp <- round(m + eft1 + efb1 + simu1$data, 2)
tab1 <- data.frame(trat, blo, resp)
local <- rep(1,25)
blo_local <- interaction(blo, local)
sim1 <- data.frame(local,coord,blo_local,tab1)

### LOCAL 2 ###
eft2 <- rnorm(t,0,sqrt(vt))
efb2 <- c(rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep
(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t))
trat <- c(sample(seq(1:5)),sample(seq(1:5)),sample(seq(1:5)),
sample(seq(1:5)),sample(seq(1:5)))
eft2 <- eft2[trat]
simu2 <- grf(25,grid="reg",xlims = c(1, 5),ylims = c(1,5),cov.pars=c(p_sill,phi),
lambda = 0, nugget=nug, cov.model="spherical", nsim=1)
resp <- round(m + eft2 + efb2 + simu2$data, 2)
tab2 <- data.frame(trat, blo, resp)
local <- rep(2,25)
blo_local <- interaction(blo, local)

```

```

sim2 <- data.frame(local,coord,blo_local,tab2)

### LOCAL 3 ###
eft3 <- rnorm(t,0,sqrt(vt))
efb3 <- c(rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep
(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t))
trat <- c(sample(seq(1:5)),sample(seq(1:5)),sample(seq(1:5)),
sample(seq(1:5)),sample(seq(1:5)))
eft3 <- eft3[trat]
simu3 <- grf(25, grid="reg", xlims = c(1, 5), ylims = c(1, 5),
cov.pars=c(p_sill,phi), lambda = 0, nugget=nug, cov.model="spherical", nsim=1)
resp <- round(m + eft3 + efb3 + simu3$data, 2)
tab3 <- data.frame(trat, blo, resp)
local <- rep(3,25)
blo_local <- interaction(blo, local)
sim3 <- data.frame(local,coord,blo_local,tab3)

### LOCAL 4 ###
eft4 <- rnorm(t,0,sqrt(vt))
efb4 <- c(rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),
rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t),rep(rnorm(1,0,sqrt(vb)),t))
trat <- c(sample(seq(1:5)),sample(seq(1:5)),sample(seq(1:5)),
sample(seq(1:5)),sample(seq(1:5)))
eft4 <- eft4[trat]
simu4 <- grf(25, grid="reg", xlims = c(1, 5), ylims = c(1, 5),
cov.pars=c(p_sill,phi), lambda = 0, nugget=nug, cov.model="spherical", nsim=1)
resp <- round(m + eft4 + efb4 + simu4$data, 2)
tab4 <- data.frame(trat, blo, resp)
local <- rep(4,25)
blo_local <- interaction(blo, local)
sim4 <- data.frame(local,coord,blo_local,tab4)

### Grupo de 4 experimentos:
dados <- rbind(sim1,sim2,sim3,sim4)

##### Modelos com diferentes estruturas #####
### MODELO BF - 01
M01 <- gls(resp ~ 1+local+trat+local:trat+local/blo,
method="REML",na.action=na.omit, data=dados)
### MODELO BFH - 02
M02 <- gls(resp~1+local+trat+local:trat+local/blo,
weight=varComb(varIdent(form=~1|local)),
method="REML",na.action=na.omit,data=dados)
### MODELO BFExp - 03
M03 <- gls(resp~1+trat+local+local:trat+local/blo,
correlation=corExp(form=~1a+1on|local,
metric="euclidean",nugget=FALSE),
method="REML",na.action=na.omit,data=dados)
### MODELO BFExpH - 04
M04 <- gls(resp~1+trat+local+local:trat+local/blo,
weight=varComb(varIdent(form=~1|local)),

```

```

        correlation=corExp(form=~la+lon|local,
        metric="euclidean",nugget=FALSE),
        method="REML",na.action=na.omit,data=dados)
### MODELO BFGau - 05
M05 <- gls(resp~1+trat+local+local:trat+local/blo,
        correlation=corGaus(form=~la+lon|local,
        metric="euclidean",nugget=FALSE),
        method="REML",na.action=na.omit,data=dados)
### MODELO BFGauH - 06
M06 <- gls(resp~1+trat+local+local:trat+local/blo,
        weight=varComb(varIdent(form=~1|local)),
        correlation=corGaus(form=~la+lon|local,
        metric="euclidean",nugget=FALSE),
        method="REML",na.action=na.omit,data=dados)
### MODELO BFEsf - 07
M07 <- gls(resp~1+trat+local+local:trat+local/blo,
        correlation=corSpher(form=~la+lon|local,
        metric="euclidean",nugget=FALSE),
        method="REML",na.action=na.omit,data=dados)
### MODELO BFEsfH - 08
M08 <- gls(resp~1+trat+local+local:trat+local/blo,
        weight=varComb(varIdent(form=~1|local)),
        correlation=corSpher(form=~la+lon|local,
        metric="euclidean",nugget=FALSE),
        method="REML",na.action=na.omit,data=dados)

aic_mbf <- c(summary(M01)$AIC,summary(M02)$AIC,summary(M03)$AIC,summary(M04)$AIC,
summary(M05)$AIC,summary(M06)$AIC,summary(M07)$AIC,summary(M08)$AIC)
aic_bf <- matrix(c(modelos_bf,round(aic_mbf,2)),ncol=2)

return(cbind(aic_bf))}
resultados<-list()
for(i in 1:50){
result<-try(analise())
if(class(result)=="matrix"){
    resultados[[i]]<-result
    cat("final da iteraçao ", i)
    cat("-----")
    cat(" ")
}
}
resultados<-resultados[!unlist(lapply(resultados,is.null))]
resultados
resumo<-function(matriz){
    c(matriz[which.min(matriz[,2]),1])
}
final<-lapply(resultados,resumo)
final<-matrix(unlist(final),byrow=T,ncol=1)
final<-as.data.frame(final)
summary(final)

```